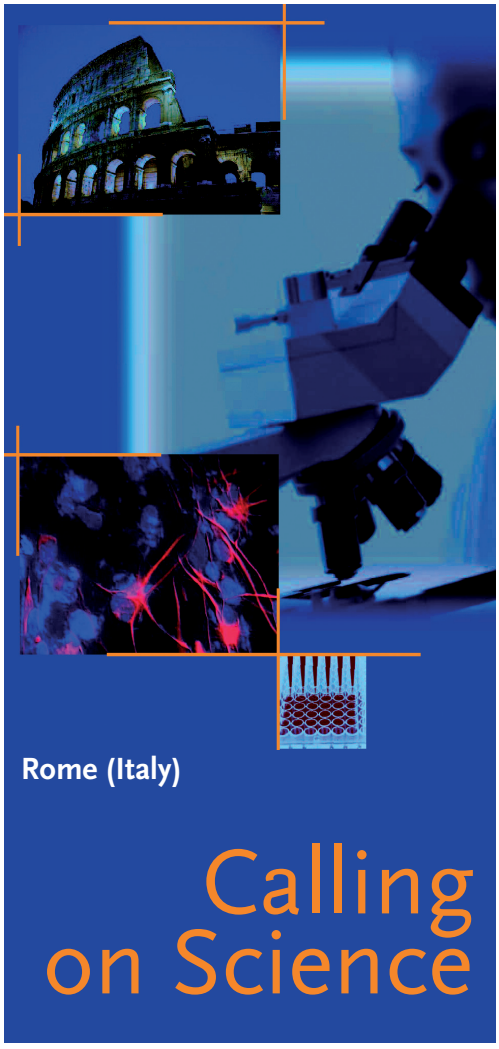
Proceedings 7th World Congress Rome 2009
Volume 27, Spec. Issue 1
ISSN 1868-596X
1-334 (2010)

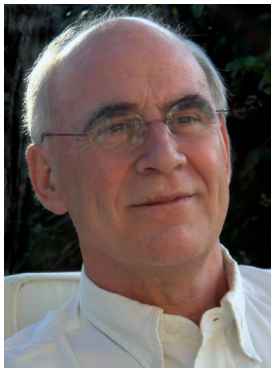
ALTEX

ALTERNATIVES TO ANIMAL EXPERIMENTATION



Congress Chairmen's Preface

Thomas Hartung



Herman B. W. M. Koëter

Dear colleagues, participants and supporters of WC7 and other readers of these proceedings

Ten months have passed since we spent five exciting days of science, networking and friendship together in beautiful Rome, Italy. It is with great pleasure (and some pride) that we now present the proceedings of this conference. This completes the commitment we took on in 2005 on occasion of WC5. We have enjoyed the support of many remarkable individuals in the steering groups, as corporate supporters and in the AIM group. Together we shaped and executed a programme that has received praise from many sides. The interest of both the scientific and lay community was evidenced by reports in *Nature* and many other journals, websites, blogs, etc. Those who participated in WC7 will share positive memories of days of scientific exchange, recollection of challenges lying around their work and ahead as well as the infusion of enthusiasm to carry on with our common endeavours. We are confident that with these proceedings also those who were not able to participate can appreciate the diversity and high standard of presentations offered at WC7.

It is the first time that the proceedings are made available electronically instead of in printed form. All registered participants will receive them as a DVD. In addition they are available on the websites of the journal ALTEX (www.altex.ch) and the Center for Alternatives to Animal Testing webportal AltWeb (http://altweb.jhsph.edu/). These portals make them accessible to a broad community in form of open source material, thus further feeding the lively developments in the field of alternatives to animal experiments. Selected articles chosen by the Editor of ALTEX will be published in addition as “Highlights from WC7” in the printed issues of the journal.

The conference took place exactly 50 years after the publication of Russel & Burch’s “The Principles of Humane Experimental Technique”, establishing the 3Rs principle and thus a cornerstone of alternatives to animal testing. It was therefore an appropriate time to appoint three prominent individuals who have steered the implementation of the 3Rs like no others as “Patrons of laboratory animal welfare in the life

sciences": Michael Balls, Alan Goldberg and Horst Spielmann. Though now retired from the positions that allowed them to shape our field, they are still sources of enthusiasm and insight to those who follow them, as witnessed again at WC7.

The congress took place at a time when major developments all over the world were driving the field of alternatives: The ban of animal testing for cosmetics in Europe, the start of REACH registration, the intense discussion on the "Toxicity Testing in the 21st Century" vision of the US National Academy of Science, the creation of the Korean Center for the Validation of Alternative Methods and the first chair for alternative methods in India are only a few examples of the developments of 2009. WC7 also saw the first announcements of the largest call for proposals in history, i.e. € 50 million shall be made available by COLIPA and the European Commission for systemic toxicity testing, and the first transatlantic 3Rs centre, CAAT-EU. Finally, WC7 also put a finger on another very important animal welfare issue not addressed at earlier World Congresses, namely that of corporate social responsibility (CSR), an ethical approach aiming, among other aspects, at an ethical or fair treatment of animals, be it food producing animals or experimental animals in the production of goods, foods and chemicals such as pharmaceuticals. We hope this broader approach will pick up further and take root in World Congresses to come, as it does increasingly in our society.

Most importantly, however, WC7 has shown that the field of alternative methods is embracing science and science is embracing alternative approaches. We saw how modern technologies are shaping a second generation of alternative methods and how some recent developments in the life sciences were fuelled by the wish to overcome animal experimentation. The area presents itself as one of the liveliest in the life sciences, at the same time bridging to practical applications. The regulatory use of these new approaches is imminent and few fields are resonating as much with political discussions.

Experimental animal welfare and safety assessments are more than a matter of science; they are key societal interests. WC7 has mirrored this energy with stakeholders coming to the congress with different views and approaches and leaving with increased understanding, compromise and a vision for the future. It was a privilege to serve WC7 by preparing for it, moderating it and now harvesting its fruits. We would like to wish WC8 and its contributors all success and hope that this series of congresses continues with the same level of enthusiasm for our shared goals.

Thomas Hartung,
Director, Center for Alternatives to Animal Testing (CAAT),
Baltimore, USA
thartung@jhsph.edu

Herman B.W.M. Koëter,
Managing Director,
Orange House Partnership npo,
Brussels, Belgium
herman.koeter@orangeOhouse.eu

Contents Proceedings WC7, Rome 2009

Congress Chairmen´s Preface

Plenary Lectures

Early Morning Sessions

MS3 Animal welfare associations

MS5 3R centers

MS6 Associations

Satellite Meeting

A satellite symposium to the 7th World Congress on Alternatives and Animal Use in the Life Sciences, organized under auspices of CELLTOX (Associazione Italiana di Tossicologia *in vitro*) and ESTIV (European Society of Toxicology in Vitro)

Plenary Lecture

PL3

Brueghel's Two Monkeys: Passing the Final Exam in the History of Mankind

Ingrid E. Newkirk

People for the Ethical Treatment of Animals, Norfolk, VA, USA

Summary

As society's ethical values expand over time, we understand that we must have consideration for more than just ourselves, our race, our gender, and our species. This talk confronts our biases and provides food for thought in moving beyond our current understanding of human-animal relations. History provides a lens through which our current norms can be viewed. It allows us to discern how our behavior might be perceived by future generations, and this perspective can help us understand how to improve our behavior. While it is easy to be appalled by what has been done in the past, it is more challenging to uncover the actions taking place today that will be regarded with horror in the future and, most important by far, to be a part of the necessary change. By modifying one's perspective to include a more empathetic view of other animals, our obligations and potential become clear.

Keywords: animal rights, animal experimentation, non-animal testing, alternatives, empathy, fear, ethics, ICCVAM, EPA

1 What kind of animal are we?

When I was in school, we didn't consider the rights or protection of animals an issue. I went to a convent in the Himalayas in Kodaikanal, Southern India, and I gassed animals for dissection class without much thought. Looking out at this room, I see some of you are my age or thereabouts and so probably, like me, missed out on the animal rights and ethics courses offered in universities today. In other words, we are late in wrapping our brains around these concepts. The rest of you have no excuse!

Are most human beings inherently kind? Well, I don't have any empirical evidence that many people are unkind, but I do have a lot of anecdotal evidence that some are. And it strikes me that, on the strength of that – as with criminal law, where a collection of circumstantial evidence is allowed to win a conviction – we can pretty much characterize the human race, of which I am a living, breathing part, as being, quite often, "a species behaving badly" – especially when no one is looking.

In Amsterdam, for instance, the honor system of putting out public bicycles had to be dissolved, because so many bicycles have been stolen, repainted, and sold. And then there's the person who stole my wallet. Luckily they left the Euros and only took the dollars, which are almost worthless. But such is human nature that even friendly, well-educated people who should know better often behave badly, and I'm not just talking about hedge fund managers.

In a study of visitors to Antarctica, it was found that people with a university or postgraduate education were significantly more inclined to harass seals and trample plants than those with less formal education. And then there is "pack behavior": Abu Ghraib isn't the only place where people have engaged in conduct that would be seen as truly depraved under normal circum-

stances but which became normal and routine, even amusing, as the "group mind" muted out decency.

So, average people are capable of more cruelty than we'd like to believe, sometimes not even seeing their behavior as wrong. And sometimes this cruelty becomes institutionalized and is not even seen as abnormal any more. That means, of course, that animals, elderly people, children, and any others placed in a vulnerable position or in an institution, such as an orphanage, a nursing home, or a laboratory, are far more likely to be abused.

Let me give some examples of why I say this:

Look at those who not only have taken a solemn vow to be good but who are expected to set the bar on goodness. The church is an institution more respected than any other. Yet the Roman Catholic Church, not only in the US but in Europe, has had to learn that going along with and, indeed, covering up abuse may one day catch up with you. We all know now about the little boys, and some girls, who were too scared to speak up and who had their lives and bodies interfered with. But, if pillars of the community demonstrate no regard for the feelings of children in their care, it would be mad to think that animals in laboratories, who certainly don't enjoy anything like the revered status of children in our society, are being well treated!

And no one is naïve enough to believe that research is somehow an exception – that it is the one place on earth where those in charge can be trusted to police themselves. That's like leaving a 5-year old in charge of a chocolate cake. Which, actually, researchers have done and the results are not surprising!

Let me offer some idea of what our investigators and others have found: When they thought no one who cared was looking: UNC researchers put live animals into the freezer bins and cut off rodents' heads with scissors without any effort to lessen their suffering; researchers who couldn't be bothered to walk to the gas chamber room chose to violate their protocols and kill the animals by breaking their necks against the cage cardholders. Technicians at Huntingdon Life Sciences in the U.K. were filmed punching beagles in the face and simulating sex with each other as they tried to inject a frightened dog at the same time. Researchers at the largest contract laboratory in the US, Covance, were caught slamming petrified monkeys into steel cages or, having tied them down to the table, stuffing bottles in their mouths and mocking them while whooping it up to loud music. Covance's reaction was not to decry the behavior and fling the abusers and their supervisors out the door, but to sue us – both in the UK and in the US – to get us to stop showing those videos. They lost their cases.

These are dismaying examples of people shamelessly abusing their power over others in their care and then trying to cover it up. And it's not just at Covance. Every single time we go into an institution undercover – including at AAALAC-accredited institutions – we come out with footage of atrocities.

Animals, like institutionalized people, are often simply forgotten; they become wallpaper. Once, I was touring the National Institute of Health (NIH) model facility in Poolesville, Maryland – a "model" facility, mind you – when I found some baboons being kept in small metal isolation cages. I found them because I was walking along one of the corridors and heard a fearful banging noise. That made me look through the little window into their room to see a huge male smashing his head into the stainless steel back wall of his cage. He was a Hamadryus baboon: huge, with a big snout like a dog, and such a colorful coat, sticking out all over, that he looked like a man who had been plucked off the street on his way to a fancy dress ball. I enquired about these baboons, each sitting in a standard, small metal isolation cage in this stark, barren room, with nothing to do or see or touch, no contact with each other, unable to even walk two steps. And I discovered that the researcher who had been using them in a cancer study had accepted another job two years earlier and had moved way! He was living his life in a new town, driving around, shopping, watching TV, talking to his kids, and the baboons were sitting there, day in and day out. They had been plucked from their homes, troupes, and families in Africa seven years earlier, shipped to Russia and then to the US on what could only have been frightening journeys for them, and then locked in see-through boxes in a room, and left to stare at the walls for seven years. Once a day, men in masks entered and hosed down the room and put food in their metal bowls.

Except for the absence of water boarding, they might as well have been at Guantanamo Bay.

Who was to blame for this casual oversight that caused these bright animals so much misery? The grant provider? The researcher who left? The technicians? The IACUC? All of the above, surely? Not one of them had thought the baboons important enough to wonder about, even those who had seen these animals every day. This kind of appalling neglect, in which a living being is left to experience needless mental strain, goes on all the time. So, we must be vigilant, each of us in a position to do so, to spot it and stop it. Otherwise, what kind of animal are we?

People in institutions often get used to things that disturb other people. They are like the lighthouse keeper who was so accustomed to hearing a gun go off under his nose, every six minutes, every night, to warn ships at sea that he slept through it. One night, the mechanism failed and the gun didn't go off. The lighthouse keeper woke with a start, sat up in bed and said, "What's that?"

We need to be vigilant to what's happening to others around us, because those over whom we have control can only depend on us to notice their circumstances. We cannot allow animals to become the wallpaper we don't see any more.

Those of you here who are striving to get animals out of research or at least treated with some understanding of *who* they are and what they need, have to deal with those who can't empathize (neurophysiologists can debate whether that means that their "mirror neurons" are underdeveloped or not), and with others who think it's perfectly fine to wait until some unspecified time in the future to make changes or that change is just not a priority. And we all have to deal with people on IACUCs who are untrained, don't appreciate the important responsibility vested in them, or who get too busy – personally or professionally – to take a proper look when a protocol is presented that could result in animals being put through painful, uncomfort-

able, and worrisome procedures that should have been modified or rejected outright. That is unconscionable.

2 Today's conduct as *future* "past" horrific behavior

It is said that the only thing we learn from history is that we *don't* learn from history. Let me examine with you how history applies to our behavior toward animals today.

One of the most studied cases in research ethics is, as you know, the Tuskegee experiment, in which poor black men in the southern US were purposely *not* told by their doctors that they had syphilis and were simply used as research subjects. This is a good example because it involves a marginalized group – individuals with no recourse, who didn't seem to count for much, and whom the dominant group did not understand or consider important. And because the researchers may indeed have had some sort of good intentions but didn't realize that, at some later point they'd be condemned for such a lack of empathy.

But it's not only the Tuskegee men. Human orphans were used in tuberculin tests and trials of low levels of radiation; poor Irish women immigrants to the US were used in gynecological practice surgeries which, when perfected, were performed on the rich, paying classes. Even human neonates were operated on without anesthesia until quite recently.

In the US, enlisted men (GIs) were used in LSD experiments without their knowledge; some thought they were going mad and killed themselves. GIs were used because they are a pool of often low-income men considered cheap, disposable, and replaceable. Sound familiar?

It is easy to be appalled by what has been done in the past, but callous behavior seen as acceptable just yesterday is now seen for what it is: ignorant and wrong.

For truly uncomfortable reading, there is "The Nazi Doctors," by Robert J. Lifton. In his study of what made doctors able to live with themselves while doing to prisoners exactly the sorts of ghastly things that are done to animals in today's laboratories, including drowning experiments and teratogenicity experiments, he discovers that the comfortable idea that they were just "some madmen" ruled by another madman is absolutely unsupported. In fact, vast numbers of people somehow rationalized these extreme cruelties and killings in a variety of ways and not only accepted it but participated in it.

Lifton's conclusion is that these experimenters were ordinary people like us. One of them, Dr. Siegmund Rascher, even felt comfortable enough to put on paper a formal request to Berlin to move his experiments on inmates from Dachau to Auschwitz because, he wrote: "In Auschwitz, the freezing process is faster because it is colder there. Moreover, the camp is bigger, so that the subjects' howling can hardly be heard."

The analogies are clear. Nobel Laureate Isaac Bashevis Singer, whose family fled the Nazis, became a vegetarian because he looked out of his window above a slaughterhouse in Chicago, watched the cattle shackled together, being prodded and poked down the ramp to their deaths and wrote, "To animals, all men are Nazis."

Lucy Kaplan Rosen, who wrote the introduction to *Eternal Treblinka*, tells the story of her father, who was transported in a cattle car to Birkenau-Auschwitz in 1944. This was after he had witnessed the murder of his wife and two daughters. He survived six camps. Ms. Rosen says that what she loved most about her father was that when, in 1945, at only 100 lbs and bearing the injuries of years of Nazi abuse, he emerged with his previous compassion for animals enhanced precisely because he realized that he had been treated like one.

So, we see that past atrocities weren't necessarily an aberration, and ordinary people can do hideous things if they don't discipline themselves not to or are not stopped by someone else.

The trick, however, is not in simply looking backwards but in figuring out what is being done *today* that will be looked back upon in the future with disgust. The trick is to be one of the people who finds a way to reach those who can't or won't *relate* to the being on their *hotplate*.

Lifton's study is also instructive in another way. It suggests that – if the same observations of human nature apply today – of every three people hired as animal caretakers, two will not object, even silently, when an animal is abused or neglected, and one of the three will be easily capable of joining in flagrant abuse. Let's ask ourselves then, of every ten people appointed to an IACUC, how many will speak up when an unnecessary or unnecessarily cruel experiment is proposed?

On the encouraging side, we teach our children that "Might Does Not Make Right," and we ask them to obey the Golden Rule of "Do Unto Others as You Would Have Them Do Unto You." Presumably, we mean those things. But when you think about it, only "might" allows someone to pick up a rat, a mammal every bit as sentient as any dog or cat or me, and bleed him from the eye and then toss him back into a shoebox-sized container, as if the rat's experience, his fears and sensations aren't real. And only a lack of empathy would allow someone to joke that now the animal will need a white stick when he goes out on the town.

We all know that scientists want to be thought of as precise and particular. Perhaps you know the joke about two scientists driving along in Australia when they pass a flock of sheep. "Look," said one, "Those sheep have all been sheared." "Well," says the other scientist, "On *one* side of their bodies, anyway." Like most stereotypes, I'm not sure this business about particularity is true, and I'll tell you why:

A few years back, we ran a photo in PETA's *Animal Times* magazine that showed rats stuffed into narrow plastic inhalation tubes like so many socks. You can imagine how frightening it must be to have giant aliens, whose intentions are not benign, stuff you into a tube so tightly that your nose and ears are squashed against the sides. The picture was from an article published in the journal of the National Institute of Environmental Health Sciences, and the "cute" caption under the original photo read: "Nosing Around." It reminded me of a caption I saw on a photograph of a group of black males, unable to find work in a South African township, with the caption, "Lounging Around." The article in the NIEHS magazine described an experiment in

which rats were forced to breathe mercury vapors in an attempt to *duplicate* the reproductive effects *already seen for years* in dental hygienists exposed to mercury in fillings. Incidentally, the Principal Investigator concluded: "We weren't able to reproduce any of those effects in our animal model."

When we showed this photo to the head of ICCVAM, he shook his head and said, "that's terrible." But it turns out he didn't mean that the experiment was terrible or what was done to the rats was terrible; he meant that it was terrible that the photo had been put in the magazine for everyone to see!

PETA researchers attend many toxicology conferences. They hear the jokes, the references to boondoggles, the acknowledgements that certain experiments have no value whatsoever. On a tour of the USUHS, the US Uniformed Health Services facility, the chief veterinarian in charge pointed to the desert tortoises being used to study TMJ (temporo mandibular joint pain). Now, the desert tortoise's jaw is not like a human jaw at all; it is especially arranged for a fibrous, plant-based diet – there are no teeth! People chew their food, tortoises do not, and there are many other critical differences. Furthermore, the NIH says that TMJ treatment should not be surgical, rather it begins with simply changing jaw movement and avoiding teeth-grinding. Considering that a desert tortoise cannot be reasonably expected to speak our language, and we don't understand hers, it will be difficult to explain this to her and to know when she feels better.

PETA did a double take and got the funding on that experiment pulled. If we hadn't, who knows how many other tortoises would have had their jaws damaged and how many more tax dollars would have been wasted.

I was once invited to lecture at the USUHS and made the mistake – or clever move – of arriving early enough to sit unnoticed in the back before the earlier session ended. The laboratory chief told his students: "When you fill out the government form as to why you chose to use rats or mice, do not write 'because they are cheap, easy to handle, and few people care about them.' The form is supposed to show there is a good science-y sounding reason for your choice." But the truth is, there is no "science-y" reason.

At a conference on neurotoxicity, a panelist discussing the EPA's developmental neurotoxicity test (which uses at least 1,300 animals every time it is conducted) joked that the "FOB" – which is the acronym for the "functional observation battery" that is used in neurotoxicity testing – really stands for, and I quote: "functional observation bullshit." In that government scientist's words, "we do it because the EPA tells us to," *regardless of relevance*. In my words, "*They* die for *our* sins."

The EPA official on the panel acknowledged: "We know the rat isn't the right model. But it's like being in a bad marriage – you know you should get out but you don't because there's so much history there." That would be funny, but it isn't if you are the subject of a painful experiment.

So, the gig is up on pretending that results from one species apply across the board to others. I've always said that "When it comes to feelings, like hunger, pain, and thirst, a rat is a pig is a dog is a boy." That's just a plain old fact. What *isn't* a fact is that when it comes to *physiology*, a rat is a pig is a dog is a boy. Meaning that it's time for the so-called "gold standard" of animal testing to be recognized as the lump of coal it is.

One more example of how casually the animals' suffering is viewed: Several years ago, at the National Academy of Sciences' Institute for Laboratory Animal Research workshop on federal reporting requirements for pain and distress in animals used in laboratories, panelists who included animal researchers from the US Department of Agriculture (USDA), the NIH, and various universities and professional associations, were often seen yukking it up over animal suffering. Empathy was in short supply, if there at all. One of the panelists spoke about the importance of proper training. Reciting a "can-ya-believe-this-one" story, he gave a litany of botched jobs by under-trained or under-skilled lab staff, culminating in a story of someone improperly restraining a mouse so that, as the animal's skull was being drilled into, his body spun around with the drill-bit. Throughout the list of examples, other members of the panel chuckled in recognition, and at this last example, several panelists laughed openly.

I read a story in the newspaper about a group of young people who had been caught standing around a BBQ pit, poking a live kitten into the burning coals and laughing as the kitten cried. Who wouldn't be upset at the mental health of these individuals? But, when you think about it, the place is unimportant. Whether suffering is seen as a joke in the lab, at a conference, or at the barbeque pit, it must stop.

Cruelty doesn't have to mean being the person who pokes the kitten deeper into the coals, or being the person who wields the scalpel or syringe. It also means being the person who ignores the plight of the pain-wracked or lonely animal in the cage. Where there is acceptance of the perpetuation of wrong, there is *complicity*. Perhaps the Nazi doctors could not have spoken up without being shot, but we are not in that position – not any of us.

What allows this kind of jocular, cruel nonsense to go on is a lack of empathy (that undeveloped "mirror neuron") or the group acceptance of unacceptable behavior. It will continue until every one of us who cares – and we are many – gets a moral backbone, gets truly interested in real science, and speaks up, complains, refuses to accept this kind of thinking, and puts an end to it.

It's history again, isn't it? Someone gave me a book of parlor games from the Southern US. One game in it is described this way: "A distinguished lady is chosen to address the group. She must explain that a baby has been orphaned, and she will suggest that the club rise to the occasion by chipping in to raise the child for its first year. Of course, everyone will agree, and one member must be asked to volunteer to be the first to take the child home. As the volunteer comes forward, everyone claps with appreciation. Have the maid bring a swaddled baby into the room. When the lucky volunteer is handed the infant, all will howl with laughter when the cloth is pulled back and the lady finds she is holding a Negro child." The book continues: "If a Negro child is not readily available, you can achieve the same effect by using a baby pig."

We find that beyond horrific now, but it was all good fun down South in that bastion of civil liberties, the United States of America, not so long ago. And it is a great illustration of why change must keep come right on coming.

3 The capacity for pleasure and fear

I'm not so much condemning the people who didn't "get it" then and who don't "get it" now; rather, I'm offering it as a blunt reminder to ourselves that we are making history every day that we live and breathe. Today's conduct can be seen as *future* "past" horrific behavior, if you know what I mean. It was just 30 or so years ago, when I was already smoking cigarettes, that physicians in white coats appeared on TV, advising us to smoke low tar cigarettes to soothe a sore throat. It was just 30 years ago that, as Dr. Jane Goodall points out, scientists openly ridiculed the idea that chimpanzees were intelligent, had social needs, engaged in tool making and use, and had language. Or more recently, take the octopuses commonly used in laboratories:

It is only after years of electro-shock experiments on them that experimenters have conceded that these bizarre – to us – animals are so emotionally upset by their loss of control over their destiny and their inability to flee the pain meted out to them, that they commit suicide by pecking themselves to death. Jacques Cousteau first revealed how dolphins in captivity would sometimes take their own lives; now we see that behavior in other captive species, including cephalopods. Yet these animals have been treated in laboratories as if they were inanimate. Or actually, that's not so, because if they were thought to be inanimate, no one could design an experiment to hurt them and see what they would do, could they? So, people recognize that octopuses have feelings but are deliberately ignoring the fact.

Recently, an extensive study of pleasure in the animal kingdom showed that "from tickling to playing catch, animals engage in certain behaviors just for fun, even enjoying sensations that are unknown to humans." The author of the findings, published in *Applied Animal Behavior Science*, believes scientists, conservationists, and others should not overlook animal joy. "The capacity for pleasure," the author writes, "means that an *animal's life* has intrinsic value, that is, value to the individual independent of his or her value to anyone else, including humans."[1] And what isn't mentioned here, is that one of the most overlooked areas of animal suffering is fear. We talk about caging size and other considerations, but the "Fear Factor" isn't just an American TV game show.

This animal at this podium knows a bit about fear. My father was a very daring man, quite an adventurer who went out in fierce storms and into war zones and was at Bikini Atoll to help set up nuclear testing. He once took my mother in a jeep across the Little Rani of Kutch in India. This is an area that, at certain times of the year, is extremely dangerous, pitted with pockets of quicksand that are impossible to see but which can gobble you up. My mother only found that out, and found out that no one else had had dared accompany my father on this expedition, when they were well out in the middle of it. She remembers yelling at him and my father saying, "Oh come on, where's your spirit!" To which she replied, "At home in the drinks cabinet, but damn you, if I'd known that we might die, I'd have brought some with me."

I know what she meant, because when I was a tiny tot, he took me up a glacier many thousands of feet high. Despite the physical pain of the cold, which was intense (we didn't have polar fleece back then), what was far worse was the fear: the fear of tumbling thousands of feet down the side of that vast wall of ice, the fear that something terrible was about to happen to me. That is the experience animals in laboratories live with all the time: The mother monkey clutching her baby to her chest in her small metal box, wondering if that giant, powerful animal entering the room to take her blood is also going to steal her child. And he is, just as if he were taking a box off a shelf, ignoring her fear grimace, her desperately chattering teeth and her little begging sounds. Yet her love for her child is indisputably as strong as any human mother's love for her infant. What must her suffering be like when she cannot protect her own child? Who will speak up for her?

In a BBC documentary, scientists refer to cuttlefish as "aliens from inner space." It's a fascinating term, because, of course, our species is out there in space, spending a great deal of time, money, and effort searching for intelligent life, yet it is all around us: from those cuttlefish who communicate in waves of color (able to create a magnificent pattern on one side of his body that lures a prospective mate while creating another pattern on the other side that wards off a competitor); to the smallest desert mouse who rolls a stone in front of her burrow to collect dew; to the Indigo buntings who navigate by learning the constellations, fix their position by the height of the sun and, if blown off course, reset their paths by the phases of the moon and the rising and setting of the stars; to the rhino who communicates by altering his breathing.

And should any Cartesians be among us and laugh such things off as programming, they must also laugh off their own loves, desires, and fears, their own programmed behaviors. The Cartesians also would have to ignore studies like the one this very month reporting how five crows all were able to figure out how to use a short stick to get hold of a medium one and the medium one to reach a large one – the only one that could allow them to retrieve a food reward. All five figured it out without training, and four of the birds did it in the first try.

4 A slippery slope is just another term for progress

So, what if we do come across intelligent life during our space explorations? If it is stronger than we are, we will undoubtedly beg for mercy and understanding and insist that "We come in Peace." But if we find intelligent life out there that is *not* as strong

[1] "Animals just want to have fun, survey finds: From tickling to playing catch, animals do some things simply for enjoyment," Jennifer Viegas, Discovery Channel.

as we are, what will happen to those noble protestations? Our governments will want to do to *them* what we've done to all the intelligent life forms on *this* planet: Capture them, cage them, dissect them, and deny them any consideration. Perhaps snack on a few of them just as we snack on the sea slug who, like a fat opera diva, so gracefully dances among the rocks, her cloak floating behind her, her mind on who knows what. But we don't have to be like that computer Pacman, gobbling up everything in our path; we can be considerate of those with less power than we have.

I hear people who want to cling to the *status quo* say, "Don't concede that primates need social enrichment or that rats and mice must be afforded protection, *it's a slippery slope*." And of course it is, but isn't a "slippery slope" that leads us away from treating others badly just another term for progress? A society can't evolve if it is afraid of the slippery slope. Looking back, it was a blink ago in time that a noted Harvard surgeon was deeply worried about the "absurd" idea that a woman might be allowed into the operating room, let alone be trained as a physician. In the time of the Suffragettes it was said that "If you give *women* the right to vote, you might as well give asses the right to vote." Today, we have fine female physicians and scientists and – as for women voting – well, we should be entitled to make the same mistakes at the polling booth as any man! And frankly, looking at who we elect sometimes, one wonders if asses couldn't do a better job than the lot of us.

So, what can we do? You may think that I want all animals out of the laboratories now and you'd be right. I do think it is morally indefensible – given what we know in this day and age – for us to inflict pain and suffering and fear on any other living being *simply because we can*. That is the lesson I take from history. And if you don't "get it," look at the animals, learn about the animals, and if you still don't "get it," look again.

However, one can still help enormously without having to embrace that belief.

5 Regulatory testing: the obstacles and the movement forward

I polled the PETA staff who work on these issues, asking them to give me basic starting points. I am only focusing here on regulatory and toxicity testing. Let me go through the list and I hope you will agree:

1. Where there is a non-animal alternative, use it.
2. Don't automatically default to animal testing: profiling a chemical's biological activity using a suite of non-animal methods will allow for thoughtful toxicology by identifying the most hazardous chemicals and providing information that can guide further testing and, of course, NO experiment using animals should EVER be carried out if the information is available or can be derived elsewhere.
3. Where there isn't an alternative yet, work to find one and keep an eye on what's going to be available soon.
   Alternatives for many biological endpoints are in sight, and for the more complex endpoints in the areas of toxicokinetics, chronic toxicity, and carcinogenicity, there are efforts that urgently need promoting, including PETA's planned workshop focusing on replacing the rodent cancer bioassay.
4. The EPA and NIH are beginning to act on the vision set forth in 2007 by the National Academy of Sciences, which recognizes that the near-exclusive reliance on animal testing that has characterized chemical testing programs to date are costly, time-consuming, and not up to the task of accurately and adequately assessing the toxicity of tens of thousands of chemicals. Both the NAS report and the 2009 EPA strategic plan recommend moving away from dependence on animal tests to a process that relies more heavily on *in vitro* assays to predict human health effects.
   EPA and NIH have created formal collaborations to develop and implement this approach, such as the ToxCast and Tox 21 initiatives.
   It is encouraging to note that federal agencies and some industry consortia are stepping up to the plate to help fund these initiatives, in addition to PETA which, up till now has tried to step into the void by providing massive donations to *in vitro* laboratories and QSAR experts, even though our budget is mere manicure money to federal agencies such as the EPA.
   May I solicit your support to encourage and fund the use of these technologies in current and future testing programs? There is still so much work to be done as current testing programs have been slow to incorporate this new approach.
   For example, there is the hideous mess that is the EPA's Endocrine Disruptor Screening Program. The first phase could kill more than 40,000 animals – and not one of them will die quickly or painlessly – to test just 67 chemicals, all of which are either pesticides or High Production Volume chemicals that have *already been heavily tested*. The additional testing is highly unlikely to provide any useful information for additional regulation of these chemicals, especially when the EPA still can't say how it is going to use the resulting information! This program is so out of tune with where we are now that it is like designing an iPhone app using carrier pigeons. This program needs to be redesigned from the bottom up to take advantage of the latest technology and new approaches. Please take a look at our poster, number 539, that describes an integrated approach to endocrine testing and see how you can use it.
   The US National Toxicology Program continues to kill thousands of animals every year to test well characterized chemicals, even natural substances such as ginseng and green tea, and it does so if the substance is nominated by anyone, even by a single anonymous person.
   And what of ICCVAM? In a decade, ICCVAM has gone from beloved baby to Frankenstein monster. Instead of doing the job the US Congress intended it to do – namely to facilitate the incorporation of non-animal methods into government regulatory programs – it has become the chief obstacle in the US to the use of non-animal testing methods.
   Recently, ICCVAM rejected the work of a consortium of companies that worked together to develop a non-animal method for assessing eye irritation. The work was so promising that the EPA launched its own pilot program, accepting

data using this method. So, now the EPA is making more progress than the federal entity whose job it is to do so.

Another shameful example is that nine years ago, an international workshop concluded that *in vitro* cytotoxicity could be used *immediately* as a dose setting measure to reduce the number of animals poisoned in lethal dose tests. The experts also concluded that, with interest and funds, the test could be validated as a complete replacement method for lethal dose tests within 2-3 years. Yet it took until 2008 for ICCVAM to issue formal recommendations to agencies to use the cytotox method and then *only as a reduction method to set the starting dose for poisoning animals*.

NIEHS has not made appropriate funding of ICCVAM a priority, and ICCVAM does not do the sort of independent research performed by ECVAM and ZEBET. It appears that ICCVAM members, who are drawn from the federal agencies, are being allowed to misuse ICCVAM to perpetuate their antiquated biases in favor of animal tests. This is evident through ICCVAM's continued presumption that animal tests are the "gold standard" of toxicology to which all non-animal assays must measure up (or down as the case may be) and through emails leaked to PETA in which ICCVAM representatives discuss circling the wagons against evidence-based toxicology. It should therefore come as no surprise that the US lags so far behind Europe and some other countries, not only in real football but in implementing non-animal testing methods too. We documented that disgrace in an extensive report last year that ended up on the front pages of the *Washington Post*.

5. The current US toxic chemicals legislation, the Toxic Substances Control Act, is about to be revised. While this is sure to lead to requirements for information similar to REACH in the EU, it also provides an unprecedented opportunity to incorporate these new approaches and new technology into toxics legislation – something we at PETA are working hard to do and which you will hear about in the next session on chemicals and pesticides.
6. In Europe, there is much work to be done as well. In spite of the deadlines imposed by the Cosmetics Directive to eliminate animal testing and a number of EU initiatives, there are still no accepted alternatives for eye irritation nor any completely non-animal methods for acute toxicity, and non-animal replacements for chronic and developmental toxicity will not be in place for the 2013 deadline. REACH will have an enormous impact on the number of animals used in testing. There are some animal reduction provisions in REACH, but with thousands of chemicals requiring base data sets by 2010, there will be animal suffering on an unprecedented scale.

   Even without REACH yet in full swing, the numbers of animals used in the UK has risen every year since 2000, with a dramatic increase of 42% in the past decade. Most of this is due to the increasing use of transgenic mice – with thousands of them being used to breed and maintain each line – even though these so-called "models" of human disease are of questionable relevance. And in spite of public support for a ban on primate experiments, the use of primates in the UK actually rose 16% in 2008 over 2007. Surely, intelligent and dedicated people can find a way to reverse this trend.

   While the continuing revision of Directive 86/609 provides an opportunity to address long-standing problems with the use of animals in European experiments, there is a worrisome de-regulatory agenda that threatens to gut the proposal of its most progressive measures. Europe must not miss the opportunity to put in place a rigorous and comprehensive system of regulation that reduces animal suffering; fosters the development, validation, regulatory acceptance, and uptake of replacement methods; and, most importantly, provides the basis on which meaningful progress will be made towards the goal of eliminating all animal experiments as soon as possible.

That is the end of the list.

6 Passing the final exam in the history of mankind

One good thing that came out of the 1960s, and I'm not talking about tie-dye, was the expression, "If you aren't part of the solution, you are part of the problem." Society is deeply indebted to each of you who are part of the solution. Those of you who are in government, it is vital to rock the boat; those of you who are on committees and in funding agencies who opine that animal tests are not only ethically flawed but often conducted out of habit, obstinacy, laziness, and because no one has spoken up, thank you for living.

Now, let me return to where I started. If we believe what social scientists have told us – that ordinary humans are capable of extraordinary cruelty – we must recognize that laboratories are exactly the place where such things will occur. It is not enough to regulate vigorously, although we must do at least that. As a group, human beings are far too tolerant of cruelty and far too unpredictable to be entrusted with the lives of truly vulnerable beings, yet we are in that position and must be vigilant about our conduct.

And finally, if you are thinking, "What was that bit about Brueghel's Two Monkeys?" – it refers to the poem by Wislawa Szymborska, who wrote:

This is what I see in my dreams about final exams;
Two monkeys, chained to the floor, sit on the windowsill,
The sky behind them flutters,
The sea is taking a bath.
The Exam is the History of Mankind
I stammer and hedge.
One monkey stares and listens with mocking disdain,
The other seems to be dreaming away –
But when it's clear *I don't know what to say*
He prompts me with a gentle
Clinking of his chain.

The animals are all around us, intelligence and emotions shining from their eyes. They are prompting us – with the clinking of their chains – to lead a life we will be proud of when the

time comes, as it always does to mortal beings, to take that final exam.

No one knows when that will be, but a time *will* come to all of us – those of us who answer to a god and those of us who answer to ourselves – when we look back on our careers and our lives with pride or with regret.

Society's ethical values expand as we come to understand that we not only have the capacity but are duty-bound to extend consideration beyond just ourselves, our families, our races, and, without a doubt, our species. To understand, as Dr. Albert Schweitzer said, "Ethics are complete, profound and alive only when addressed to all living beings."

To put it in practical perspective, I was in England recently and, as I was reading the Sunday paper, I came across a column written about dogs. The columnist wrote: "Contrary to what Buddhists would have you believe, remember, a dog is just a dog: he will never write a great book or compose a great symphony." I thought "Hang on a minute!" I'm going to bet that this columnist will never write a great book or compose a great symphony, and one thing I know is that he will never detect a cancerous tumor with his nose, and he certainly wouldn't be able to find his way home over hundreds of miles without the benefit of a GPS, a map, a street sign or advice from another human being. Perhaps what separates humans from other animals is the desperate quest that our species has to find something that distinguishes us from the other animals.

Maybe the question should be "When will we all start seeing ourselves as just one of the many musicians in this vast orchestra of life, one no more special than the others?"

When we take those final exams, may we all be able write that we contributed to the History of Mankind by bravely confronting our biases and by helping our species evolve from undisciplined bully to compassionate citizen. May we be able to say that each of us had the nerve, the backbone, the principle, and the vision to say what needs to be said about the use of animals, the suffering of animals, and the appropriateness of the behavior of those around us. I wish you all the best in everything you do to pass that exam.

Correspondence to
Ingrid E. Newkirk, President
People for the Ethical Treatment of Animals
501 Front Street
Norfolk, VA 23510
USA
e-mail: jessicas@peta.org

The Principles of Humane Experimental Technique: Timeless Insights and Unheeded Warnings

Michael Balls
FRAME, Nottingham, UK

Summary
In The Principles of Humane Experimental Technique*, Russell and Burch said that "the central problem is that of determining what is and what is not humane, and how humanity can be promoted without prejudice to scientific and medical aims". They then explained how the Three Rs can be used to diminish or remove direct inhumanity ("the infliction of distress as an unavoidable consequence of the procedure employed") and contingent inhumanity ("the infliction of distress as an incidental and inadvertent by-product of the use of a procedure"). They concluded that "Replacement is always a satisfactory answer, but Reduction and Refinement should, whenever possible, be used in combination".*
Many of the commonsense insights in The Principles *are no less relevant today than they were in 1959. However, their warnings about the limited value of models and, in particular, the danger of succumbing to the high-fidelity fallacy (whereby it is assumed that the best models for humans are always placental mammals, because they are more like humans than other animals), appear to have largely gone unheeded. Of particular importance is their discussion on toxicity testing, which they saw as one use of laboratory animals "which is an urgent humanitarian problem, for it regularly involves considerable and sometimes acute distress". How, then, can it be that mammalian models are still routinely used in attempts to detect chemical carcinogens and reproductive toxins, despite the fact that the relevance to humans of the data they provide has not been, and perhaps could never be, satisfactorily established? Nevertheless, there are signs that some significant changes in attitude are taking place, particularly in the USA, which could be more in line with the main thrust of The Principles, the belief that good science and human technique inextricably go hand-in-hand.*

Keywords: animal experimentation, reduction, refinement, replacement, Russell & Burch, Three Rs, toxicity testing

1 Introduction

On 21 August 2007, in Tokyo, Japan, during the Opening Ceremony of the *6th World Congress on Alternatives and Animal Use in the Life Sciences*, we paid tribute to the rich and varied life of W. M. S. Russell, who, with R. L. Burch, gave us the Three Rs concept, in their book, *The Principles of Humane Experimental Technique* (Russell and Burch, 1959; Balls, 2008). Now, two years later, in Rome, Italy, toward the end of the *7th Congress,* we are celebrating the 50th anniversary of the publication of *The Principles*.

I firmly believe that *The Principles* contains timeless insights into how we should think about the use of laboratory animals for research and testing, which are as relevant today as they were in 1959, and which can guide us as we seek to achieve genuine progress, whilst maintaining the highest standards in terms of both scientific methodology and animal welfare. The book also contains warnings about how fundamental mistakes can be made, which compromise the value of the science and threaten the welfare of the animals.

My concern is that, although a large number of people say they are committed to supporting the Three Rs concept

of *Reduction*, *Refinement* and *Replacement*, as put forward by Russell and Burch, most of them are unaware of the detailed implications of these insights and warnings, because they have not read the book itself. The result is that I am disappointed that the great benefits afforded by a careful consideration and dedicated application of *The Principles* have not been achieved. I therefore hope that this Congress will mark a new beginning – a much-needed, renaissance of the Three Rs.

As one of the initiatives to celebrate its own 40th anniversary, FRAME has made an abridged version of *The Principles* available, with the cooperation and support of Cleo Paskal, W. M. S. Russell's Literary Executor. The principal aim of *The Three Rs and the Humanity Criterion* (Balls, 2009) is "to seek to retain the remarkable concepts and flavour of the original, whilst clarifying some of the English language employed, as well as reducing some of the lengthy discussions based on uses of animals in the 1950s which are no longer practised".

What I plan to do here, is to list some of the insights, then say why the failure to heed some of the warnings severely limits the impact of the Three Rs, and, as a result, compromises the high standards of scientific practice and animal welfare which Russell and Burch sought. Quotations from *The Principles* are shown in italics in the sections 2 to 4.

2 The insights

The main principle on which Russell and Burch based their analysis is that *it is widely recognised that the humanest possible treatment of experimental animals, far from being an obstacle, is actually a prerequisite for successful animal experiments*. They considered the central problem to be that of *determining what is and what is not humane, and how humanity can be promoted without prejudice to scientific and medical aims*.

They began with *the concept of inhumanity and its relation to those of pain and distress,* then turned to *the positive aspect – the analysis of methods of diminishing inhumanity in experimentation*.

They said *that we must first distinguish direct and contingent inhumanity. By the former, we mean the infliction of distress as an unavoidable consequence of the procedure employed. By the latter, we mean the infliction of distress as an incidental and inadvertent by-product of the use of the procedure, which is not necessary for its success*.

Their thesis was that *inhumanity can be, and is being, diminished or removed under the three broad headings of Replacement, Reduction, and Refinement – the Three Rs of humane technique:*

- *Reduction means reduction in the numbers of animals used to obtain information of a given amount and precision*.
- *Refinement means any decrease in the incidence or severity of inhumane procedures*.
- *Replacement means the substitution for conscious living higher animals of insentient material. It is always a satisfactory answer, but reduction and refinement should, wherever possible, be used in combination*.

3 The warnings

The Principles also contains a number of important warnings, but it will only be possible to discuss two of them here.

First, the tendency to misunderstand the nature of models, and especially, the use of animals as models for man: *A perfect model of the human organism would obviously be indistinguishable by any test from its original. Any other* in vivo *model must depart in some degree from the original. There are two factors governing the way in which the model differs from the original. Fidelity means overall proportionate difference, and discrimination means the extent to which the model reproduces one particular property of the original*.

The point is that, however great the overall similarity between the original and a model may be, if there are significant differences in the specific properties being studied, the model will not be useful. Also, however, great the differences between the original and a model may be, if there are sufficient similarities in the specific properties being studied, the model may be a useful one. Clearly, high fidelity/high discrimination models are most useful, but, where this is not possible, a low fidelity/high discrimination model is preferable to a high fidelity/low discrimination one.

Russell and Burch go on to say that *Progress in replacement has been restricted by certain plausible, but untenable assumptions* about models, which have led to *the high-fidelity fallacy*. The major premise is that *the highest possible fidelity is always desired in medical research and testing*, and that, *for man*, a *member of another placental mammal species would be a model of higher fidelity than a bird or a microbe*. This assumption can have disastrous consequences in terms of the data produced, and can also lead to unnecessary, and therefore unacceptable, animal suffering.

Most of the macaques used in the UK are involved in toxicity testing for the pharmaceutical industry. When asked about the 16% increase in 2008, a senior scientist from Global R&D of a leading pharmaceutical company said that this was driven by a move toward more biological medicines (Gill, 2009): "These treatments need to be tested in a human-like model, and old world primates are closer relatives of humans than new world primates." But what about the TGN 412 scandal, where the "human-like model" did not reveal the acute adverse effects of this humanised product, which later occurred in human volunteers? Macaques should not be used merely because of their overall similarity to humans, but only when it has been established, in advance, that they are appropriate models for use in a particular study.

Russell and Burch were concerned about toxicity testing on more-general grounds, since they considered it to be *an urgent humanitarian problem, for it regularly involves considerable and sometimes acute distress, and to be an activity where the high-fidelity fallacy may be more prevalent and influential at the legal level, rather than at the laboratory level*. They clearly foresaw the problem of persuading regulators to accept the use of scientifically-advanced, replacement alternative methods instead of the animal tests with which they are more comfortable.

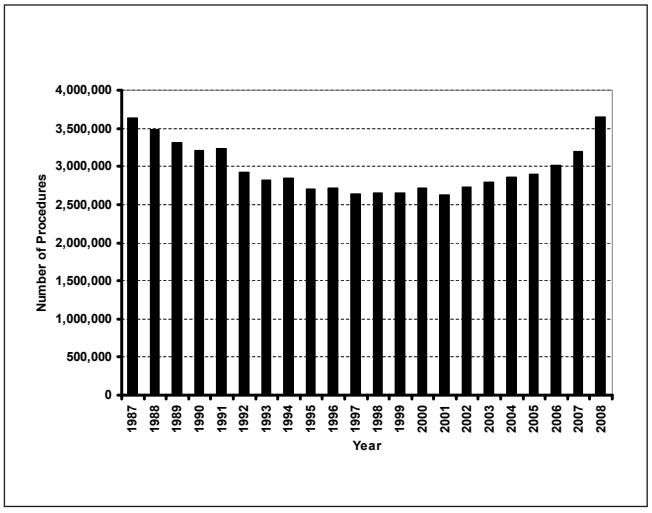


Fig. 1: Numbers of scientific procedures on living animals in Great Britain, 1987 to 2008*.

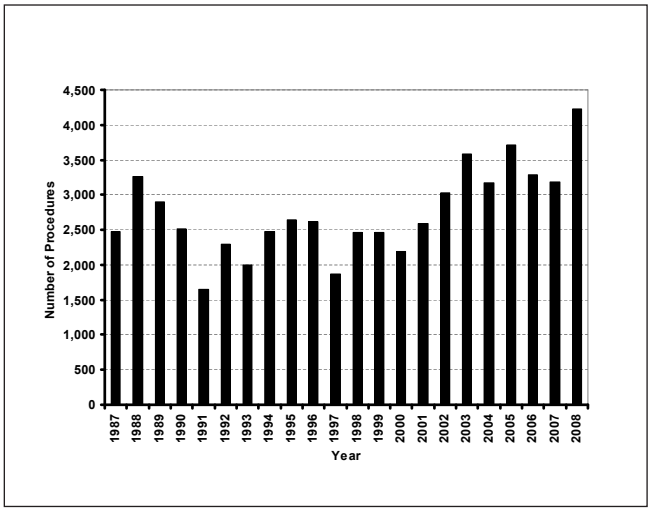


Fig. 2: Numbers of scientific procedures on macaques in Great Britain, 1987 to 2008*.

4 Progress of the Three Rs

4.1 Reduction

At the 5th World Congress, in Berlin in 2005, I said that "the progressive *reduction* in the numbers of animal experiments which had been foreseen when the new legislation was passed in the 1980s seems to have come to an end, especially as more and more mice are sacrificed on the altar of genetic exploitation. Also, far from working together toward the zero option of the use of non-human primates, there is pressure to build more and more primate research centres (Balls, 2006)." Sadly, the situation has worsened since 2005, rather than improved.

In Britain, the number of scientific procedures on living animals in 2008 was higher than the number in 1987 (Anon, 2009), the first year after the *Animals (Scientific Procedures) Act 1986* came into force (Fig. 1), and the number of procedures applied to old world primates (macaques) in 2008 (4230) was much higher than in 1987 (2470; Fig. 2).

The situation in Britain is no doubt mirrored in other countries, and the production and use of genetically modified mice is widely used to excuse the overall situation. The chilling prospect is that there is much worse to come, given a recent report in *Nature* (Abbot, 2009): "European investment could see knockout rats catching up with mutant mice in medical research. The European Commission has approved the world's first major systems-biology programme to study the rat. Known as EURATRANS – for European large-scale functional genomics in the rat for translational research – the multimillion-€ project includes collaborators in the United States and Japan." Will this be followed by a move to set up programmes for producing large numbers of transgenic non-human primates?

One of the main scientific points of emphasis in *The Principles*, the need for high quality experimental design and statistical analysis, has been largely ignored. Indeed, there is much evidence to support the contention that scientists, regulators, universities, industries, governments and grant-giving bodies are content to tolerate bad science. There are some praiseworthy efforts to redress this situation, such as the training schools run by the FRAME Reduction Steering Committee, in collaboration with the University of Manchester, and with the support of the European Commission's COST programme (Howard et al., 2009).

Virtually no time has been specifically devoted to *Reduction* at this 7th World Congress, so it is impossible to avoid the conclusion that it is the forgotten R, even though Russell and Burch saw it as *of great importance, and of all the modes of progress, it is the one most obviously, immediately, and universally advantageous in terms of efficiency*.

4.2 Refinement

There has been considerable progress concerning the husbandry and use of laboratory animals, not least because of greater recognition of the importance of laboratory animal technicians and laboratory animal veterinarians.

That is to be welcomed, but there is also a danger that refinement can be used as a convenient way of showing commitment to the Three Rs, whilst ensuring that animal experimentation is seen as respectable and can be allowed to continue, while the fundamental ethical questions raised by it are avoided.

This is not my area of expertise, so I will not dwell on it further. However, I do wonder whether the activities linked to ethical review processes and institutional animal care and use committees, however positive they may be in terms of refinement, have any significant effects in relation to reduction and replacement. We should remember that Russell and Burch said that, *in general, refinement is never enough, and we should always seek further for reduction and if possible replacement*.

* Figures 1 and 2 were kindly provided by Michelle Hudson, and are based on the annual statistics of animal procedures regulated under the terms of the *Animals (Scientific Procedures) Act 1986*, published each year by The Stationery Office, London.

4.3 Replacement

The position adopted by academia and research-funding bodies has long been that new techniques emerge during the natural development of a science, so deliberately seeking replacement alternatives for the vast array of procedures applied to laboratory animals in the basic sciences is not necessary. However, it could be argued that the legislation which regulates animal experimentation imposes legal and ethical obligations on all concerned, which should not be so easily avoided. From the scientific point of view, the high-fidelity fallacy deserves far greater recognition and resultant action, especially in the case of animal models of human disease, where insufficient about the disease is known for sound judgements to be made about the relevance, or otherwise, of the model.

Nevertheless, it is toxicity testing where, as Russell and Burch recognised, the greatest concerns arise. For example: Why is it believed that the rodent bioassay can tell us what chemicals are likely to be carcinogenic in humans, when dosing is based on the maximum tolerated dose, and the mouse is a poor model for the rat and vice versa? Why is it believed that the current regulatory reproductive toxicity tests can tell us what chemicals are likely to be reprotoxic in humans, when so many false positives and false negatives occur that it is impossible to judge whether the test procedures can even identity reprotoxins in the animal models themselves? Why are animal data still widely regarded as the "gold standard" to be matched by non-animal tests, when the reliability and relevance of the tests concerned cannot be established, even for the animals? The Draize rabbit eye irritancy test data are so variable that the test cannot reliably be used to identify potential eye irritants in rabbits, so why is it believed that the data have any relevance to humans?

5 Moving backward in Europe

Despite much positive talk by politicians and senior officials about the importance of the Three Rs and their commitment to them, it could be argued that, in actual fact, Europe is going backwards. The great promises of the 1990s have not been delivered. Three examples will suffice, although this is not the place to discuss them in detail.

1. The REACH system: totally unworkable, proposed by ill-informed ambitious civil servants, taken up by ill-informed ambitious politicians, and then by ill-informed ambitious governments. It was clear from the early drafts of the Commission's White Paper that nobody had any coherent or defensible idea of the numbers of chemicals that would need to be registered, the number of additional animal tests that would be required, or how human health and the environment would be afforded greater protection. There was no mention of non-animal tests or their validation. Later on the potential value of replacement alternatives was grudgingly accepted, and, as it became clear that the numbers and costs of various aspects would be much, much higher than had been expected, they came to be seen as a way of saving face and reducing embarrassment. We now have an expensive agency in Helsinki, which is producing thousand upon thousand of "guidance" documents. What is the value of accumulating so-called "missing" data, if its value and genuine usefulness have not been established? There is a likelihood that, since the science is being driven by the politics, the validation process itself will be corrupted. Instead of waiting until they have been independently shown to be reliable and relevant for their stated purposes, replacement alternative tests may come to be accepted because they are "suitable" (i.e. politically convenient). But aren't plausibility and suitability, based on the high-fidelity fallacy, among the reasons why we have so many useless animal tests? What will be the consequences, if "suitable" tests are found, in time, to have been "unsuitable" after all, and who will accept the responsibility for their failure?

2. The 7th Amendment to the Cosmetics Directive: a ruse of no value, seemingly designed to convince politicians and a gullible public that something is being done. The situation with regard to cosmetic ingredients is no less unsatisfactory. Many of the chemicals used in cosmetics are also used for other purposes, and the REACH system will apply to them. If the testing of cosmetic ingredients in compliance with the Cosmetics Directive comes to be banned, which companies will admit to doing any animal testing for that purpose? Won't they say that the testing was done for compliance with the REACH system, and won't some of them try to stick to "not tested in animals for cosmetics purposes" labelling, while conveniently and dishonestly omitting the last three words? In addition, the definition of a "cosmetic" used in Europe is increasingly unsatisfactory, as cosmetic products are produced which actively alter the biological properties of the components of the skin.

3. Draft proposals for a Directive to replace Directive 86/609/EEC: one of the worst pieces of draft legislation ever published, which even foresees circumstances in which Member States could permit experiments on Great Apes. The Commission's proposals were produced after years of discussion with all kinds of stakeholders, but what emerged was not a draft directive at all. Rather than having its intentions spelled out clearly, in a way which could be implemented as a law, the result was a curate's egg mishmash of ideas which were either ill-conceived or in need of further discussion and development. As a result, hundreds of amendments have been put forward, and the result is a threat to both good science and sound animal welfare. Despite this totally unsatisfactory situation, there is great political pressure to get something in the statute book.

I mention these three points, because I fear that they illustrate the fact that Europe is going down a slippery slope as far as the Three Rs and a sensible balance between science and animal welfare are concerned. In the long run, fine words and catchy slogans count for nothing – it is sustainable actions of high quality that matter. As Jesus said, "Ye shall know them by their fruits" (Matthew's Gospel, 1611 translation). The problem is, who has the power and the desire to intervene and see that the downward trend is reversed?

6 Moving forward in the USA?

Meanwhile in the USA, a number of very promising developments are taking place, and in particular, the follow up to the publication by the US National Academy of Sciences of *Toxicity Testing in the 21st Century – A Vision and a Strategy* (National Research Council, 2007) and by the US Food and Drug Administration of the *Critical Path Initiative* (FDA, 2004). What these two documents have in common is the recognition that animal models can no longer be relied on in drug development and in toxicity testing in general, and that more effort should be put into the development, evaluation, acceptance and use of what we would describe as replacement alternative methods and strategies, particularly when they are of direct relevance to humans.

To be fair, I must recognise that promising developments are also taking place in Europe in relation to speed and safety in drug discovery, as was shown, for example, at a symposium held in London in 2008 (Gard and Clotworthy, 2009).

7 Concluding remarks

The published proceedings of the 7th World Congress will reveal a wealth of activity, mainly focused, perhaps, on the possibility of replacing animal procedures by more-modern methods, based on the remarkable progress being made in cellular and molecular biology. In a way, then, the Three Rs concept is to the fore as we celebrate the 50th anniversary of the publication of *The Principles*.

Nevertheless, as I have pointed out, there are grave causes for concern, especially as the number of animal procedures conducted each year continues to increase, and legislative changes, especially in Europe, threaten to perpetuate and expand that increase even more.

I hope that many of the grandiose statements made in apparent support of good and ethical science based on the Three Rs will lead to identifiable and excellent outcomes, which will demonstrate a genuine renaissance in line with Russell and Burch's outstanding concept.

In particular: significant *reduction* in animal use should be achieved, without further delay, through better experimental design and statistical analysis; *refinement*, however welcome, should not be seen as an end in itself; and much greater resources should be invested in the dedicated search for *replacement* alternatives. Meanwhile, Russell and Burch's warning about the high fidelity fallacy should be taken much more seriously and acted upon.

The way in which Russell and Burch put it cannot be repeated too often: *If we are to use a criterion for choosing experiments, that of humanity is the best we could possibly invent. The greatest scientific experiments have always been the most humane and attractive, conveying that sense of beauty and elegance which is the essence of science at its most successful.*

So, let us all take this opportunity to renew our commitment to live up to this ideal, with total sincerity, then go home, and get on with the job.

References

Abbot, A. (2009). Return of the rat. *Naturenews 460*, 788. Published online on 11 August 2009.

Anon (2009). *Statistics of Scientific Procedures on Living Animals*, Great Britain 2008. HC 800. London: The Stationery Office.

Balls, M. (2009). *The Three Rs and the Humanity Criterion*. Nottingham: FRAME.

Balls, M. (2008). Professor William Russell (1925-2006): Doyen of the Three Rs. *AATEX 14, Spec. Issue*, 1-7.

Balls, M. (2006). The Three Rs: Looking back … and forward. *ALTEX 23, Spec. Issue*, 29-32.

FDA (2004). *Challenge and Opportunity on the Critical Path to New Medicinal Products*. Available at: http://www.fda.gov/oc/initiatives/critical path/whitepaper.pdf. (Accessed 15.04.05). Silver Spring, MD: US Department of Health and Human Services, Food and DRUG Administration.

Gard, T. and Clotworthy, M., eds (2009). Proceedings of the Speed and Safety in Drug Discovery Symposium. *ATLA 37, Suppl. 1*, 1-65.

Gill, V. (2009). Upward jump in lab animal tests. *BBC Online*, 21 July 2009. Available at: http://news.bbc.co.uk/1/hi/sci/tech/8161253.stm (Accessed 21.07.09).

Howard, B., Hudson, M. and Preziosi, R. (2009). More is less: reducing animal use by raising awareness of the principles of efficient study design and analysis. *ATLA 37*, 33-38.

National Research Council (2007). *Toxicity testing in the 21st century: A vision and a strategy*. Washington, DC: National Academy Press.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen.

St Matthew (1611). Jesus teaches with authority. Chapter 7, verse 16. New Testament, Authorised King James Version. London: Church of England.

Correspondence to

Prof. Michael Balls
FRAME
Russell & Burch House
96–98 North Sherwood Street
Nottingham NG1 4EE, UK
e-mail: michael.balls@btopenworld.com

The Principles of Humane Experimental Technique: Is It Relevant Today?

Alan M. Goldberg

CAAT, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Summary

In the section on replacement in their 1959 book, The Principles of Humane Experimental Technique, Bill Russell and Rex Burch state: "As new fields of biology open in the future, it may become a matter of routine to apply the lessons of the past and turn as soon as possible to the techniques of replacement." They foresaw in vitro techniques, in their infancy at that time, as the science of the future. Today, in the US, the National Academy of Sciences publication of Toxicity Testing in the 21st Century: A Vision and a Strategy proves their point. This pivotal publication recognizes that the future of toxicity testing lies in the use of human cells in culture and in methods that Bill Russell and Rex Burch could not have possibly conceived of in 1959 but which they identified generically as "the future."

To truly establish this 21st Century approach will require very specific training in translational toxicology (the use of clinical observations to develop in vitro methods to understand pathways and systems biology), the development of transnational programs, and ways to evaluate the accuracy, validity and importance of new and/or traditional studies. These evaluations are known at evidence-based toxicology (EBT).

Science is "the art of the question." The concepts identified above are the tools to answer these questions – and to lead us to the next round of questions. The principles that Bill Russell and Rex Burch developed during the 1954-59 writing of "The Book" may be more important today than ever before. They argued that the newest science, the most humane science is also the very best science. This hypothesis is being proven now, as each of us contributes to the world's body of knowledge.

1 Introduction

On this the 50th anniversary of Bill Russell and Rex Burch's book, *The Principles of Humane Experimental Technique (The Principles),* I share the honor, with Michael Balls (FRAME), of presenting my thoughts on the relevance of this landmark publication to the conduct of science today. In brief, these *Principles* are more important than ever.

I had the pleasure of meeting Bill Russell and his wife Claire in their home in Reading, England in 1992. I met Rex Burch a few days later in Sheringham, England. With each of them, I shared stories, books, news clippings, and other publications illustrating the profound impact of their book, *The Principles of Humane Experimental Technique*, over the 30-plus years since its publication.

I visited Bill and Rex shortly after I was awarded the Russell and Burch Prize by the HSUS in 1991. At this point, Bill and Rex had not seen each other since the publication of their book. The two scientists went their separate ways, and neither had any idea of the impact of their work until Martin Stephens of the HSUS called them to get approval to use their names on the award.

Many of us subsequently had the pleasure of getting to know Bill, and those interactions have been amply recorded in two sets of publications – both of which are readily accessible in ATLA. The first is a series of obituaries of Bill, collected shortly after he died (Balls et al., 2006). The second is a collection of personal reflections given as part of FRAME's 40th Anniversary celebration (2009, in press).

Rex Burch, on the other hand, is less well known. Unfortunately, few of us had the opportunity to spend much time with him. When I met Rex, he was the sole owner and sole employee of a small microbiology testing laboratory in the basement of the Town Hall of Sheringham, England. Since it was a one-person operation, Rex worked 7 days a week, 52 weeks a year. He was an exceedingly gracious host.

Rex's contribution to *The Principles of Humane Experimental Technique* lay primarily in conducting interviews and collecting data. We have little knowledge of his role in the planning of the book, organization of material, or actual writing of the manuscript. Perhaps, as the Russell archive is evaluated, we may learn more.

Michael and I each will share how we see the impact of The Principles. I have chosen to address the relevance of Bill and Rex's thinking and why their book is a monumental contribution that anticipates where we are today.

To answer the question, is *The Principles* relevant today, I will focus on four areas of activity essential to the future of *in vitro* sciences as they apply to Toxicology and Risk Assessment. These areas are: 1) training of scientists; 2) The NAS report: Toxicity Testing in the 21st Century; 3) the Transatlantic Think Tank for Toxicology (t^4): working together; and 4) humane sciences: the art of the question.

2 Training of scientists

The principles of humane experimental technique reside in two concepts: humane science is the best science, and to practice humane science one must incorporate the 3Rs: reduction, replacement, and refinement. I list them alphabetically, rather than in the order suggested by *The Principles*, to emphasize that all Three Rs are critically important to good science (Goldberg, 2004).

Most scientists have not been taught the 3Rs or the principles one must follow to practice humane science. For most, humane science is the untaught standard that the scientist strives for. Bill Russell and Rex Burch provided the tools; now we must make sure these tools are made available to all scientists. We also must make sure that we teach our students that failure to practice the most humane science can and will compromise the quality of their results.

I have the privilege of co-teaching a course titled "Animals in Research: Science, Policy and Law." It is offered twice per year – once as a classroom course and once as a distance education course with electronic communication and live talks. In the past two years, about 50 students have taken the course. These students get it. Their questions, their discussions, and their defense of proper use of animals in research all tell me that animal research will only be accepted if humane science is practiced.

3 Toxicity Testing in the 21st Century: a Vision and a Strategy

The US Environmental Protection Agency (EPA) contracted with the National Academy of Sciences (NAS) to address the question of what toxicology testing will look like in the 21st century. The resultant final publication is remarkable (Committee on Toxicity Testing and Assessment of Environmental Agents, NAS, 2007). It truly looked into the future and suggested a clear and uncompromising path to fulfill the vision. The publication has been widely recognized prior to this meeting, and at this meeting several awards have been presented to the Academy and the Authors.

The report has four basic conclusions:

- Animal research is expensive
- Animal research is not always predictive of human consequences
- The future of toxicology lies in the use of human cell systems in culture
- To understand toxicity we must understand pathways (mechanisms) and systems biology

These are the same concepts that the Johns Hopkins Center for Alternatives to Animal Testing (CAAT) has focused on since our founding in 1981. From the outset, we funded research on mechanisms and human cells in culture. I wish I could take credit for all of it, but the Advisory Board of CAAT was more than a co-equal partner.

Two additional aspects of a 21st Century toxicology warrant consideration here: translational toxicology and evidenced-based toxicology (EBT). The NAS report does not directly address these topics, but both are important as we complete the paradigm shift in toxicology.

Translational toxicology involves, first, using observations in clinical or toxicological events and designing *in vitro* and/or mechanistic approaches to understand the observation; then, using the newly developed *in vitro* methods to evaluate different chemicals with these methods; and finally, developing approaches to predict *in vivo* consequences from *in vitro* data. Clear examples of this approach have been developing over the last few years (Sawada et al., 2005).

Evidence-based toxicology is an extension of evidenced-based medicine. It is a structured approach to literature evaluation. Evidenced-based approaches serve to eliminate studies that do not meet the criteria for inclusion in summary evaluations and to strengthen those studies that do meet the criteria. Studies or methods that do not have scientific validity or scientific rigor are eliminated. In clinical medicine this allows one to identify "best" treatment options. In toxicology it will allow better risk assessment and management of chemicals. One example of an evidenced based approach is the Cochrane Collaboration (The Cochrane Collaboration, founded 1993). The Cochrane Collaboration provides a wonderful example of the depth of information that can be achieved with evidence-based approaches.

The Principles anticipated what has become the toxicology of the 21st century. It states, "As new fields of biology open in the future, it may become a matter of routine to apply the lessons of the past and turn as soon as possible to the techniques of replacement" (p.104). Bill and Rex recognized that *in vitro* techniques, in their infancy at that time, would become the science of the future.

4 t4 – Transatlantic Think Tank for Toxicology

In the early 90s, when I was a Trustee of the Doerenkamp-Zbinden Foundation, I pushed the foundation to establish endowed chairs in *in vitro* toxicology to guarantee the future of the field and to enhance the impact of the foundation's work. The first chair was awarded to Marcel Leist at the University of Konstanz, Germany. My term as a Trustee with the Foundation was completed, but the idea was in place and well supported. The next chair was established at the University of Utrecht in the person of Bas Blaauboer, who was followed most recently by Thomas Hartung at the Johns Hopkins University.

Prior to his arrival at Hopkins, Thomas had already established the Transatlantic Think Tank for Toxicology (t^4) to serve as an incubator for new ideas and to provide a forum where concepts can be discussed at an international level. Its agenda is to examine ideas and concepts in light of changing science and regulation. Thomas asked me to be a principal member of t^4, along with the Doerenkamp-Zbinden Professors. The initial activities were defined at the time of the 2009 SOT meeting in Baltimore, as Thomas arrived and was installed officially as the Doerenkamp-Zbinden Professor and Endowed Chair for Evidence-based Toxicology.

The first publication identified as coming from the t^4, "Re-evaluation of animal numbers and cost for *in vivo* tests to accom-

plish REACH legislation requirements for chemicals – a report by the Transatlantic Think Tank for Toxicology (t^4)," written by Constanza Rovida and Thomas Hartung, was published in the 3/09 issue of *ALTEX*. A two-page version of the paper appeared in *Nature* (27 August 2009). This paper cogently argues that, since the initial evaluation of cost and numbers of animals required to meet the requirements of REACH, many intervening circumstances have resulted in significantly higher figures on both counts. Rovida and Hartung demonstrate that the actual costs of meeting the REACH requirements will require some 54 million animals and $ 9.5 billion. Both figures are unacceptable and undoable. This makes it all the more imperative that we fulfill the NAS Vision and Strategy.

The t^4 also will participate in the implementation of the NAS report, *Toxicity Testing in the 21st Century*. These related but separate activities form the basis of the paradigm shift in toxicology. *In vitro* toxicology has been established on a solid foundation of activities that span the globe. The t^4 is a major building block of this foundation.

5 Humane science: the art of the question

As we move from animal studies to *in vitro* studies, the goal of humaneness is generally achieved. As Michael Balls points out in his companion piece, *The Principles* require that refinement and reduction be used together.

Science can be defined as the art of the question. The better the question, the better the science. Humane science – the 3Rs in practice – requires a very different set of questions. In the US, the responsibility for posing the appropriate questions is shared between the investigator and the Institutional Animal Care and Use Committee (IACUC). The investigator has the responsibility to ask whether the experiment will answer the question asked, whether the species selected is the best choice, and whether the experimental design, including the number of animals, is planned so as to maximize the benefits of the question. The IACUC, by law, does not look at the science. Its role is to examine whether the above identified parameters are fulfilled. It asks such questions as: Are methods available that would make these experiments more humane? Are non-invasive methods available? Are they being used? Have humane endpoints been considered? Is this the best approach to answer the question being addressed?

In *The Principles*, Bill and Rex hypothesize that humane science is the best science. Over the last 50 years, the scientific community has proven their hypotheses to be true. The title of this talk: *The Principles of Humane Experimental Technique*: Is It Relevant Today?

The answer: A resounding yes.

References

Balls, M., Paskal, C., Kirkwood, J. et al. (2006). An appreciation of the life of Professor William Russell. *ATLA 34*, 469-487.

Committee on Toxicity Testing and Assessment of Environmental Agents, Board on Environmental Studies and Toxicology, Institute for Laboratory Animal Research, National Research Council of the National Academies (2007). *Toxicity testing in the 21st century: a vision and a strategy*. Washington, D.C.: The National Academies Press.

Goldberg, A. M. (2004). The sixth annual FRAME lecture on animals and alternatives: societal expectations and scientific need. *ATLA 32*, 545.

Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature 460*, 1080-1081.

Personal reflections on Bill Russell. FRAME's 40th anniversary celebration (2009), in press.

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals. *ALTEX 26(3),* 187-208.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen & Co. Special edition published by Universities Federation for Animal Welfare (UFAW), 1992.

Sawada, H., Takami, K. and Asahi, S. (2005). A toxicogenomic approach to drug-induced phospholipidosis: Analysis of its induction mechanism and establishment of a novel in vitro screening system. *Toxicol. Sci. 83*, 282-292.

The Cochrane Collaboration: The reliable source of evidence in health care. See: http://www.cochrane.org.

Correspondence to

Alan M. Goldberg, Ph.D.
Bloomberg School of Public Health
Johns Hopkins University
615 N Wolfe Street, Room7032
Baltimore, MD 21205
USA
e-mail: agoldberg@jhsph.edu

PL8

Calling on Science: Making "Alternatives" the New Gold Standard[1]

Melvin E. Andersen

The Hamner Institutes for Health Sciences, Research Triangle Park, NC, USA

Summary

All of life's great journeys start with a goal in mind! The 2007 NAS report, Toxicity Testing in the 21st Century – A Vision and A Strategy, has proposed a clear goal. This report envisions a not-so-distant future where all routine toxicity testing for environmental agents will be conducted in human cells in vitro evaluating perturbations of cellular responses in a suite of toxicity pathway assays. Dose response modeling would utilize computational systems biology models of the circuitry underlying each toxicity pathway; in vitro to in vivo extrapolations would use pharmacokinetic models, ideally physiologically based pharmacokinetic models, to predict human blood and tissue concentrations under specific exposure conditions. Results from these toxicity pathway assays and associated dose response modeling tools rather than those from high dose studies in animals would represent the new gold standard for chemical risk assessment. This talk focuses on some of the scientific challenges required to make this vision a reality, including characteristics of assay design, prospects for mapping and modeling toxicity pathways, assay validation, and biokinetic modeling. All of these tools necessary for this transformation of toxicity testing to an in vitro platform are either available or in advanced development. Science must lead the transformation. The scientific community, animal alternatives groups, regulatory agencies, and funding organizations will also have to muster the resolve to work together to make this vision a reality.

Keywords: gold standard, toxicity pathways, in vitro biology, computational systems biology, toxicity testing transformation

1 Introduction

This year marks the 50th anniversary of the publication of *The Principles of Humane Experimental Technique* by William Russell and Rex Burch. Their contribution focused attention on the 3Rs – replacement, reduction, and refinement. In toxicity testing, the primary initiative with the 3Rs in the intervening decades has arguably focused on reduction of animal usage while holding firm the belief that results from animal studies provide a "Gold Standard" for making decisions about possible human health risks of compounds. The very wording, "alternatives", has often been regarded by many in toxicology as those test methods that will reduce animal usage even though the result from the tests are not necessarily optimal for risk assessment decision-making. A second challenge in reduction of animal use through mechanistically-based testing arises from the idea of validating "alternatives". The process of validation with alternatives, in general, focuses on the ability of a test or a series of tests to give results consistent with those that would be obtained through testing in animals. In this context, all alternatives will fall short of the mark of complete concordance with *in vivo* outcomes. Are all efforts to reduce animal use significantly doomed to disaster as they are dashed against the "gold standard" barrier?

The recommendations of a recent report (NRC, 2007) from the US National Academy of Sciences, *Toxicity Testing in the 21st Century: A Vision and A Strategy* argues that it is time to redefine the toxicity testing paradigm, moving away from high dose studies in animals to *in vitro* assays assessing perturbations of toxicity pathways by environmental agents. In essence, the report supports a sweeping redefinition of our "gold standard." The author of this present paper was a member of the NAS toxicity testing committee. Since the publication of the NAS report in June 2007,

[1] Several of the ideas in the introductory portion of this paper reflect those from two previous contributions (Krewski et al., 2009; Andersen and Krewski, 2009). The interested reader should also consult these two papers.

several of the NAS committee members have presented aspects of the report at more than 40 venues in North America and Europe. These presentations and the lively debate engendered on these occasions have sharpened ideas about the use of results from *in vitro* toxicity pathway assays in risk or safety assessments. The NAS report, although published in 2007, was essentially completed in fall 2006. Advances in several key technologies in the past three years – especially stem cell biology, computational systems biology, and pathway mapping and modeling – appear likely to be key catalysts for moving the vision forward. Finally, the transformation from current, traditional approaches to new *in vitro* methods based on human biology will not come easily. Who will step up to assist in the transformation to a new approach to testing and risk assessment? Several initiatives within the United States, both in federal government research organizations and in the private sector, look likely to accelerate implementation. These topics – (1) the recommendations from the NAS report, (2) the manner in which the *in vitro* toxicity pathway data can be organized for risk/safety assessments, (3) the call to the alternatives community to embrace 21st century computational and bioinformatics methodologies in designing and interpreting *in vitro* results, and (4) the institutional opportunities to accelerate implementation of the NAS vision – are discussed in turn in this current paper.

2 Toxicity Testing in the 21st Century: a Vision and a Strategy

The US Environmental Protection Agency and the US National Institute of Environmental Health Sciences asked the US National Research Council (NRC) to provide guidance on new directions in toxicity testing, incorporating emerging technologies such as genomics and computational systems biology into a new vision for toxicity testing. In 2004, the NRC convened a 22 person committee for this purpose (Tab. 1). The committee produced two reports. The committee's interim report (NRC, 2006) provided an overview of testing methods and approaches that could incrementally improve traditional toxicity testing. This report noted that health protection agencies and the public had experienced increasing frustration with the failure of current approaches to toxicity testing to provide timely, relevant information to support informed regulation of environmental agents. These toxicity testing strategies relied primarily on the observation of adverse health responses in laboratory animals treated with high doses of these agents. Estimating risks to human populations based on high dose animal studies require difficult extrapolations, first from high doses to environmental levels that are usually orders-of-magnitude lower than those used in the animal studies, and then from animals to humans. These traditional toxicity testing approaches and methods for their interpretation date back some 30 to 60 years, and were developed at a time when knowledge of biology – and of the manner in which chemical exposures perturbed biological processes – was primitive. While there have been steady, incremental improvements in toxicity testing over the years, there has been no comprehensive evaluation of the manner in which advances in cellular and molecular biology might improve toxicity testing practices.

The final report of the toxicity testing committee (NRC, 2007) outlined design criteria that needed to be considered in any revisions of practices for toxicity testing. In choosing among various toxicity testing options, four criteria are important: (1) achieving broad coverage of chemicals, chemical mixtures, outcomes, and life stages, (2) reducing the cost and time required for toxicity testing, (3) developing a better scientific basis for assessing human health effects of environmental chemicals, including knowledge of modes of action, and (4) minimizing use of animals in testing. The consideration of how these criteria should guide a modern approach to toxicity testing led the committee to propose a new framework for toxicity testing that would entail a major overhaul of current practice.

Toxicity testing and targeted in life studies

The NAS committee vision consisted of several key technology areas (Fig. 1). While also including *in silico* methods for assessing structure activity relationships and population assessments, the transformative parts of their new toxicity testing paradigm was the types of toxicity testing and the manner in which results from these tests could be organized to support human health risk assessment. This vision centers on defining dose-response relationships for toxicity pathway perturbations that would be

Tab. 1: The Roster of the NRC Toxicity Testing Committee

Daniel Krewski (*Chair*), University of Ottawa, Ottawa, ON
Daniel Acosta, Jr., University of Cincinnati, Cincinnati, OH
Melvin Andersen, The Hamner Institutes for Health Sciences, Research Triangle Park, NC
Henry Anderson, Wisconsin Division of Public Health, Madison, WI
John Bailar III, University of Chicago, Chicago, IL
Kim Boekelheide, Brown University, Providence, RI
Robert Brent, Thomas Jefferson University, Wilmington, DE
Gail Charnley, HealthRisk Strategies, Washington, DC
Vivian Cheung, University of Pennsylvania, Philadelphia, PA
Sidney Green, Howard University, Washington, DC
Karl Kelsey, Harvard University, Boston, MA
Nancy Kerkvliet, Oregon State University, Corvallis, OR
Abby Li, Exponent, Inc., San Francisco, CA
Lawrence McCray, Massachusetts Institute of Technology, Cambridge MA
Otto Meyer, Danish Institute for Food and Veterinary Research, Søborg, Denmark
D. Reid Patterson, Reid Patterson Consulting, Inc., Grayslake, IL
William Pennie, Pfizer, Inc., Groton, CT
Robert Scala, Exxon Biomedical Sciences (Ret.), Tucson, AZ
Gina Solomon, Natural Resources Defense Council, San Francisco, CA
Martin Stephens, The Humane Society of the United States, Washington, DC
James Yager, Jr., Johns Hopkins University, Baltimore, MD
Lauren Zeise, California Environmental Protection Agency, Oakland, CA

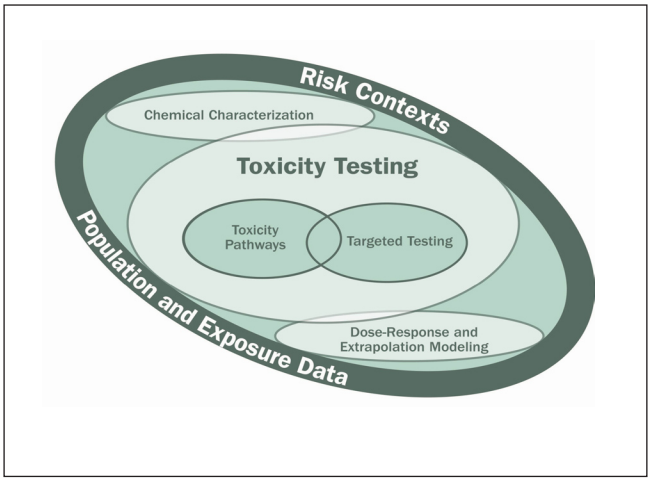


Fig. 1: Components of the vision for Toxicity Testing in the 21st Century (NRC, 2007).
The key elements in this proposal are related to Toxicity Testing, which includes the types of *in vitro* tests and short term *in vivo* tests to evaluate perturbations on toxicity pathways, and Dose-Response and Extrapolation Modeling, which provides the requisite tools for interpreting toxicity testing results for assessing human health risk assessment. Reproduced from the NRC report (NRC, 2007) with permission.

expected to lead to adverse health outcomes if the perturbations were maintained *in vivo* at a sufficient level of intensity and for a sufficient duration of exposure. The key component of the vision is assaying perturbations of toxicity pathways, which are simply normal biological signalling pathways that may be perturbed by chemical exposures. Toxicity pathway testing would require a suite of *in vitro* tests that could identify the range of significant perturbations of human pathways that might occur as a result of chemical exposure (Fig. 2). Biologic responses are viewed as results of an intersection of exposure and biologic function. The intersection results in perturbation of biologic pathways. The circuitry affected by the chemical is expected to determine shapes of dose response relationships for these perturbations. Ideally, these assays would be conducted in human cells, cell lines or in engineered human tissues. The committee believed that the use of a comprehensive array of *in vitro* tests with human cells would markedly reduce the need for whole animal testing, and provide much stronger, mechanistically-based tools for human health safety assessment. It was recognized that the conversion to an *in vitro* basis had challenges and the committee also suggested that targeted *in vitro* testing was also likely to continue for some time where such studies could provide information about metabolism, possible metabolite toxicity, toxicity pathways, etc. Metabolism has been recognized as a particular challenge for developing *in vitro* testing alternatives (Coecke et al., 2006).

Dose-response and extrapolation modeling

How will results from a comprehensive suite of toxicity pathways inform quantitative risk/safety assessments for environmental agents? In this new toxicity testing strategy, *in vitro* concentration response curve would cover multiple orders of magnitude (Inglese et al., 2006, 2007) and evaluate responses in cells/tissues from humans, the species of primary interest. The broad range of concentrations permit the definition of dose ranges resulting, or not resulting, in significant alterations of normal biological function. While low dose and interspecies extrapolations are not as problematic, new challenges arise in understanding the mechanistic bases for dose-response behaviors of the toxicity pathway assays, in calibrating expected blood/tissue concentrations in humans against the vitro concentrations used in the toxicity pathway assays, and in understanding the linkages of early perturbations to adverse responses expected in exposed people. The report identified key technologies that will assist dose response and *in vitro-in vivo* extrapolations, includ-

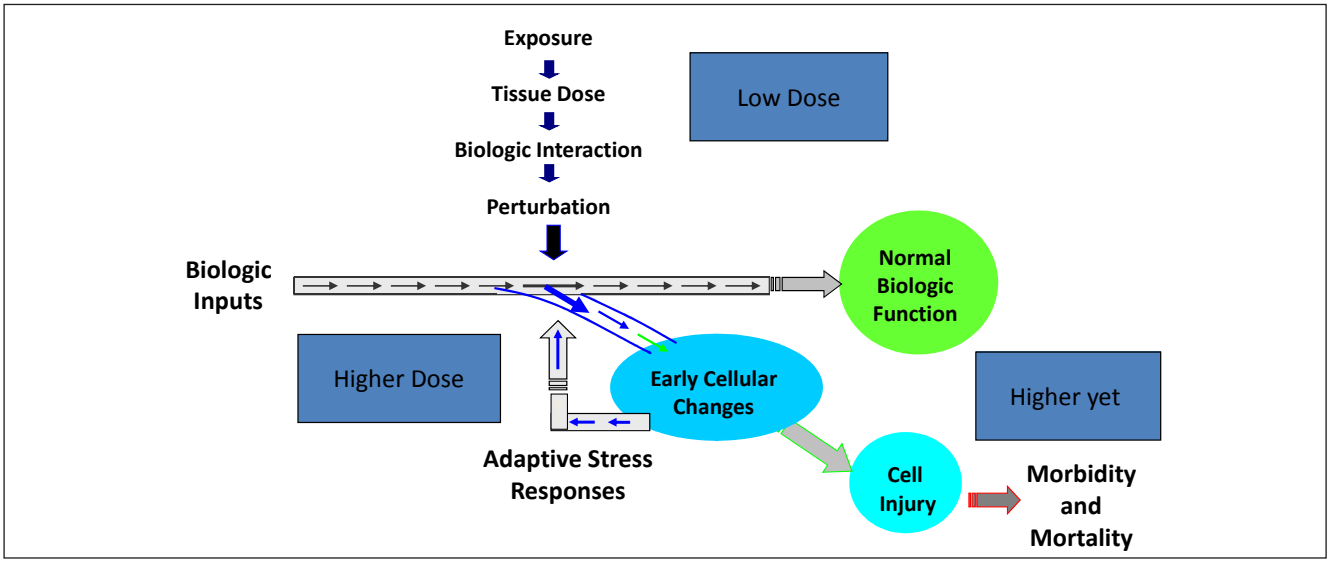


Fig. 2: The progressive activation of toxicity pathways from perturbation of initial targets, through activation of stress controlling pathways, to overtly toxic responses (apical endpoints).
Adapted from Andersen et al. (2005).

ing (1) empirical dose-response models based on results from the *in vitro*, mechanistically based toxicity pathway assays, (2) mechanistic dose-response models based on knowledge of toxicity pathway circuitry and dynamics of pathway function, and (3) physiologically based pharmacokinetic (PBPK) models to equate tissue-media concentrations with tissue dose in exposed people. Two recent perspectives on the NAS report provide good overviews of the report and directions for implementation (Krewski et al., 2009; Andersen and Krewski, 2009).

3 The new gold standard in practice

Over the past three years, there has been continuing discussions about the NAS report with diverse stakeholder audiences. During these discussions, many questions were directed at the manner in which the pieces of the new test paradigm would integrate together to provide quantitative approaches for risk or safety assessment. The NAS report did outline two hypothetical cases of "assessments" that might arise from a battery of *in vitro* test using examples of a reactive gas and of a compound with estrogenic activity. These examples were cursorily developed, but indicated how various parts of the testing and analysis would likely contribute to health assessments. It is possible today to provide a more complete picture of how these pieces might be integrated (Fig. 3).

The core component of the testing will be the suite of toxicity pathway assays (Fig. 3; section i). These assays would be developed for human cells, human cells in culture, or human three-dimensional tissue surrogates. The toxicity test assays themselves need to be capable of evaluating the progression from initial activation of the pathway on through degrees of perturbation that would be considered sufficiently large to be associated with likely toxicity if maintained over a period of time in an intact organism. For most, if not all assays, concentrations are expected to range from sub-threshold through those causing initial pathway activation, on to regions of adaptation, and finally to those causing adverse cellular consequences. To cover these various degrees of response, each assay would likely provide different levels of biological readout as a function of concentration and duration of treatment.

Each pathway assay is expected to have specific dose response characteristics depending on the organization of the circuitry that determines the action of compounds on the toxicity pathway. The dose response behaviors should arise from the underlying biology of the circuitry. These core signaling processes include the initial signal recognition and then the larger scale network through which the initial perturbation progresses to generate toxicity in the test system. Computational systems biology (Alon, 2006, 2007) provides the tools for describing these circuits and the differential behavior of the circuits with increasing degrees of perturbation.

The process of validating toxicity pathway assays would be to study its behavior for positive control compounds and to extract the network structure and network dynamics that determine dose response. The sequential passage from sub-threshold, to adaptive, to toxic conditions represents dose-dependent transitions in modes of action in an *in vitro* system. Dose-dependent

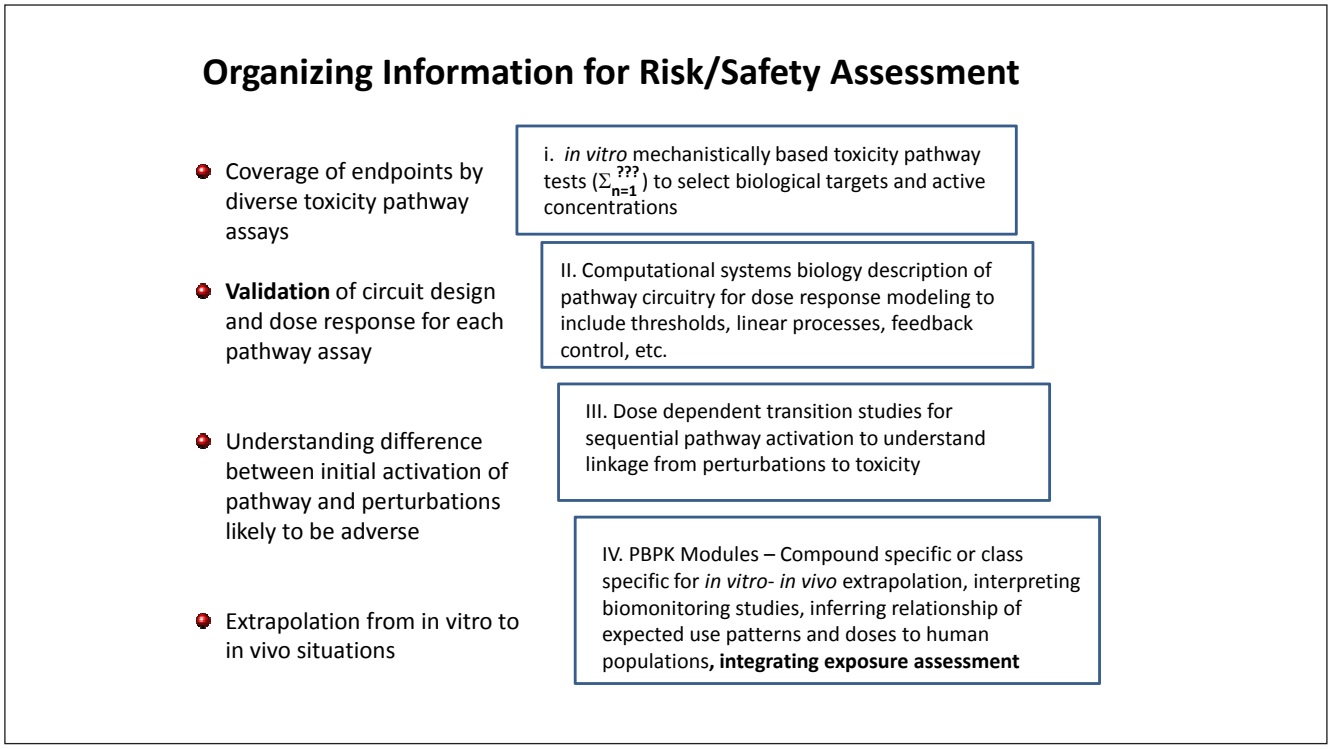


Fig. 3: The progressive activation of toxicity pathways from perturbation of initial targets, through activation of stress controlling pathways, to overtly toxic responses (apical endpoints).
Adapted from Andersen et al. (2005).

transitions in *in vivo* toxicology studies are well known (Slikker et al., 2004a, 2004b). This type of response cascade has been described *in vitro* by Xiao et al. (2003) and by Nel et al. (2006) in work on hierarchical oxidative stress. Feedback process control and dose response have also been examined more theoretically for anti-stress gene regulatory networks (Zhang and Andersen, 2007). Alon (2007) has provided a good overview of network motifs in signaling pathways and in what were termed "developmental" networks. In practice, toxicity pathway characterization would optimally include standard operating procedures (SOPs) for preparing cells, conducting specific assays, generating read-outs, and the detailed process by which the pathway structure, circuit, and dynamics had been evaluated to support dose-response modeling. The detailed pathway characterization (essentially the process of validation of the pathway behavior) would be the mainstay of dose response analyses. For the safety assessment, primary attention would focus on pathways affected at the lowest concentration (Fig. 3; sections ii and iii).

How do we relate concentrations affecting cells *in vitro* with exposures in human populations likely to cause similar responses in an intact individual? Human biomonitoring for chemicals in blood and excreta is becoming more widespread. In some instances, concentrations of exogenous compounds in humans may be available. Comparisons could be made between those concentrations seen in exposed populations and those affecting cells in the toxicity pathways assays in order to estimate a "margin of safety" or "margin of exposure." This comparison is unlikely to be possible with very many compounds. A more general methodology would be development of biokinetic models (DeJongh et al., 1999) to determine the human exposure situations expected to give cell and tissue concentrations similar to those affecting the human cells in the *in vitro* pathway assay test (Fig. 3; iv). These approaches are extensions of the physiologically based pharmacokinetic types of models that have been of interest both with toxic substances and pharmaceuticals (Reddy et al., 2005; Bouvier d'Yvoire et al., 2007). A coordinated effort is required to develop a larger suite of PBPK models and to enhance efforts in reverse dosimetry, i.e., estimating the exposure levels in a human population that produce specific blood/tissue concentrations (Clewell et al., 2008). Current efforts to improve dosimetry methods are also advancing *in vitro-in vivo* extrapolation tools (Gulden and Seibert, 2003; Heringa et al., 2004).

The risk assessment process would entail running the suite of assays for a compound to see the pattern of activation of pathways and the concentrations at which effects were noted in various pathway assays. The most sensitive hits from the suite of assays would then be organized to support both dose response modeling and *in vitro-in vivo* extrapolation. The pattern of activity across the suite of assays could also provide signatures to indicate the types of toxic endpoints that might be observed *in vivo* (Dix et al., 2007). For example, specific signatures might indicate a high likelihood of reproductive toxicity or of hepatic toxicity in a qualitative manner.

Risk assessments completed based on results from these toxicity pathway assays are likely to be quite different from those arising from current approaches. Today, we see effects in animals, usually at fairly high doses, and estimate the likely incidence of response at lower doses in exposed populations. For cancer, we might try to estimate the expected concentration estimated to give a 1/1,000,000 level of population risk. This process has two less than desirable outcomes – first, labeling compounds based on high dose hazard studies and (2) providing a false sense of precision regarding our ability to extrapolate across doses and species. For instance, if Compound A causes cancer, at a maximally tolerated dose, it becomes labeled as carcinogen regardless of considerations of exposure levels. Secondly, the public is led to believe that the estimates of the low dose extrapolations are scientifically valid without any appreciation of the uncertainties about these estimates. In contrast, the assessments based on the *in vitro* toxicity pathway assays would be more directed at safety assessment, estimating regions of exposure where no appreciable perturbations are expected in human cells or human tissues in culture.

4 Calling on 21st century science

The NAS committee discussed a variety of key technology areas for toxicity testing in the 21st century. While the broad suite of new tools are likely to influence many areas of toxicology research and to greatly improve understanding of cell signaling pathways, it is important to ask more narrowly how specific technologies and advances will contribute to the four components noted in Fig. 3. The three areas most likely to benefit immediately are in assay design, using stem cell technology, pathway mapping and modeling, and computational systems biology for assessing expected dose response behaviors.

Assay design

A frequently voiced concern after publication of the report was the difficulty in obtaining and working with primary human cells and the caveats associated with use of human cell lines. The past few years have provided optimism in the ability to obtain tissue-specific human and rodent stem cells from which more mature cell types can be generated (Alonso and Fuchs, 2003). The stem cells can be stored and grown as needed for assays and will likely become available for a wider and wider suite of tissues (Reya and Clever, 2005; Gaudio et al., 2009). Embryonic and fetal amniotic fluid stem cells can be used and differentiated through frequently tedious, multi-step processes to multiple cell types (DeCoppi et al., 2007). With tissue-specific stem cells, the route to mature cells is shorter and requires less manipulation (Wang et al., 2009).

In addition to availability of tissue specific stem cells, other advances bringing biomedical and small-scale manufacturing processes offer opportunities to utilize human 3-dimensional tissue in higher throughput contexts. For instance, Khetani and Bhatia (2008) discuss the application of semiconductor manufacturing microtechnology for fabrication of microscale tissues. A miniaturized, multiwall culture system for human liver cells with optimized microscale architecture maintained phenotypic functions or several weeks. These organotypic cultures could be useful in insuring better correspondence between *in vitro* tests and expected behaviors *in vivo*.

A major emphasis is required to produce appropriate assays with the right level of detail and an ability to provide appropriate read-outs across different responses levels. For risk/safety assessments with a single compound, rapid, *in vitro* testing for the suite of pathways is essential. High throughput and high data content methods were emphasized in the NAS report. In this usage, high throughput assays allow evaluation of hundreds or thousands of compounds across multi-point dose response in a period of just a few days. Some assays such as the organotypic liver assay above may not be amenable to high throughput. For toxicity testing, it is useful to distinguish the need for high throughput methods for testing large numbers of compounds from efficient *in vitro* tests that can be done over the course of days but may not be easily scalable to the ultra-high throughput. For evaluating the chemical space active for a particular pathway, high throughput permits evaluation pathway perturbations for large compound libraries, leading to better *in silico* modeling of structure activity relationships.

Mapping and modeling toxicity pathways

Assay outputs can be diverse as clearly evident from the US EPA ToxCast group of assays (Dix et al., 2007). Nonetheless, the area where the diverse array of new technologies has the greatest possible for contribution is in mapping and modeling the underlying signaling networks for specific toxicity pathways. The vast majority of perturbations are associated with networks that affect transcriptional control. Such a conclusion is obvious for so-called receptor-mediated toxicants, such as dioxin and the aryl hydrocarbon receptor, but is equally valid for stress response pathways. Antioxidant response signaling starts with oxidants reacting with cellular sensors – primarily Keap1. The modification of Keap1 leads to its dissociation from a complex with Nrf2, allowing Nrf2 and other partnering proteins to form a promotional complex altering expression of genes controlling cellular anti-oxidants (Motohashi and Yamamoto, 2004).

As toxicity pathway circuitry becomes better understood over time, it will be possible to create computational systems biology models for expected dose-response relationships for each of the assays used for toxicity testing following similar principles. Over the past decade, tools for mapping and modeling have blossomed. In a recent paper, Bromberg et al. (2008) described the network by which cannabinoid receptor (CB1R) controls neurite outgrowth. Activation of several hundred transcription factors within the nucleus after cell stimulation was measured to understand the logic of the signaling network. Bioinformatic methodologies connected CB1R to 23 activated transcription factors. Experiments with pharmacological inhibitors of kinases revealed a network organization of partial "OR" gates regulating kinases stacked above AND gates that control transcription factors. As in most instances of current research in systems pharmacology and network modeling, the goal of these studies was not dose-response as would be a primary interest for toxicity pathway analyses. This example provided a glimpse of the structure of the network without attempting a quantitative computational model. The epidermal growth factor (EGF) signaling network is particularly well studied. Amit and colleagues (2007) used a suite of experimental and bioinformatic tools to determine the forward signaling and feedback processes controlling the EGF network. The network was dissected by transcriptional profiling coupled with reverse phase protein lysate assays that assessed phosphorylation states of proteins within the EGF pathway. The analysis provided the structure of the logic of the circuitry for the early, immediate and later stage portions of the network.

Computational systems biology

It appears likely that a major contribution of 21st century science will be the application of an array of technologies to elucidation, mapping and modeling the behavior of the test systems for assessing toxicity pathway dynamics. The tools would include mRNA, transcription factor and phospho-protein time course profiling, coupled with bioinformatic technologies to extract network structure. The outcome would provide dynamics of the signaling networks and the dose-and time dependence of expected consequences of perturbations by test compounds, including positive controls for each of the pathways. Dynamic behavior of signaling networks have been described quantitatively using computational systems approaches focusing on models of transcriptional control (Alon, 2007; Aldridge et al., 2006). Theoretical descriptions of networks leading to better understanding of modular design elements in biological circuits have refined our vocabulary – concepts of ultrasensitivity, bistability, network gain, feedback and feed forward motifs, noise, stochasticity, and sequential levels of early, mid-term and late gene expression – to allow discussion of network behaviors with some commonality of terminology. These concepts are more extensively elaborated in a course text on "Computational Systems Biology and Dose Response" available at the Hamner Institutes web-site (http://www.thehamner.org/education-and-training/drm_workshop.html).

5 Creating the transformational mindset

In a Figure (5-1) in Chapter 5, the NAS report discussed a strategy for implementation, including ballpark estimates of the time (1 to 2 decades) and costs ($ 1-2 billion) for transitioning from current animal intensive toxicity testing to a toxicity pathway based approach. The report stressed the need for an organization to have the lead responsibility for overseeing the technology development to support the transition – a role that could eventually be played by an appropriate laboratory within the US National Institutes of Health. The overall timeline was shown in the report in a linear fashion leading to a transition to new approaches after completion of technology development for assays and achieving some confidence that the suite of assays would provide adequate coverage of possible pathway perturbations. In the current global economic climate and with a variety of competing interests for biomedical research, is it reasonable to expect federal agencies or the private sector to support such a long-term, expensive initiative?

Some aspects of the NAS vision are embedded in other programs. Three federal US agencies with responsibilities for

health-related research – the Environmental Protection Agency (EPA), the National Institute of Environmental Health Sciences' National Toxicology Program (NTP), and the National Institutes of Health Chemical Genomics Center (NCGC) – have a memorandum of understanding to conduct research necessary to advance the NRC committee vision for the future of toxicity testing. Collaboration among these organizations in the US will be essential in establishing a national commitment to develop the scientific foundation of the vision. This collaboration (Collins et al., 2008) focuses on research (1) to develop high throughput test methods, (2) to identify toxicity pathways, (3) to pursue targeted testing in short-term *in vitro* tests, and (4) to develop dose-response and extrapolation models. New approaches for *in vitro* toxicity testing and toxicity profiling are key parts of several federal programs in the US (Dix et al., 2007; National Toxicology Program, 2004). The US EPA ToxCast program (Dix et al., 2007) is using a variety of high throughput tests and computational methods to enhance prioritization of compounds for targeted testing in animals. A professed goal of the new interagency collaboration is predicting high dose results in animals. Prioritization and predicting high dose results are not part of the NAS vision. Nonetheless, the tools and approaches being developed in this collaboration will be important for achieving the long-term vision for transforming toxicity testing. Other tools will mature from efforts that are today primarily focused on animal alternatives (e.g., Spielmann et al., 2000).

In the past year, the Humane Society of the United States (HSUS) and its affiliates, the Humane Society Legislative Fund (HSLF) and Humane Society International (HIS) have taken steps to enlist partners to a stakeholder consortium – The Human Toxicology Project Consortium. The goal of this group is to facilitate the global shift to a cell response pathway paradigm for chemical safety assessments. This shift, in the words of the consortium, holds great promise for more rapid predictions of human health outcomes while superseding traditional animal testing for environmental agents and pharmaceuticals. The goals of this consortium is to (1) promote dialogue, information sharing and establishment of a research and development roadmap, (2) lobby for, coordinate and provide resources to support transatlantic efforts necessary to fulfill NAS vision, (3) engage in collaborative outreach to legislative, regulatory, corporate, academic and public interest audiences, and (4) to urgently develop a targeted research program to jump-start the transformation.

This targeted research plan, focusing on proof of concept efforts, would first focus on prototype compounds and provide examples of the application of results from toxicity pathway assays for risk/safety assessments sequentially rather than waiting 10 to 20 years to bring a totally new risk assessment paradigm on line. The proposed research over a 5 to 10 year period would provide examples with ten to fifteen pathway assays and generate opportunities for diverse stakeholders to gain experience in collecting and using these results for safety assessments. The outline of steps for this more targeted research program includes several components.

– Select about 10 prototypes compounds/pathway. These compounds would be chosen based on the breadth of information about animal toxicity and of the expected toxicity pathway targets, serving as a test bed for examining relationships be-

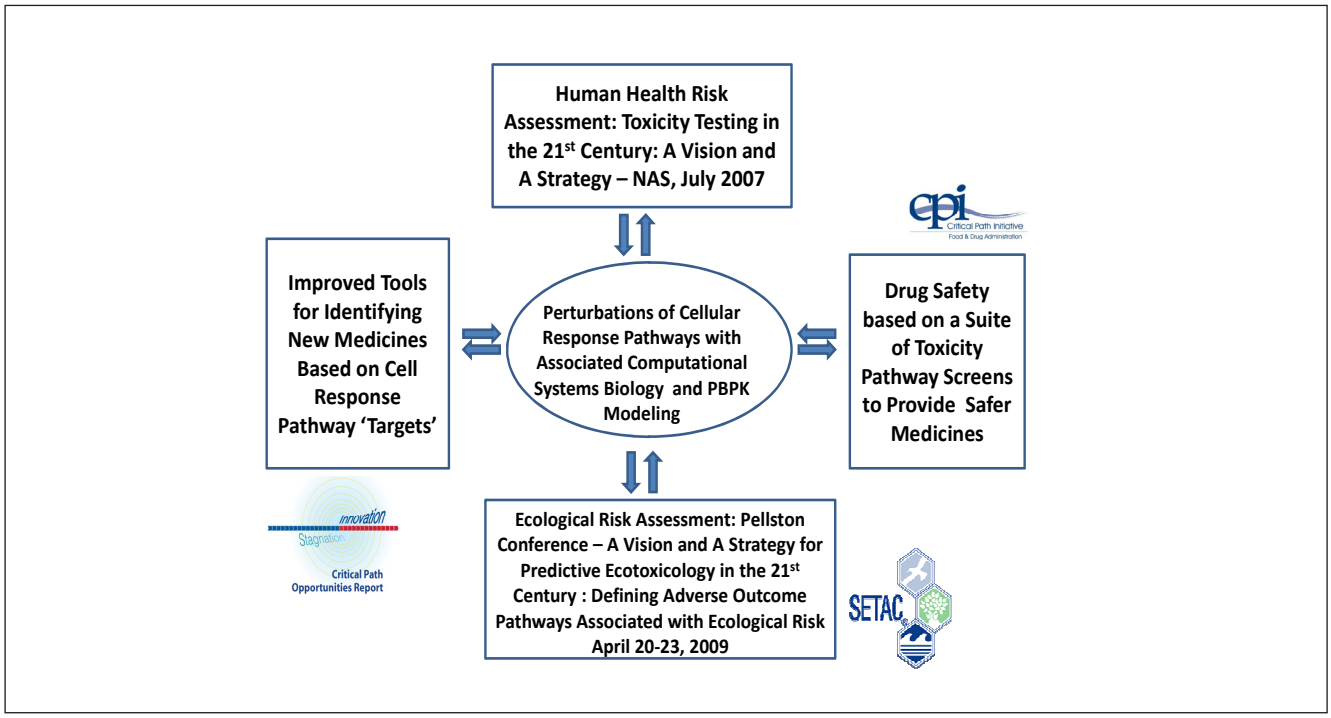


Fig. 4: Commonality of systems approaches to examining perturbations/modulations of normal biology for safety testing with environmental agents, pharmaceuticals, consumer products, and foods and for drug development. SETAC: Society for Environmental Toxicology and Chemistry.

tween *in vitro* toxicity test assays and historical information regarding *in vivo* results.
- Design appropriate cell-based toxicity assays. For these prototype compounds, test assay systems would focus on both rodent and a human assay, preferentially using stem cells or mature cells derived from these stem cells. With one or more of the prototypes, 3-dimensional tissue systems could be used for the assays.
- Develop the next generation quantitative risk assessment tools. These assays would be subjected to mapping and modeling analysis to uncover pathway circuitry, the dynamics of pathway responses to positive controls, and the dose response behaviors expected from different levels of perturbation.
- Examine relationships between perturbations and toxicity for prototypes. The assay design would require consideration of cascades that contain initial target activation, adaptive responses, and adverse responses with prolonged levels of perturbation.
- Integrate results from studies to provide representative health risk/safety assessments. The outcome of each of the prototypes would be risk/safety assessments that would be compared to more conventional approaches from animal toxicity data sets.
- Within the first 3-5 years expand from the first 10-prototypes to a larger suite of pathways/compounds. This transition should also allow some mid-course correction in the strategy, stemming from a continuing evaluation of successes and challenges in applying the new science in assisting human health safety assessment.
- With success in getting the program jump-started through the consortium, other partners, including toxicity testing organizations, regulatory agencies, and federal research organizations, could be enlisted as partners in moving forward with the transformation.

Regardless of which organization seizes leadership for the efforts to create the technology base for shifting to a new "Gold Standard," the central question is whether such an initiative is a good public health investment. From the point-of-view of sparing animal use and a more humane infrastructure for testing the answer has to be yes. Is it also likely to be a good investment in terms of its likely scientific value? The answer here is also a resounding yes! Our primary investment in toxicity testing today is simply box-checking, becoming a bit more mechanistically oriented for high value chemicals that show responses in animal toxicity tests. The *in vitro*, human biology approach, elaborated here and arising from the NAS vision, has a much reduced emphasis on rote testing and much increased emphasis on generating detailed understanding of the signaling pathways affected by chemicals and how perturbations/modulations in these pathways affect biological outcome. These tools and approaches will be just as valuable in drug safety/drug development, in evaluation of safety of food and consumer products, and in ecotoxicology (Watanabe et al., in press). In addition to the broader applicability for human health outcomes for modulation of these pathways (Fig. 4), the organization of information on pathway structure and function is a natural post-genome program that could provide a better understanding of health, disease and susceptibility within the human population – a much preferred investment compared to today's approach of *in vivo* testing and the cataloging of testing results.

6 Conclusions

Toxicity testing and much of the discipline of toxicology have reached a tipping point. Old practices focusing primarily on high dose studies to evaluate end-organ toxicity in animals are giving way to modern practices that assess how chemicals are likely to affect human biology and the concentrations under which these effects might be expected in exposed humans. This change will not occur easily. Even though the current toxicity testing is far from optimal, it is difficult to move away from entrenched traditional practices to a new footing. Change of this magnitude is discomfiting for most everyone. There are, of course, serious challenges to consider in such a transformation. They should not be dismissed or diminished. Chemical toxicity may relate to metabolism. How will the *in vitro* tests adequately assess metabolites with new compounds undergoing *in vitro* screening? Which of the observed perturbations will be considered appropriate for the safety assessment – will it be target activation, adaptive responses, or only some clear definition of overt toxicity in the cells? Will it be possible to describe circuitry for most toxicity pathways in enough detail to be confident in expected dose response behaviors? Finally, where will we find the scientists and regulators with the training and background to be comfortable with new practices? These issues are all important, legitimate questions that need to be considered. Yet, they should not divert us from the goal – to move toward a redefinition of the toxicity testing gold standard that focuses on human biology and perturbations of human toxicity pathways *in vitro*. We must bear in mind all the challenges, but push relentlessly toward the goal of a modern approach to human safety assessment.

References

Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nat. Cell. Biol. 8*, 1195-1203.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics 8*, 450-461.

Alon, U. (2006). *An Introduction to Systems Biology: Design of Biological Circuits*. Boca Raton, FL.: CRC Press.

Alonso, L. and Fuchs, E. (2003). Stem cells of the skin epithelium. *PNAS 100*, 11830-11835.

Amit, I., Citri, A., Shay, T. et al. (2007). A module of negative feedback regulators defines growth factor signaling. *Nature Genetics 39*, 503-512.

Andersen, M. E. and Krewski, D. B. (2009). Toxicity testing in the 21st century: Bringing the vision to life. *Toxicol. Sci. 107*, 324-330.

Andersen, M. E., Dennison, J. E., Thomas, R. S. and Conolly., R. B. (2005). New directions in incidence-dose modeling. *Trends Biotechnol. 23*, 122-127.

Bouvier d'Yvoire, M., Prieto, P., Blaauboer, B. J. et al. (2007).

Physiologically-based Kinetic Modelling (PBK Modelling): meeting the 3Rs agenda. The report and recommendations of ECVAM Workshop 63. *ATLA 35*, 661-671.

Bromberg, K. D., Ma'ayan, A., Neves, S. and Iyengar, R. (2008). Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science 320*, 903-909.

Clewell, H. J., III, Tan, Y. M., Campbell, J. L. and Andersen, M. E. (2008). Quantitative Interpretation of Human Biomonitoring Data. *Toxicol. Appl. Pharmacol. 231*, 122-133.

Coecke, S., Ahr, H., Blaauboer, B. J. et al. (2006). Metabolism: a bottleneck in in vitro toxicological test development. The report and recommendations of ECVAM workshop 54. *ATLA 34*, 49-84.

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science 319*, 906-907.

DeCoppi, P., Bartsch, G., Jr., Siddiqui, M. M. et al. (2007). Isolation of amniotic stem s-cell lines with potential for therapy. *Nat. Biotechnol. 25*, 100-106.

DeJongh, J., Forsby, A., Houston, J. B. et al. (1999). An integrated approach to the prediction of systemic toxicity using computer-based biokinetic models and biological in vitro test methods: overview of a prevalidation study based on the ECITTS. *Toxicol. In Vitro 3*, 549-554.

Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci. 95*, 5-12.

Gaudio, E., Carpino, G., Cardinale, V. et al. (2009). New insights into liver stem cells. *Digestive and Liver Disease 41*, 455-462.

Gulden, M. and Seibert, H. (2003). In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations in vitro. *Toxicology 189*, 211-222.

Heringa, M. B., Schreurs, R. H. M. M., Busser, F. et al. (2004). Toward more useful in vitro toxicity data with measured free concentrations. *Environ. Sci. Technol. 38*, 6263-6270.

Inglese, J., Johnson, R. L., Simeonov, A. et al. (2007). High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol. 3*, 466-479.

Inglese, J., Auld, D. S., Jadhav, A. et al. (2006). Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. USA 103*, 11473-11478.

Khetani, S. R. and Bhatia, S. N. (2008). Microscale culture of human livers cells for drug development. *Nat. Biotechnol. 26*, 120-126.

Krewski, D., Andersen, M. E., Mantus, E., and Zeise, L. (2009). Toxicity testing in the 21st century. *Risk Analysis 29*, 474-479.

Motohashi, H. and Yamamoto, M. (2004). Nrf2-Keap1 defines a physiologically important stress response mechanism. *Trends Mol. Med. 10*, 549-557.

National Research Council (NRC) (2007). *Toxicity testing in the 21st century: a vision and a strategy*. Washington, DC: National Academy Press.

National Research Council (NRC) (2006). *Toxicity testing for assessment of environmental agents: interim-report*. Washington, DC: National Academy Press.

National Toxicology Program (NTP) (2004). *A national toxicology program for the 21st century: A roadmap for the future*. NTP, NIEHS, Research Triangle Park, NC. (http://ntp.niehs.nih.gov/vision).

Nel, A., Xia, T., Madler, L. and Li, N. (2006). Toxic potential of materials at the nanolevel. *Science 311*, 622-627.

Reddy, M. B., Yang, R. S. H., Clewell III, H. J. and Andersen, M. E. (eds). (2005). *Physiologically Based Pharmacokinetics: Science and Applications*. Hoboken, NJ: John Wiley & Sons, Inc.

Reya, T. and Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nature 434*, 843-850.

Slikker, W. Jr., Andersen, M. E., Bogdanffy, M. S. et al. (2004a). Dose-dependent transitions in mechanisms of toxicity. *Toxicol. Appl. Pharmacol. 201*, 203-225.

Slikker, W. Jr., Andersen, M. E., Bogdanffy, M. S. et al. (2004b). Dose-dependent transitions in mechanisms of toxicity: Case studies. *Toxicol. Appl. Pharmacol. 201*, 226-294.

Spielmann, H., Müller, L., Averbeck, D. et al. (2000). European Centre for Validation of Alternative Methods. *ATLA 28*, 777-814.

Wang, Y., Yao, H. L., Barbier, C. et al. (2009). Matrix signals replacing lineage-dependent mesenchymal reeders for growth versus differentiation of hepatic stem cells to adult fates. *J. Experimental Med.*, in review.

Watanabe, K. H., Andersen, M. E., Basu, N. et al. Development of computational models of known adverse outcome pathways: domoic acid and neuronal signaling as a case study. *Environmental Toxicology and Chemistry*, in press.

Xiao, G. G., M. Wang, N. Li, J.A. et al. (2003). Use of proteomics to demonstrate a hierarchical oxidative stress response to diesel exhaust particle chemicals in a macrophage cell line. *J. Biol. Chem. 278*, 50781-50790.

Zhang, L., Theise, N., Chua, M. and Reid, L. M. (2008). The Stem Cell Niche of Human Livers; Symmetry between Development and Regeneration. *Hepatoloy 48*, 1598-1607.

Zhang, Q. and Andersen, M. E. (2007). Dose response relationship in anti-stress gene regulatory networks. *PLoS Comput Biol. 3(3):e24*. Epub 2006 Dec 22.

Correspondence to

Melvin E. Andersen, PhD, CIH, DABT, FATS
Director, Program in Chemical Safety Sciences
The Hamner Institutes for Health Sciences
Six Davis Drive, PO Box 12137
Research Triangle Park, NC 27709-2137, USA
e-mail: MAndersen@TheHamner.org

Early Morning Sessions

Session MS3: Animal welfare associations

Foundation Animalfree Research: 33 Years of Commitment to Alternatives to Animal Experiments

Stefanie Schindler
Animalfree Research, Switzerland

Summary
This article describes the beginnings and the history of our foundation, some of the main projects of the past, its current activities as well as the foci of the future.

Keywords: Animal welfare, 3Rs, animal experimentation

1 Introduction

The foundation Animalfree Research recently celebrated its 33rd birthday. During all these years, it has dedicated itself exclusively to the abolishment of animal experimentation via the development and implementation of alternative methods that replace or at least reduce the numbers of experimental animals, informing the public and influencing political opinion. From the first moment this approach was regarded as the only viable way to achieve sustained success in the abrogation of *in vivo* experimentation.

2 The founding of Animalfree Research (Fonds für versuchstierfreie Forschung, FFVFF)

In the 1970s, when hair-raising stories and pictures of animal experiments rattled people all over the world, one of the later founders of Animalfree Research, Irène Hagmann, then editor of the Zurich animal welfare organisation (*Tierschutzbund Zürich*), came upon an insert in an Italian newspaper on animal experimentation. Horrified by this report, she established contact with a journalist, Susi Goll, who had recently written an article on dolphins used as living explosive devices by the US Military. She too had seen a brochure on animal experiments distributed at a booth of the Zurich animal welfare organization (*Tierschutzbund Zürich*). The two women joined forces and, with a sum of CHF 5,000, established the independent "*Fonds für versuchstierfreie Forschung*", an organisation dedicated exclusively to the support of research methods without the use of animals. To this day, the financing relies entirely on private donations and contributions of other welfare organisations.

3 Aims and objectives

Animalfree Research was founded to actively replace and reduce animal experimentation, to sensitise public and political opinion makers to animal welfare issues and to impress on people their duties and responsibilities towards experimental animals. Inspired by documents of FRAME, the collaboration with science, politics and the public was a major focus from the very beginning. Likewise, close networking with other animal welfare organisations and participation in expert groups and commissions was the order of the day.

The range of activities, accordingly, was and still is very broad. Partial or complete financing of research projects, meta-analyses, the organisation of workshops, and the development of an argumentary against the use of animals, be it from an ethi-

cal, scientific or legal point of view, as well as participation in political debates, constitutes the work of our foundation. Furthermore, we are currently working on establishing ourselves as a counselling centre for the public and the media, and for researchers who are interested in implementing in vitro techniques in their projects.

4 Examples of activities and projects

The very first project was a literature study about the questionable scientific value of the single dose oral acute LD_{50} test, performed by Prof. Dr. Gerhard Zbinden, at that time Director of the Toxicological Institute in Zurich (Zbinden and Flury-Roversi, 1981). Among other things the study demonstrated that the test is without any relevance for substances with a very low acute toxicity and yields variable results depending on species and strain, age, sex, procedures, etc. The study had great impact in the scientific world and marked the beginning of the end of a scientific experiment that had until then been regarded as indispensable. The OECD Council adopted the decision to delete the respective OECD Guideline TG 401 on 17 December 2002.

From the beginning, the importance of distributing scientific results in the field of the 3Rs was clear. In 1984, a small journal named "*Alternativen zu Tierexperimenten*" was published for the first time, edited and financed by the FFVFF. One of the, if not the, most important project of our foundation, the journal later came to be known as ALTEX. In 1999, the annual awarding of the ALTEX prize for the best and most relevant article in ALTEX was made an institution.

The development and distribution of the PharmaTutor by Daniel Keller at the Institute of Pharmacology, University of Zurich, proved another milestone success. The Tutor was made freely available and turned out to become, in its time, the most widely used simulation programme for education worldwide.

The project "Serumfree", which was co-financed by FFVFF and two other animal welfare organisations, with Prof. René Fischer of Zurich aimed at stopping a development which had started to impair the successes of the 3Rs with concern to animal welfare: the continuing shift to cell and organ culture methods instead of *in vivo* experimentation led to an ever-increasing demand for fetal calf serum. In 1995, the consumption was estimated to be around half a million litres per year (Hodgson et al., 1995), corresponding to around 1 million unborn calves. A PhD study further elucidated questionable practices in obtaining and marketing these materials, shedding light on the varying composition and quality of different lots of serum (Jochems, 1997). The development of cell culture methods that do not require calf serum for growth and maintenance and a procedure for cryopreservation of cells in serum-free media could be reported (Gonzalez Hernandez and Fischer, 2007).

5 Current projects

In 2009, Animalfree Research began to finance a follow-up project, which uses the results of the previous year's study on the impending rise of animal experiments in the field of nanotechnology and the possible *in vitro* alternatives (Sauer, 2008). Throughout Europe, governments have established action plans and workshops in order to assess risks and health concerns of nanoparticles. In this project, the responsible project leaders are approached and persuaded to focus on non-animal methods and animal welfare issues in their safety and toxicity assessment.

With its contribution, Animalfree Research helps to develop an exposition system for the controlled deposition of substances on cells at the air-liquid interface, which focuses on lung epithelial cells. This project optimises a standardised inhalation chamber for diverse possible applications: e.g. animal-free development of inhalable drugs, screening for toxicity of soluble and non-soluble substances, and biological effects of nanoparticles. The final goal is the establishment of a cost-effective, relevant and easy to use system, which is suitable for widespread application, dosimetrically exact and capable of imitating the *in vivo* situation in the lung as closely as possible. Once developed, the chamber can be adapted to other research fields which involve substance-cell interactions.

In a joint effort, Animalfree Research and the Doerenkamp-Zbinden Foundation support a project which establishes an animal-free method for the safety testing of tetanus vaccines. After inactivation, these vaccines have to be tested for residual toxoid activity in guinea pigs, which are then examined for symptoms of tetanus. At the Paul-Ehrlich Institute, an *in vitro* system that can measure active Tetanus neurotoxin is being developed (Behrensdorf-Nicol et al., 2008).

In October 2009, Animalfree Research hosted a public discussion forum dedicated to the memory of Prof. G. M. Teutsch, focussing on the concept of the animal's dignity, which was introduced into Swiss legislation in September 2008, and its practical implications on authorisation procedures and research in Switzerland. It focused on primate use and the preservation of the animals' dignity from the viewpoint of research and animal welfare. The forum aimed to illuminate obstacles and difficulties in maintaining respect for the animal in the routines of research and to demonstrate possible improvements and solutions.

6 Outlook

In the next years, Animalfree Research aims to work on a stop to severely distressing animal experiments, which we consider irreconcilable with the concept of the animal's dignity introduced into Swiss legislation in 2008. In the same vein, we will focus on the abolishment of primate experimentation, since the

welfare issues of neither the experiments themselves nor the associated stress, e.g. with capturing and housing, can outweigh the possible uses of the research results. The outcome of the Federal Court's decision on the ban of specific primate experimentation will strongly influence the speed and efficiency with which these experiments can be phased out.

In addition, we will continue to broach the issues of transgenic animals, a topic with steadily increasing urgency as animal numbers continue to rocket.

References

Behrensdorf-Nicol, H. A. et al. (2008). Residual enzymatic activity of the tetanus toxin light chain present in tetanus toxoid batches used for vaccine production. *Vaccine 26*, 3835-3841

Gonzalez Hernandez, Y. and Fischer, R. W. (2007). Serum-free culturing of mammalian cells--adaptation to and cryopreservation in fully defined media. *ALTEX 24(2),* 110-116.

Hodgson, J. (1995). To treat or not to treat: that is the question for serum. *Biotechnology (United States) 13(4),* 333-334.

Jochems, C. E. A. (1997). Use, trade and harvest of livestock sera. Thesis. Utrecht University, The Netherlands (http://www.vet.uu.nl/nca/userfiles/other/report_use_trade_and_harvest_of_livestocksera.pdf)

Lenz, A.-G. et al. (2009). A dose-controlled system for air-liquid interface cell exposure and application to zinc oxide nanoparticles. *Particle Fibre Toxicology*, submitted.

Sauer, U. (2009). Animal and non-animal experiments in nanotechnology – the results of a critical literature survey. *ALTEX 26(2),* 109-128.

Zbinden, G. and Flury-Roversi, M. (1981). Significance of the LD50-Test for the toxicological evaluation of chemical substances. *Arch Toxicol 47*, 77-99.

Correspondence

Dr. med. vet. Dr. rer. nat. Stefanie Schindler
Hegarstrasse 9
Postfach 1766
8032 Zürich
Switzerland
e-mail: schindler@animalfree-research.org

CAAT: A 3Rs Center for the 21st Century

Carol J. Howard

The Johns Hopkins Center for Alternatives to Animal Testing (CAAT), Baltimore, Maryland, USA

Summary

The Johns Hopkins Center for Alternatives to Animal Testing (CAAT), founded in 1981, is one of the oldest of the 3Rs Centers. For 28 years, CAAT's innovative programs have served to promote the creation, development, validation, and use of alternatives to animals in research, product safety testing, and education. CAAT also is, in many ways, one of the newest of the 3Rs Centers, with a new director – former Head of ECVAM Thomas Hartung – and an ever-expanding array of new programs and projects.

CAAT's diverse activities and resources range from our long-standing research grants program, workshops and symposia, and awards to such new developments as a Transatlantic Think Tank for Toxicology (t^4); the union of Altweb and the journal ALTEX (now in English), a new 3Rs Centers website, policy and outreach programs, implementation of the report, Toxicity Testing and Assessment in the Twenty-first Century: A Vision and a Strategy, and more.

Keywords: CAAT, alternatives, 3Rs, toxicology, Altweb, ALTEX

1 Introduction

For nearly 30 years, the Johns Hopkins Center for Alternatives to Animal Testing (CAAT) has been working to promote the creation, development, validation, and use of alternatives to animals in research, product safety testing, and education.

In 1993, CAAT hosted the first World Congress on Alternatives and Animal Use in the Life Sciences, held in Baltimore, USA. Planning for that Congress started in 1990 at CAAT's 7th annual symposium, where 33 leaders in the field from 13 countries shaped the vision that led, 19 years later, to WC7 in Rome, Italy.

2 What is CAAT?

CAAT is an academic, science-based center within the Johns Hopkins Bloomberg School of Public Health. Established in 1981 with a grant from the Cosmetic, Toiletry and Fragrance Association (now the PCPC), the Center is dedicated to the proposition that humane science is the best science.

CAAT's mission is to:

- Promote and support research in the development of in vitro and other alternative techniques
- Serve as a forum to foster discussion among diverse groups leading to creative approaches to facilitate acceptance and implementation of alternatives
- Provide reliable information on the science, philosophy, and public policy of alternatives to academia, government, industry, and the general public
- Educate and train in the application of alternatives

CAAT's research grants program serves to provide critical seed money for scientists interested in developing alternative methods. To date, the Center has funded 300-plus grants for a total of more than $ 6 million. CAAT grants are available to researchers world-wide.

Since its founding, CAAT has held more than 20 major workshops and symposia, bringing together diverse stakeholders to find common ground and devise ways to make research and testing as effective and humane as possible. The Center currently is organizing a workshop for the American Chemistry Council, to be held in Baltimore in July 2010.

CAAT faculty members are involved in teaching about alternatives, both at Hopkins and elsewhere, and have developed a free online course to make this information more broadly available. In 2007, CAAT launched a new program in public policy, education, and outreach. The Center also spearheaded the development of a certificate program in Humane Sciences and Toxicology Policy, offered through the Bloomberg School of Public Health.

CAAT's information program produces a variety of 3Rs-related materials, in both print and electronic media. Print pub-

lications include brochures, posters, flyers, reports, editorials, and letters to the editor. CAAT also created and maintains two websites: the CAAT Website (http://caat.jhsph.edu) and Altweb, the Alternatives to Animal Testing Website (http://altweb.jhsph.edu).

The CAAT site is devoted to the Center's programs, activities, and publications. Resources available on this site include:
– Information about CAAT grants
– Proceedings and technical reports from CAAT workshops and symposia
– Alternatives books and articles by CAAT staff
– Free online course: Enhancing Humane Science / Improving Animal Research
– Information about the Hopkins certificate program in Humane Sciences and Toxicology

Altweb, the Alternatives to Animal Testing Website, was created to serve as a gateway to alternatives resources, information, and news – on the Web and beyond. The site came about as result of a 1997 ECVAM report (Janusch et al., 1997) that called for "a central reference point" for alternatives information. CAAT took on that challenge, launching Altweb on October 1, 1997. The site is continually updated and expanded, including a redesign and reorganization in 2009.

CAAT manages Altweb on behalf of an international project team, which helps provide guidance, direction, and feedback for the site. The Project Team has grown from 10 member organizations in 1997 to 26 in 2007, representing academia, industry, animal welfare, and government regulatory organizations. The site is truly international in scope, drawing visitors from more than 130 countries.

Altweb serves a broad range of audiences, including:
– Biomedical researchers
– Industry
– The international regulatory community
– IACUCs, Ethics Committees, and others who review animal protocols
– Veterinarians, lab technicians, and others who work with laboratory animals
– Educators
– Students
– The animal welfare community
– The international alternatives community
– The general public

Altweb offers a broad array of resources, including:
– Alternatives news
– Meetings calendar
– Journal abstracts (ALTEX, ATLA, In Vitro Animal, Laboratory Animals, Toxicolgy in Vitro, and more)
– Proceedings
– Technical reports
– Books
– Fact sheets
– Databases
– Regulations information from around the world
– Links to many other relevant resources

The Altweb homepage provides the latest in alternatives news, as well as information on upcoming meetings in the field. The Frequently Asked Questions (FAQs) section, available in both English and Spanish, is one of the site's most frequently visited pages. It covers basic introductory information and is particularly useful for students and the general public.

Special Features of Altweb:
– Full text of Russell and Burch's *Principles of Humane Experimental Technique*, the classic book that launched the field of alternatives
– Guide to searching for alternatives
– Alternatives in education: an introduction and overview
– Spanish content (FAQs, glossary, links to other resources en Español)
– Special section on monoclonal antibodies
– Special section on refinement

The "Search for Alternatives" guide is a particularly valuable resource, providing a step-by-step approach to the search process, with detailed information on relevant databases, policies and regulations in various countries, sample searches, and more. The section on databases describes each database and the topics covered, as well as indicating whether a database is free, proprietary, or government/regulatory, and which may be most useful for research, for teaching, or for testing.

The special section on monoclonal antibodies (MABs) was developed with the goal of organizing and synthesizing information on this complex subject. It consists of an introductory text explaining key topics in non-technical language, accompanied by a set of links to relevant databases, websites, books, articles, abstracts, etc. The questions and concepts addressed in the MABs section include:
– What are monoclonal antibodies? Definition, description, uses, history, issues
– Ascites method versus *in vitro* methods: cost, effectiveness, pain and distress
– MAB production: description of methods
– Policies, regulations, guidelines, recommendations
– Where to get *in vitro* MABs: academic core centers and commercial facilities
– Reports and proceedings

The special section on refinement is modeled after the Altweb section on monoclonal antibodies, providing introductory text plus links to key resources on the following concerns:
– What is pain and distress? Definitions, biology, and physiology
– Recognition and assessment of pain and distress
– Alleviation and prevention of pain in animals
– Humane endpoints
– Euthanasia
– Enrichment

3 New at CAAT

New CAAT Director
- Thomas Hartung, MD, PhD
- Former Director of the European Centre for the Validation of Alternative Methods (ECVAM)
- Alan Goldberg, founding director of CAAT, now serves as chairman of the CAAT Advisory Board

New: Doerenkamp-Zbinden Professor & Chair for Evidence-based Toxicology (EBT)
- Thomas Hartung installed in May 2009 as inaugural Doerenkamp-Zbinden Professor & Chair for EBT
- EBT is modeled on the "Cochrane Collaboration" of evidence-based medicine (EBM)
- EBM works to provide valid & accepted consensus information on clinical trials & methodologies
- Will apply a similar approach to toxicology

New: Implementation of the NAS report: *Toxicity Testing in the 21st Century: A Vision and a Strategy*
- In 2007, U.S. National Academy of Sciences (NAS) released the report: *Toxicity Testing in the 21st Century: A Vision and a Strategy*
- Advocates sweeping & transformative changes in regulatory toxicity testing
- Shift from current whole-animal methods to testing in vitro methods, human cells in culture, and mechanisms of toxicity as understood through systems biology
- Will change the way toxicology is practiced in the future

New: Transatlantic Think Tank for Toxicology (t^4)
- Think tank aimed at implementing the NAS report
- Collaborative effort between CAAT and Doerenkamp-Zbinden Foundation
- Four leaders in EBT to serve as "ambassadors" of t^4
- Marcel Leist (University of Konstanz); Bas Blaauboer (Utrecht), Thomas Hartung & Alan Goldberg (Johns Hopkins)
- High quality analyses of toxicological problems, workshops, reports & scientific papers

New: Research Lab: Developmental Neurotoxicity (DNT)
- Thomas Hartung is establishing a lab at Johns Hopkins to develop alternative methods
- Initial focus will be on developmental neurotoxicity (DNT)

New: Altweb + ALTEX Collaboration
- Altweb has joined forces with the journal ALTEX
- Formerly published in a mixture of German and English, ALTEX is now all in English
- ALTEX articles will be freely available on Altweb
- ALTEX includes regular "Food for Thought" column by Thomas Hartung

New: 3Rs Centers Website
- To be launched at WC7: website designed to serve as information hub for 3Rs & alternatives organizations around the world
- Searchable using keywords, tags & a global map
- Grew out of discussion at 3Rs centers meeting at WC6 in Tokyo

New: CAAT Information Days: Working with Industry
- Day-long workshops focusing on a specific issue facing industry
- Brings together stakeholders, experts, and solution-providers
- First information day held 8th July focused on the 7th Amendment of the EU Cosmetics Directive
- November 2009 information day addressed issues faced by US industry members in complying with REACH legislation.

New: EU Center of Excellence
- CAAT is now part of the American Consortium on European Studies (ACES), an EU Center of Excellence based at the Johns Hopkins School of Advanced International Studies (SAIS)
- CAAT established a Humane Sciences and In Vitro Alternatives component
- CAAT held its first ACES-sponsored Congressional briefing on Capital Hill in June

New: CAAT Europe
- In December 2009, CAAT announced the formation of the Johns Hopkins Center for Alternatives to Animal Testing-Europe (CAAT-EU)
- A collaboration between the Johns Hopkins School of Public Health (US) & the University of Konstanz Department of Biology (Germany)
- To coordinate transatlantic activities to promote humane science
- To be housed at the University of Konstanz

5 CAAT online

For more information, please visit:
- Altweb + ALTEX http://altweb.jhsph.edu
- CAAT website: http://caat.jhsph.edu
- 3Rs Centers website: http://caat.jhsph.edu/international_alternatives
- A Boundless Ethic (CAAT blog) http://aboundlessethic.com

Correspondence to
Carol J. Howard
The Johns Hopkins Center for Alternatives to Animal Testing (CAAT)
615 N. Wolfe Street, W7032
Baltimore, Maryland, USA
e-mail: choward@jhsph.edu

A Proposal to Establish a Brazilian Center for Validation of Alternative Methods (BraCVAM)

Octavio Presgrave[1], Chantra Eskes[2], Rosaura Presgrave[1], Eloisa Alves[1], João Carlos Freitas[1], Cristiane Caldeira[1], Izabela Gimenes[1], Ronald Silva[1], Saulo Nogueira[1], Jadir Nunes[3], Ekaterina Rivera[4], Vanessa Sá-Rocha[5], Sandra Coecke[6] and Thomas Hartung[7]

[1]National Institute of Quality Control in Health (INCQS)/Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil; [2]Independent consultant, Italy; [3]Brazilian Association of Cosmetology, São Paulo, Brazil; [4]Biological Science Institute, Federal University of Goiás, Brazil; [5]Natura, São Paulo, Brazil; [6]ECVAM, Institute for Health and Consumer Protection, EC DG-JRC, Ispra, Italy; [7]Center for Alternatives to Animal Testing (CAAT), Johns Hopkins University, Baltimore, USA

Summary

Brazil recently approved law number 11.794/2008, which regulates the use of animals for scientific purposes. Many products on the Brazilian market are still required to be controlled by animal testing. In Brazil, there is no improved mechanism for funding collaborative studies, and there is no institution responsible for managing and coordinating such studies. Oswaldo Cruz Foundation (FIOCRUZ) assembles all the conditions to be the headquarter of the Brazilian Center for Validation of Alternative Methods (BraCVAM), since it is an internationally recognised scientific institution uniting a large number of scientific fields, including basic and applied research, drug and vaccine production, quality control, teaching, hospitals, etc. The multidisciplinary scientific infrastructure of FIOCRUZ could be used to establish a network in different fields of knowledge. The creation of BraCVAM would facilitate the development and validation of tests, not only in Brazil but also in South America and the Caribbean.

Keywords: BraCVAM, alternative methods, validation, regulatory acceptance

1 Introduction

The Brazilian Government approved law 11.794 in October 2008. This law, known as "Lei Arouca" (Arouca's Law) in homage to the physician that first proposed it, regulates the use of animals in research and teaching (Brazil, 2008). Before this law, some other legislation touched upon this subject, i.e. laws that regulate environmental crime, veterinarian professionals, etc. Law 11.794 is specifically focused on the use of animals in experimentation and teaching, but not for agronomical purposes.

Law 11.794 creates the National Council of Animal Experimentation (CONCEA) that includes in its responsibilities monitoring of and evaluating the introduction of alternative techniques that replace the use of animals in teaching and experimentation (Marques et al., 2009). However, the law does not state that whenever an alternative method exists, animals may no longer be used (Machado et al., 2009). CONCEA is composed of members of related Ministries and scientific societies, including representatives of animal protection societies (Brazil, 2008). However, CONCEA lacks technical support.

Therefore, the creation of a Brazilian Center comparable to ECVAM, ICCVAM and JaCVAM would be an important step for such studies, which could then be coordinated in an efficient manner. In addition, the Center would be invaluable for contact and interaction with similar organisations in other countries throughout the world.

2 State of the art in Brazil

There are many animal rights protection groups in Brazil, some of them representatives of groups from other countries. Although this movement is not as intense as it is for example in Europe, the need to validate and establish alternative methods according to the 3Rs has compelled a large number of laboratories to perform studies in this field.

Currently about 15 to 20 groups are working on alternative methods. These groups are at official laboratories, universities, industry and private laboratories. Most of them are trying to validate alternative methods to replace animals in assays such as skin and eye irritation, pyrogen test and vaccine potency, while some others are studying alternatives in education, nutrition and environmental enrichment.

Brazil does not have an organisation that is responsible for the coordination of validation studies. The current research in this field is due to particular initiatives of the above mentioned institutions, which, motivated by precedents in other countries, have started working on the development of alternative methods.

Brazilian Decree Number 4725, dated 9th June 2003, dictates that one of the activities of the National Institute of Quality Control in Health (INCQS) is to establish the quality control rules and procedures for the health laboratory network (Brazil, 2003a). However, except when methods are already internationally validated, it is impossible to comply with this ruling without

conducting a collaborative study to harmonise the procedures and to define the parameters for the interpretation of the results they provide. Even when methods are already validated, some may need to be adapted for specific local needs since Brazil controls final products instead of raw material and some products are specific to the country and are not the object of international validation, etc.

Although Brazilian *Resolution No. 899*, dated 29th May 2003, published the *Guideline for the Validation of Analytical and Bioanalytical Methods* (Brazil, 2003b), the whole process is not defined, nor is the role of each entity during the various steps established, as was done in the EU initiative which established ECVAM (Anonymous, 1991) and subsequently in the OECD guidance on validation (OECD, 2005).

3 Backgrounds

It took a long time for Brazilian meetings to introduce a special session on alternative methods. Although there are no documents on it, it probably started to occur around 2000. Before this time, studies on alternative methods were presented at congresses and similar events as poster presentations in different sessions, i.e. only related to the specific subject (e.g. toxicology, animal welfare, etc.).

In 2005, INCQS organised the Brazilian Meeting on Alternative Methods to Animal Use for Regulatory Purposes (EMALT) where problems related to financial support for research in the alternative area and problems in validating assays were discussed (Presgrave and Boghal, 2005). That was the first time that the need of establishing a Brazilian Center for validating alternative methods was declared during the discussion of mechanisms of the validation process.

In 2007, at the XV Brazilian Congress of Toxicology, Prof. Dr. Thomas Hartung and Prof. Octavio Presgrave both presented the need for creating the Brazilian Center for Validation on Alternative Methods (BraCVAM) in their speeches. The BraCVAM proposal was first published in a recent article (Presgrave, 2008).

4 Validation centres around the world

Many institutions in the world work on development and validation of alternative methods and animal welfare based on the 3Rs principle. The BraCVAM idea was based upon some of them.

European Centre for the Validation of Alternative Methods (ECVAM)

ECVAM was created in 1991, as part of the European Commission's Joint Research Centre (JRC), in order to fulfil the requirements outlined in Article 23 of *EU Directive 86/609/EEC*, which aimed to stimulate the development and validation of alternative methods which would be able to supply at least the same level of information given by existing animal assays, but which would use fewer animals, or significantly reduce their suffering, or replace their use altogether (Anonymous, 1991). ECVAM's role in achieving these aims is mainly via the coordination of the validation of these alternative methods.

ECVAM was located in the JRC in Ispra, Italy, since the JRC is a multilingual establishment, with a history of acting in the promotion of international scientific cooperation. This factor facilitated the design of the correct technical structure needed for ECVAM's work, besides providing a basis for expanding the role of the JRC in research (Anonymous, 1991).

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

ICCVAM was founded in September 1994 as an ad hoc committee for presenting a report to support the requirements of the 1993 *NIH Revitalization Act*. In this Act, the National Institute of Environmental Health Sciences (NIEHS) was required to establish criteria for the validation and regulatory acceptance of toxicological alternative methods (ICCVAM, 2008). This committee became permanent in December 2000 and nowadays comprises representatives from 15 US federal regulatory and research agencies.

At present, ICCVAM receives scientific support from National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods NICEATM), which is part of NIEHS, located at Research Triangle Park, NC, USA. ICCVAM's mission is to promote the development, validation and regulatory acceptance of alternative methods, emphasising the 3Rs principles, while scientific quality and human, animal and environmental health are upheld and promoted (ICCVAM, 2008).

Japanese Centre for the Validation of Alternative Methods (JaCVAM)

JaCVAM was established in November 2005 and is affiliated to the Japanese National Centre for Biological Safety and Research (NCBSR) of the National Institute of Health Sciences (NIHS) (Ohno, 2005a; Kojima, 2007).

The mission of JaCVAM is to propose and improve the validation of alternatives and to establish international cooperation besides developing new alternative methods (Ohno, 2005b; Kojima, 2007).

Centre for Documentation and Evaluation of Alternatives to Animal Experiments (ZEBET)

ZEBET was established in 1989 and its aim is to bring about the replacement of legally prescribed animal experiments with alternative test methods, to reduce the number of test animals to the absolutely necessary level and to alleviate the pain and suffering of animals used in experiments. ZEBET is also responsible for documenting alternatives to animal experiments. By this mode of action, ZEBET is the information office for alternatives to animal experiments for the federal authorities.

ZEBET also coordinate experimental validation of methods not involving animal experiments in order to include them in official safety toxicological test directives. ZEBET has the advantage of undertaking its own research and has a separate budget to promote specific projects on the development of alternative methods (ZEBET, 2008).

5 The Brazilian Centre for the Validation of Alternative Methods (BraCVAM)

Since Brazil does not have an institution that coordinates alternative studies and these studies are important for the development of research, it is imperative to create BraCVAM (Presgrave, 2008).

Oswaldo Cruz Foundation (FIOCRUZ) possesses all of the necessary attributes for it to become the headquarters of BraCVAM, since it is an internationally-recognised scientific institution that unites a vast range of scientific and medical fields, including basic and applied research, drug and vaccine production, quality control and teaching. This is an important aspect, since the multidisciplinary scientific infrastructure of FIOCRUZ could be used to form a working platform with a wide range of available knowledge on many relevant subjects.

Between 2001 and 2007, FIOCRUZ published 6,359 articles in Brazilian and international journals. In 2007, the ratio of published articles in the biomedical area was about 60% of the total articles (Teixeira et al., 2009).

INCQS, and some other groups in FIOCRUZ, have been studying alternative methods for a long time. Since INCQS is the official laboratory of the Brazilian Ministry of Health, it has to perform assays to guarantee that products that are on the market are sufficiently safe. For this purpose, INCQS routinely performs animal toxicological tests and, since 1989, has tested most of these products with alternative methods, for comparing results and evaluating correlations between animal and non-animal data.

INCQS has frequently participated in several Brazilian and international congresses, including the *World Congress on Alternatives and Animal Use in the Life Sciences*. It comprises a group of approximately 20 professionals and students working on alternative methods, mainly for the replacement of animals in testing for skin, eye and mucous irritation, and in sensitisation and pyrogen testing, as well as in vaccine control. During this period, a range of activities, such as poster presentations, lectures, organisation of meetings and round tables, paper publications, and post-graduate studies, have been undertaken.

BraCVAM action should not be restricted to the experimental area but must encompass studies on alternatives in education, environmental enrichment, etc. The Center must act in the development of alternative research in every area in which it is possible to apply the 3Rs concept. Of course these structures should be constructed step by step. Integration with similar institutions is very important in order to start establishing mechanisms of validation studies.

The BraCVAM structure should present a coordination division and divisions that reflect areas of action as recently proposed by Eskes and coworkers (2009, see Fig. 1). The scope of these areas may be as following:

a) Promotion and Networking
   – Creation and maintenance of a centralised website & database on alternative methods
   – Monitoring and creation of a network of interested parties on alternative methods in Brazil
   – Organisation of seminars and specialised workshops
   – Publications

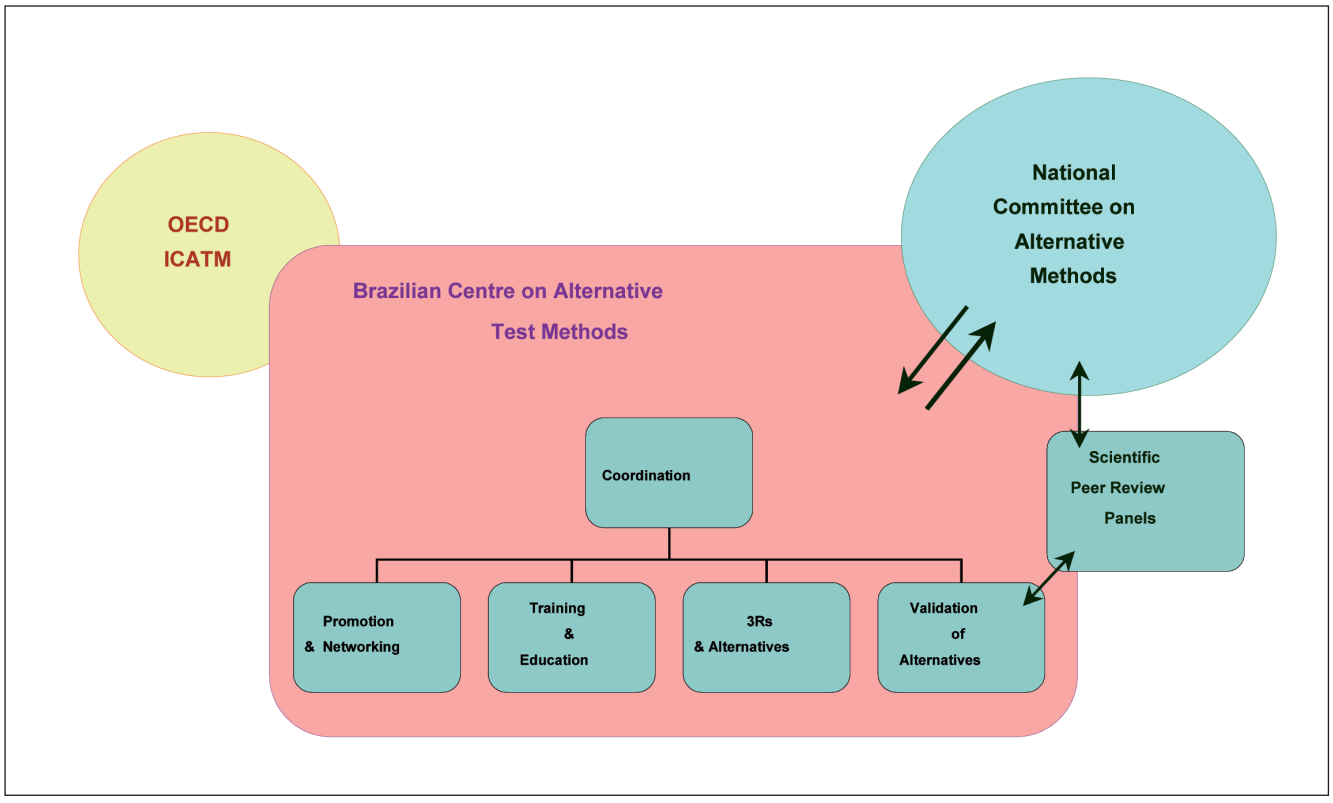


Fig. 1: Proposed organogram for BraCVAM and relationship with some external committees and institutions.

b) Training and education
  – Training on the 3Rs principles
  – Practical courses on validated alternative methods
  – Training on the internationally agreed principles of validation
c) 3Rs and alternatives
  – Identification of current regulatory, industrial and academic needs
– Identification of available alternatives
– Creation of platforms of excellence on specific fields of expertise
  – Identification of most appropriate alternatives for specific applications
d) Validation of alternative methods for regulatory purposes
  – Coordinating studies on the development of alternative methods to animal testing
  – Coordinating validation studies of alternative methods to animal testing
  – Proposing and evaluating test protocols
  – Participation in international validation studies
  – Publication of reports about the progress of studies

Once established, the activities of BraCVAM shall include:
– the development of new methods;
– the coordination of studies on the development of alternative methods;
– the coordination of the validation of alternative methods;
– the proposal and evaluation of test protocols;
– the analysis and evaluation of results obtained from studies;
– participation in international cooperation;
– the creation, maintenance and management of databases of technical procedures, substances, raw materials, finished products and any other aspects relevant to alternative methods;
– the promotion of meetings, congresses, workshops or any other route of disseminating scientific information to regulatory organisations, industries, academia and any other institutions related to alternatives;
– the publication of reports about the progress of studies; and
– the improvement of any procedure that could help in the development of alternative methods.

6 Conclusion

The creation of a Brazilian Centre for the Validation of Alternative Methods (BraCVAM) would be very important, not only for, but for the whole of Latin America, since there is no similar institution in this part of the world.

Once established, BraCVAM should determine its structure, take all necessary steps to become officially viable, and determine the start-up period for its activities.

References

Anonymous (1991). *Establishment of a European Centre for the Validation of Alternative Methods (CEVMA)*. Communication from the Commission to the Council and the European Parliament, 29 October 1991. *SEC (91) 1794 final*, 6pp. Brussels, Belgium: Commission of the European Communities.

Brazil (2003a). *Decreto No. 4725, de 9 de junho de 2003. Aprova o Estatuto e o Quadro Demonstrativo dos Cargos em Comissão e das Funções Gratificadas da Fundação Oswaldo Cruz — FIOCRUZ, e dá outras providências. Diário Oficial [da República Federativa do Brasil],* de 10 de junho de 2003, pp. 8-12. Brasília, Brazil: República Federativa do Brasil.

Brazil (2003b). *Resolução RE No. 899, de 29 de maio de 2003. Determina a publicação do "Guia para validação de métodos analíticos e bioanalíticos". Diário Oficial [da República Federativa do Brasil],* de 02 de junho de 2003, pp. 56-59. Brasília, Brazil: República Federativa do Brasil.

Brazil (2008). *Lei n. 11.794, de 08 de outubro de 2008*. Regulamenta o inciso VII de parágrafo 1o do art. 225 da Constituição Federal, estabelecendo procedimentos para o uso científico de animais; revoga a Lei n. 6.638, de 8 de maio de 1979; e dá outras providências. Diário Oficial (da República Federativa do Brasil), de 09 de outubro de 2008. Brasilia, Brasil: República Federativa do Brasil.

Eskes, C., Sa-Rocha, V. M., Nunes, J. et al. (2009). Proposal for a Brazilian Centre on Alternative Test Mrthodes. *ALTEX 26*, 265-268.

ICCVAM (2008). *About the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)*. Research Triangle Park, NC, USA: Interagency Coordinating Committee on the Validation of Alternative Methods. Available at: http://iccvam.niehs.nih.gov/about/about_ICCVAM.htm (Accessed 06.11.08).

Kojima, H. (2007). JaCVAM: an organization supporting the validation and peer review of new alternatives to animal testing. *AATEX 14, Spec. Issue*, 483-485.

Machado, C. J. S., Filipecki, A. T. P. and Teixeira, M. (2009). Current Brazilian law on animal experiments. *Science 324*, 1643-1644.

Marques, R. G., Morales, M. M. and Petroianu, A. (2009). Brazilian law for scientific use of animals. *Acta Cirurg. Brasil. 24*, 69-74.

OECD (2005). *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment, Number 34*. ENV/JM/MONO(2005)14, 96pp. Paris, France: Organisation for Economic Co-operation and Development.

Ohno, Y. (2005a). Significance of Japanese Center for the Validation of Alternative Methods (JaCVAM) establishment. *AATEX 11*,1-3.

Ohno, Y. (2005b). Organization and mission of JaCVAM (Presentation). Available at: http://ntp.niehs.nih.gov/files/JaCVAM_SACATM.pdf (Accessed 06.11.08).

Presgrave, O. A. F. (2008). The need of the establishment of a Brazilian Centre for the Validation of Alternative Methods (BraCVAM). *ATLA 36*, 705-708.

Presgrave, O. A. F. and Boghal, N. (2005). EMALT: A Brazilian Meeting on Alternative Methods to Animal Use for Regulatory Purposes. *ATLA 33*, 670-672.

Teixeira, M. O., Machado, C. S. and Filipecki, A. T. (2009). A dinâmica da organização da pesquisa em biomedicina no Brasil: anatomia de uma experiência recente na Fundação Oswaldo Cruz. *R. Eletr. Com. Inf. Inov. Saúde 3*, 47-72.

ZEBET (2008). *About the Centre for Documentation and Evaluation of Alternatives to Animal Experiments (ZEBET)*. BfR – Federal Institute for Risk Assessment. Available at: http://www.bfr.bund.de/cd/1591 (Accessed 06.11.08).

Correspondence to
Octavio Presgrave
Instituto Nacional de Controle de Qualidade em Saude
Avenida Brasil, 4365 Manguinhos
21045-900, Rio de Janeiro, RJ
Brasil
e-mail: octavio.presgrave@incqs.fiocruz.br

A National Center for Animal Alternatives in India: The Mahatma Gandhi-Doerenkamp Center for Alternatives to the Use of Animals in Life Science Education (MGDC)

Mohammad A. Akbarsha[1], Shiranee Pereira[2] and Franz P. Gruber[3]

[1]Bharathidasan University, Tiruchirappalli, India, [2]People for Animals, Chennai, India, [3]Doerenkamp-Zbinden Foundation, Kuesnacht ZH, Switzerland

Summary

The Mahatma Gandhi-Doerenkamp Center for Alternatives to the Use of Animals in Life Science Education (MGDC) has been established as a national centre for alternatives in India at Bharathidasan University, Tiruchirappalli, Tamil Nadu, a renowned university under the University Grants Commission of the Government of India. The mandate of the centre is to introduce the Gandhian Philosophy of 'Ahimsa' or 'non-violence' into the teaching /research of life sciences. The centre was established with the generous financial support of the Doerenkamp-Zbinden Foundation, Switzerland, in conjunction with the establishment of the 'Gandhi-Gruber-Doerenkamp Chair' for Alternatives in Biomedical Science Education and in vitro Toxicology. The centre was established in the knowledge that promoting humane science is an imperative scientific, legal, psycho-social, ecological and economic need of the hour.

The MGDC will strive to create a strong positive presence of alternatives to the use of animals, thereby promoting quality and excellence in life science education, research and testing by way of continuous training programmes, an alternatives knowledge bank, library and certificate / diploma / post-graduate diploma programmes in animal alternatives and in vitro toxicology testing. The MGDC will also bring together stakeholders in the 3Rs – academia, scientific community, industry, government and animal welfare personnel from national / international levels – to raise the awareness / facilitate the exchange of information / ideas on alternatives to translate the vision of 3Rs into policy and curricular changes in India as relevant to education and research. The MGDC will help by way of funding research and development of environmentally friendly pedagogical tools and in vitro alternative methods for life science teaching and research. The approach will be to encourage the use of e-tools, help establish virtual learning centres, and to establish an in-house state-of-the-art cell culture laboratory for training in non-animal methods of research and product testing. The centre will be essentially a service provider in respect to non-animal methods in learning, research and testing. It is a joint venture of the Doerenkamp Zbinden Foundation, Bharathidasan University and People for Animals (Chennai), India.

Keywords: Mahatma Gandhi, life science, biomedical education, India, alternative methods, e-learning, Ahimsa

1 History

Very few men have done so much good for their fellow creatures, be it human kind or animal kind or enriching philosophy, as did Mohandas Karamchand Gandhi, who has been called the "Mahatma" (meaning the "great soul") in India. His greatest philosophy was that of "Ahimsa" or "non-violence". He said "non-violence is not a cloistered virtue to be practiced by the individual for his peace and final salvation, but it is a rule of conduct for society if it is to live consistently with human dignity and make progress towards the attainment of peace for which it has been yearning for ages past". He went on to say, "We should not inflict cruelty on even the meanest of creatures. I also will have to answer for this in the court of the Almighty."

Only few single persons have contributed so much wealth to the protection of animals as Hildegard Doerenkamp from Germany. Ms. Doerenkamp, together with the late Gerhard Zbinden, founded the Doerenkamp-Zbinden Foundation.

To honour their greatness, the Mahatma Gandhi-Doerenkamp Center (MGDC) carries the names of these two towering personalities in order to propagate the message "do not kill animals" in education, research and testing. The centre has been established at the Bharathidasan University, Tiruchirappalli, India, which is named after the great revolutionary Indian poet, Bharathidasan, who vowed "to create a brave new world". The university's affairs are now guided by M. Ponnavaikko, the vice-chancellor, according to whom "God is an invisible distributed energy; he resides in every human being, plant and animal; the human be-

ings are endowed with the sixth sense, which should be used in realising God; you will realise God only when you love, respect and care for His creations".

It was a red letter day in the history of the University and the Doerenkamp-Zbinden Foundation, Switzerland, when these two parties entered into a Memorandum of Understanding to establish the "Mahatma Gandhi-Doerenkamp Center for Alternatives to Use of Animals in Life Science Education", on 13th July 2009. The centre was launched on 15th July, 2009.

If William Russell and Rex Burch were the ones to revolutionise the approach to animal experiments with their book *The Principles of Humane Experimental Technique*, published in 1959, wherein they demanded a humane approach to animal experimentation and introduced the 3Rs concept (replacement, reduction and refinement), which has come to be known as the concept of "alternatives", it was Jennifer Graham, a brave 15-year-old girl from California, who in 1987 refused to dissect an animal and sued her school district seeking an alternative study option, whereupon the state of California, USA, granted that right to all high school students, to be followed by other States in the US and later other countries as well, who revolutionised animal dissection and vivisection.

While a humane approach to animals in education, research and testing has been taken up seriously in the developed countries, the developing countries are yet to fall in line, although they were also signatories of the *Three Rs Declaration of Bologna*, which was adopted in 1999 by the Third World Congress on Alternatives and Animal Use in the Life Sciences and strongly endorsed and reaffirmed the principles of the 3Rs. The situation in India, the country with the second largest population in the world, has been dismal.

In 1996, both the Central Board of Secondary Education and the Indian Council of Secondary Education stopped the use of animals in teaching life sciences in secondary schools in India. However, with regard to higher education and research there was an awakening only after the turn of the century when organisations like People *for* Animals, India (PFA), I-CARE and the CPCSEA (Committee For The Purpose Of Control And Supervision Of Experiments On Animals, Govt. of India), began a campaign for non-animal methods of teaching and learning in life sciences.

In February 2003, the CPCSEA sensitised the teaching and scientific community by conducting a 3-day national symposium on alternatives in New Delhi, India. Over 300 participants attended the same and renowned scientists from across the world were invited to speak. For the first time, the science of alternatives was officially introduced to the Indian scientific community. International speakers from Johns Hopkins University, USA, the Netherlands Center for Alternatives, FRAME, UK, HSUS, USA, InterNICHE and the Doerenkamp-Zbinden Foundation were invited. Their efforts came to be rewarded, since these organisations also enlisted in the service of academics and universities across India, including the first author, M. C. Sathyanarayana from AVC College, Mannampandal, Mayiladuthurai, a college affiliated to Bharathidasan University, Ramakrishna, Professor of Veterinary Science from Chennai, R. Raveendran, a pharmacologist from JIPMER, Pondicherry, Dr. Syed Ziaur Rahman, another pharmacologist from Aligarh Muslim University, and others.

More importantly, they targeted the teachers, those who are the key role players in the academic curricular decisions, to bring about change. A workshop was conducted on the 25th Sep 2001 at AVC College, Mannampandal, Mayiladuthurai, Tamil Nadu, organised by M. C. Sathyanarayana, where the first author delivered the key note address. The participants, all college teachers, were introduced to the non-animal methods of teaching and learning anatomy in zoology class. From then on, many programs were conducted in different places throughout the country, but they were all piece meal basis, without a clear-cut direction.

In 2004, PFA in collaboration with InterNICHE and WSPA, UK organised one-day workshops in alternatives in education in 9 cities across India. The 'First Indian Congress On Alternatives To The Use Of Animals In Research, Education and Testing' was a national congress also conducted in Chennai by I-CARE in 2007, which attracted over 400 participants and the best international platforms of alternatives to India. In 2006, on the initiative of I-CARE and supported by the first author and the model curriculum of Bharathidasan University and the University Grants Commission, New Delhi, the Regulatory Authority of Higher Education in India, an epoch-making letter was sent to all universities in India, directing them to use alternatives and requiring the curtailment of use of animals in zoology teaching and learning. But little was done thereafter, since the efforts were made only by unorganised sectors and free-lancers.

The first author and his team approached the change from a different perspective as well. As senior teachers and researchers of zoology, they were members of the Curriculum Boards of several universities. Taking advantage of this position, they endeavoured to change the zoology curriculum such that animal dissection as an aspect of animal anatomy laboratory exercise was greatly reduced and even dropped in some universities / programmes.

The most exciting outcome was from Bharathidasan University, where all major dissections were dropped from the curriculum for undergraduate as well as post-graduate programmes, and in their place learning of animal anatomy using CD-Roms was introduced for the first time in the country. This came to be called the "Bharathidasan University Model" (Akbarsha, 2007). In 2003, CPCSEA again worked with the Pharmacy Council of India and brought about the decision that where alternative methods for pharmacological testing are available, *in vivo* testing protocols need not be practiced. Thus, the roles of I-CARE, People *for* Animals, *Peta* India, and InterNICHE are commendable. Yet, the vastness of the country, the variety in the higher education and the heterogeneity of the religious, linguistic and cultural heritage of the people made the task enormous and difficult for these organisations to handle.

2 The Mahatma Gandhi-Doerenkamp Center

At this time, there was an initiative from the Chennai chapter of PFA to upscale the campaign for non-animal methods in life science and biomedical science education in India. Since Bharathidasan University had already made its mark, spearheaded by the first author and M. C. Sathyanarayana, it was proposed that a National Center for Alternatives be established at this university, with the first author at the helm of affairs. Before funding agencies in India could be approached, the Doerenkamp-Zbinden Foundation (DZF) heard of the campaign and offered to participate in the venture, deciding for the first time to expand its activities outside Europe and the USA. It was proposed that a National Center for India be established in a tripartite collaboration between the DZF, Bharathidasan University, and PFA, Chennai, wherein the DZF bears the entire cost, including an academic chair and a building proposed in a budget, PFA renders moral and counselling support and Bharathidasan University houses and runs the MGDC. This initiative culminated in the signing of the Memorandum of Understanding on July 13th, 2009.

The centre is named after Mahatma Gandhi and Ms. Hildegard Doerenkamp. Mahatma Gandhi was the great leader of India during the most gruesome period of Indian history. Central to Mahatma Gandhi's vision was an impassioned conviction that at the heart of all life there is 'Truth' which sustains all creation; a 'Truth' which demands a personal response from each individual. He saw 'Truth' as a truth present in every person. In particular, he held non-violence as a basic tenet of this 'truth', a positive force that can bring about fundamental change at all levels. For Gandhi 'non-violence' was the discovery of a new kind of power. It is a well known fact that Gandhi not only played a major role in India achieving its independence but also taught a philosophy which has universal applicability. The core of that philosophy is the search for truth through non-violence – "Ahimsa". Gandhi taught respect for animals as well as humans, a non-exploitative relationship with the environment, the elimination of poverty, the limitation of personal wealth and possessions, and non-violence applied at all levels of relationships, be it man to man, man to animal or man to environment. According to him, "The greatness of a nation and its moral progress can be judged by the way its animals are treated". He said "I hold that the more helpless a creature, the more entitled it is to protection by man from the cruelty of man", and "I abhor vivisection with my whole soul. All the scientific discoveries stained with innocent blood I count as of no consequence."

Ms. Hildegard Doerenkamp dedicated her entire wealth towards the cause of animal protection, and, together with Gerhard Zbinden, founded the DZF. This foundation began by extending support to the discovery of alternatives and conferring awards for outstanding work on alternatives. Later, it started to establish academic chairs for alternative methods. It has Chairs at Johns Hopkins University, USA, University of Geneva, Switzerland, University of Konstanz and University of Erlangen, Germany, and Utrecht University, The Netherlands. The Foundation is also member and main sponsor of the society ALTEX Edition, which publishes the journal on alternatives to animal experimentation and testing, ALTEX.

Mohammad A. Akbarsha, a teacher of zoology and animal science for more than three decades and a scientist who uses *in vitro* tools, has been declared Director of the MGDC and also holder of the Gandhi-Gruber-Doerenkamp Chair for Alternatives in Life Science and *In Vitro* Toxicology.

The mandates of the MGDC are:

1. The MGDC will endeavour to advance the concept of "humane science" and implement the concept of the 3Rs in education, research and testing in accordance with Indian legislation, the international 'Declaration of Bologna' and international advances in the science of alternatives.
2. The MGDC will work to create a strong and positive presence for alternatives to the use of animals in experimentation / testing in India and pro-actively work to blend life science education with the Gandhian philosophy of non-violence.
3. The MGDC will evolve a national programme for humane education in teaching and research.

The MGDC will work to:

1. Develop a national humane education programme for all universities/ colleges/ national research institutes as part of their life science curriculum/ research, which will seek to blend the Gandhian philosophy of non-violence with life science education/ research
2. Develop a strategy for the implementation of the 3Rs in academic and research/testing laboratories
3. Conduct courses on Humane Science and Alternatives in the Use of Animals in Education and Research in affiliation with other renowned universities like the Oxford University Centre of Animal Ethics, UK and CAAT, Johns Hopkins University, USA, etc., for both Indian and international students
4. Support by way of funds and expertise high-quality research that advances the 3Rs for development of pedagogic tools, computer modelling for teaching, drug development, basic bio-medical research, product testing, etc.
5. Provide expertise and guidance on the 3Rs and laboratory animal welfare to the teaching/ scientific community by developing a range of resources, including guidelines and training material (e.g. CDs), organising working groups, workshops, symposia, etc.
6. Liaise with national education councils, like MCI, VCI, PCI, UGC, AICTE, etc., and state education departments for curricular developments to promote the use and knowledge of alternatives
7. Liaise with regulatory bodies for the acceptance of alternative methods in product testing
8. Establish a state-of-the-art tissue culture laboratory and library of alternatives to help train scientists in the use of alternatives. The former could also help generate revenue for the MGDC by way of product testing on a payment basis

9. Help create virtual learning centres at universities
10. Organise national level training workshops, seminars, national congresses, etc., to propagate *in vitro*/*in silico* methods and other alternatives with international expertise / collaboration and to train Institutional Ethics Committee members.

Following are the mission goals:

1. To establish a library of alternatives
2. To provide for lending of alternatives
3. To offer training in use of alternatives
4. Thus, to motivate the teachers, not only to introduce alternatives in teaching and research but to work on the Curriculum Boards to change the course content so as to replace animals in dissection, experiments and testing with appropriate alternative modalities
5. To conduct research to develop new alternatives
6. To work for application of alternatives in life science and biomedical science research
7. To conduct teaching programmes in "Ahimsa and Alternatives" and "*In Vitro* Toxicology"

The following facilities are being built:

- A repository of dissection CDs, videos, mannekins, models, etc.
- A state-of-the-art computational lab
- An *in vitro* testing facility and a cell line bank
- Bio-informatics tools such as predictive toxicology (Q)SAR, Read Alone, ebTrack, etc., for *in silico* analysis

To reach these goals the MGDC intends to:

- conduct workshops at identified places to train university and college teachers to use dissection alternatives;
- lend dissection alternatives to institutions to encourage change to the newer pedagogy of teaching animal anatomy;
- train teachers and researchers in *in vitro* and *in silico* alternatives;
- organise national and international seminars and conferences to popularise the concept of alternatives.

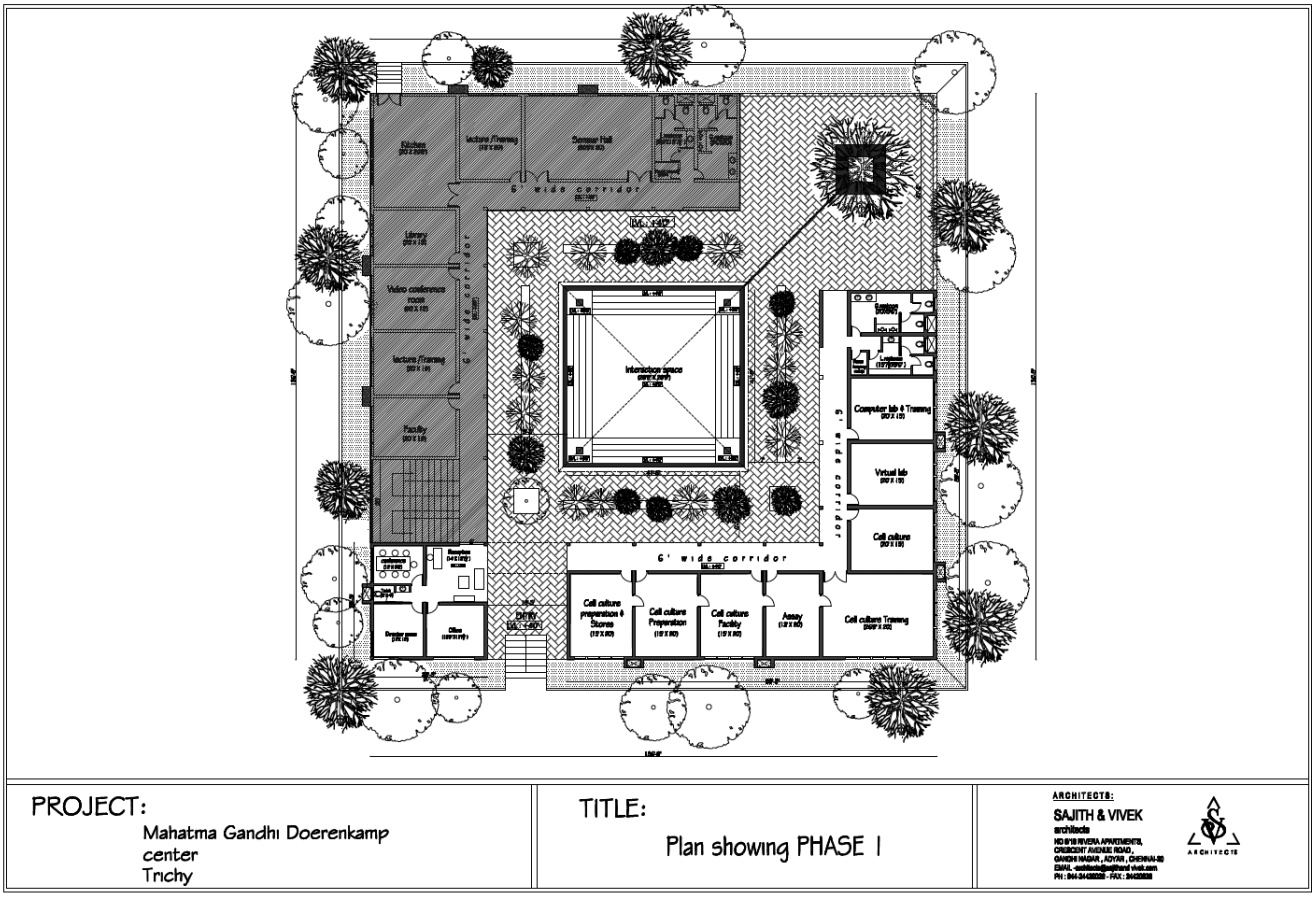


Fig. 1

The task is enormous. As already said, India is a vast country with a huge population.

Area:	3.3 million sq km
Number of States:	28
Number of Union Territories:	7
Population:	1,186.2 million
Ethnic Groups:	Indo-Aryan 72 %, Dravidian 25 % Mongoloid and others 3 %
Number of languages spoken:	about 40
Number of Central Universities:	39
Number of State Universities:	131
Number of Deemed Universities:	127
Number of Colleges:	6.289
Number of Science Colleges:	1.868
Number of Medical Colleges:	274

Yet, the MGDC is optimistic. The goal of non-animal methods in teaching, research and testing as a national policy and practice is expected to be reached soon.

With means of the DZF the MGDC will be built in the next month. The first step is fully paid by DZF, and for the second step additional sponsors are sought (see fig. 1 and fig. 2).

3 The MGDC Board

To better fulfil its obligations to bring modern teaching methods to Indian bio-medical education a governing council and an advisory board have been elected by the chancellor of the Bharathidasan University. The members meet at least once a year and give the director strategic impulses for his way forward. One of the outstanding board members is David Dewhurst, Professor for e-learning methods at the University of Edinburgh, UK, well known as a specialist in modern biomedical education (Dewhurst, 2004; Gruber and Dewhurst, 2004).

Governing council:

1. H'ble Vice-Chancellor, Bharathidasan University, Tiruchirappalli –Chairman
2. Registrar, Bharathidasan University, Tiruchirappalli

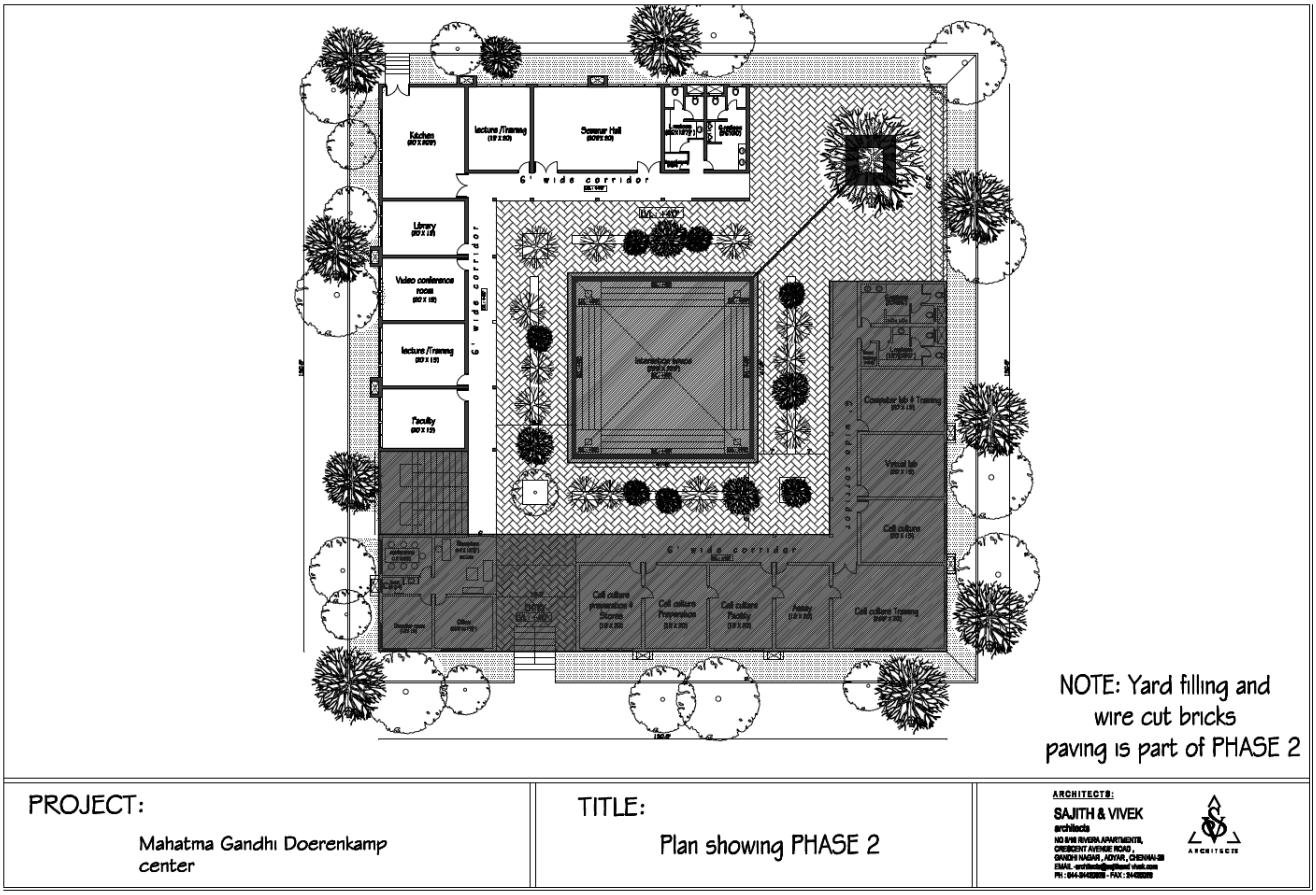


Fig. 2

3. Director, Mahatma Gandhi-Doerenkamp Center & Gandhi-Gruber-Doerenkamp Chair, Bharathidasan University, Tiruchirappalli – Secretary.
4. Dr. David Dewhurst, Professor of e-Learning, Director of Educational Information Services, College of Medicine & Veterinary Medicine, University of Edinburgh, UK – (DZF Nominee).
5. Dr. (Mrs). Shiranee Pereira, Chennai - Nominee of PFA
6. Dr. N. Tajuddin, Bioscientist Member in Syndicate, Bharathidasan University, Tiruchirappalli-
7. Dr. T. Madhan Mohan, Director, Department of Biotechnology, Govt. of India, New Delhi – Member.
8. Dr. R. S. Sharma, Deputy Director General, ICMR, New Delhi
9. Dr. B. Manivannan, Asst General Manager, Regulatory Office, A-27/B1 Extn, Mohan Corporative Industrial Estate, Madhura Road, New Delhi
10. Dr. Oommen V. Oommen, Professor of Zoology & Coordinator, UGC-SAP, University of Kerala, Thiruvananthapuram
11. Dr. S. Vincent, Member Secretary, Tamil Nadu State Council for Science and Technology, Chennai

Advisory Board:

1. H'ble Vice-Chancellor, Bharathidasan University, Tiruchirappalli – Chairman
2. Registrar, Bharathidasan University, Tiruchirappalli
3. Director, Mahatma Gandhi-Doerenkamp Center & Gandhi-Gruber-Doerenkamp Chair, Bharathidasan University, Tiruchirappalli – 620 024 - Secretary.
4. Dr. David Dewhurst, Professor of e-Learning, Director of Educational Information Services. College of Medicine & Veterinary Medicine, University of Edinburgh, UK – (DZF Nominee).
5. Dr. (Mrs). Shiranee Pereira, Chennai, Nominee of PFA,
6. Dr. Massimo Tettamanti, I-CARE, 33, Athivakkam Village, Red Hills, Chennai
7. Dr. R. Raveendran, Professor, Dept. of Pharmacology, JIPMER, Pondicherry
8. Dr. Zia-ur-Rahuman, Reader in Pharmacology, Aligarh Muslim University, Aligarh
9. Dr. Oommen. V. Oommen, Professor of Zoology & Coordinator, UGC-SAP, University of Kerala, Thiruvananthapuram
10. Dr. V. R. Muthukarruppan, Aravind Eye Hospital, Madurai
11. Dr. Michael Aruldhas, Professor & Head, Dept of Endocrinology, ALPGIBMS, Chennai
12. Dr. K. Dharmalingam, Professor & Head, School of Biotechnology, Madurai Kamaraj University, Madurai
13. Dr. Gopinath Ganapathy, Professor & Head, Dept of Computer Science, Bharathidasan University, Tiruchirappalli
14. Dr. M. Krishnan, Professor & Head, Dept of Environmental Biotechnology, Bharathidasan University, Tiruchirappalli
15. Dr. S. Parthasarathy, Reader & Head, Dept. of Bioinformatics, Bharathidasan University, Tiruchirappalli
16. Dr. P. Ilango, Professor & Head, Dept of Social Work, Bharathidasan University, Tiruchirappalli – 620 024
17. Dr. (Mrs).Radha Chellappan , Emeritus Professor, Bharathidasan University, Tiruchirappalli
18. Dr. T. Thirunalasundari, Coordinator, sixth year M.Tech Courses, Bharathidasan University, Tiruchirappalli
19. Dr. B. Kadalmani, Lecturer, Dept. of Animal Science, Bharathidasan University, Tiruchirappalli

References:

Akbarsha, M. A. (2007). Movements to curtail animal dissections in zoology curriculum: review of the Indian experience. *ALTEX 24*, 163-166.

Dewhurst, D. (2004) Computer-based alternatives to using animals in teaching physiology and pharmacology to undergraduate students. *ATLA 32, Suppl*., 517-520.

Gruber, F. P. and Dewhurst D. (2004). Alternatives to animal experimentation in biomedical education. *ALTEX 21, Suppl. 1*, 33-48.

Correspondence to:

Prof. Dr. Mohammad A. Akbarsha
Gandhi-Gruber-Doerenkamp Chair for Alternatives in Life Science Education and In vitro Toxicology & Mahatma Gandhi Doerenkamp Center (MGDC) for Alternatives to Use of Animals in Life Science Education, Bharathidasan University,
Tiruchirappalli – 620 024
India
Tel. +91 9790995854, fax: +91 431 2407045
E-mail: mgdcaua@yahoo.in

MS6: Associations

CELLTOX: The Italian Association for In Vitro Toxicology

Isabella De Angelis[1], Simonetta Gemma[1] and Francesca Caloni[2]

1Department of Environmental and Primary Prevention, Istituto Superiore di Sanità, Rome, Italy; 2Department of Veterinary Sciences and Technologies for Food Safety, Faculty of Veterinary Medicine, Università degli Studi di Milano, Italy

Summary

The principal aim of CELLTOX, the Italian Association for in vitro Toxicology founded in 1991, is the dissemination of alternative methods in toxicology. The activity is focused on the promotion of in vitro models and alternative methodologies, collaborations with private and public institutions, creation of networks with researchers from different disciplines and organisation of conferences, courses, meetings and workshops geared at spreading the 3R principles and philosophy in the scientific community. The support young scientists with fellowships and grants is one of the main goals of this association. CELLTOX has worked in collaboration with IPAM, Italian Platform for Alternative Methods, and other Italian and European organisations and is an affiliate of the European Society of Toxicology In Vitro (ESTIV).

Keywords: Celltox, Italian Association, in vitro methods, toxicology

The Italian Association for *in vitro* Toxicology, CELLTOX, was founded in 1991 and has its roots in a pre-existing but informal and enthusiastic group of scientists.

From the beginning, the Association was considered the national point of reference for researchers and public and private institutions interested and involved in the application of *in vitro* methods in toxicology. It has thus been frequently contacted to provide technical support for certain government projects and recently has worked actively with IPAM (Italian Platform for Alternative Methods) for the implementation of alternative methods in Italian legislation.

Moreover, it is closely linked to analogous European societies and affiliated with the European Society for Toxicology In Vitro (ESTIV).

Its activities and goals, still valid at present, are well described in its statutes. Its main aims are as follows:

a) to promote the use of *in vitro* experimental models and alternative methodologies in the pharmacological and toxicological field;
b) to investigate mechanisms of toxicity at the cellular and molecular level, with a special interest in cell culture models;
c) to facilitate the exchange of information and collaboration among research groups from different public and private institutions;
d) to create an information network about the ethical and practical aspects of the reduction of animal use by propagating the 3R principles and philosophy in the scientific community.

CELLTOX is managed by a board composed of eight members (and the former president). Board members, elected every three years from the group of general members, are mainly selected with the purpose of bringing in the greatest professional experience in the different fields of application of *in vitro* methods.

The current composition of the board, which represents scientists from different institutions and research groups, is reported in Table 1.

The members of the Association are mainly from the north (44%) and central parts of Italy (46%), while the south (10%) is still poorly represented. The members are quite homogeneously distributed, coming from research institutes (35%), the industrial sector (27%) and universities (15%). Students, including undergraduates, PhD students and post-docs with a reduced membership fee, represent 23% of the total members. This is an important feature of the association, as the training of young scientists has always been one of its main objectives.

In this respect, CELLTOX contributes to the diffusion of 3R concepts by helping young researchers remain updated on scientific issues by giving them the opportunity to attend international conferences and national courses with grants covering registration fees and/or travel expenses. In recent years, CELLTOX has sponsored seven grants for the participation of young members in several international meetings (AICC-CELLTOX joint meeting, 2008 ESTIV meeting) and eleven grants for the participation of young members in CELLTOX courses.

Tab. 1: CELLTOX Board

President
Isabella De Angelis (ISS[1])
Vice –President
Yula Sambuy (INRAN[2])
Treasurer
Francesca Caloni (University of Milan)
Secretary
Simonetta Gemma (ISS[1])
Councillors
Lucia Golzio(MerckSerono RBM)
Marisa Meloni (Vitroscreen)
Maria Pilar-Prieto (ECVAM[3])
Chiara Urani (University of Milan Bicocca)
Anna Zaghini (University of Bologna)
Past President
Annalaura Stammati (ISS[1])

[1] ISS – Istituto Superiore di Sanità
[2] INRAN – Istituto di Ricerca per gli Alimenti e la Nutrizione;
[3] ECVAM – European Centre for the Validation of Alternative Methods

Association activities include regularly organised meetings, practical courses and symposia which cover the principal aspects of *in vitro* toxicology research and disseminate the 3R concept.

CELLTOX has organised a cycle of workshops to focus on state-of-the-art teaching in Italian universities using alternative methods and the 3R principle, involving speakers from different institutions and associations to underline the importance of promoting alternative methods in education and providing counsel on how to integrate them into teaching. Whereas the 3R concept in research is very well known and largely applied, the use of alternative methods in education is at times limited either due to lack of information or knowledge of pedagogical advantages. The meeting has been held since 2005, starting at the University of Milan and then subsequently moving to the University of Naples, University of Bologna and University of Milan Bicocca, with the next to be held at the University of Bari.

International conferences have also been organised by the Association. Three international joint meetings were held with the Italian Association of Cell Culture (AICC) on "*In Vitro* Cytotoxicity Mechanisms" in which relevant issues on *in vitro* toxicology were presented and discussed by the participants. At the last one held in Verona in 2006, about 120 scientists and 96 contributors from several European countries attended the whole meeting, producing an excellent network for a productive exchange of ideas.

CELLTOX has organised symposia on relevant emerging topics: the first Italian symposium on nanotoxicology in collaboration with Insubria University in 2007, a round table on "Alternative methods: new prospects" in the framework of the XV National Congress of the Italian Society of Toxicology and the Satellite Symposium during the 7th World Congress on Alternatives & Animal Use in the Life Sciences "From tissue engineering to alternatives: research, discovery and development", in collaboration with ESTIV, in 2009.

Several national practical courses to introduce basic information on tissue culture and the application of toxicology have also been organised by our society recently:

1) Course on Episkin: the *in vitro* alternative model for skin irritation testing
2) Course on molecular biology techniques in the pharmaceutical industry
3) Course on cell culture in toxicology: basic techniques

All of these courses, which are included in the national training program for health workers, were attended by a large number of participants and were very well appreciated. New editions will be planned for the future.

Finally, the CELLTOX website (www.celltox.it) has, in recent years, become one of the most important tools in advertising the activities of the association and informing others about what is happening in the world of *in vitro* toxicology. It has been updated recently to make it more efficient, and a continual increase in the number of national and international contacts has been noticed.

Our hope is to further expand the Association in terms of membership and activities in the future and we expect to continue our efforts in collaboration with additional national and international entities involved in the field of alternative methods in toxicology.

Correspondence to
Isabella De Angelis MD
Istituto Superiore di Sanità
Environment and Primary Prevention Dept.
Viale Regina Elena 299
00161 Rome, Italy
e-mail: isabella.deangelis@iss.it

Tab. 2: Visits to CELLTOX website during a three month period (January-March 2009).

Visits from	total visits	average visits historically	Change	percents of visits
North America	478	532	-108	67.14
Europe	192	210	+1	26.97
Asia	41	26	+11	5.76
South America	1	-	+1	0.14

Assuring Safety without Animal Testing: Unilever's Ongoing Research Programme to Deliver Novel Ways to Assure Consumer Safety

Carl Westmoreland, Paul Carmichael, Matt Dent, Julia Fentem, Cameron MacKay, Gavin Maxwell, Camilla Pease and Fiona Reynolds

Safety & Environmental Assurance Centre (SEAC), Unilever, Bedford, UK

Summary

Assuring consumer safety without the generation of new animal data is currently a considerable challenge. However, through the application of new technologies and the further development of risk-based approaches for safety assessment, we remain confident it is ultimately achievable. For many complex, multi-organ consumer safety endpoints, the development, evaluation and application of new, non-animal approaches is hampered by a lack of biological understanding of the underlying mechanistic processes involved. The enormity of this scientific challenge should not be underestimated.

To tackle this challenge a substantial research programme was initiated by Unilever in 2004 to critically evaluate the feasibility of a new conceptual approach based upon the following key components:

1. Developing new, exposure-driven risk assessment approaches
2. Developing new biological (in vitro) and computer-based (in silico) predictive models
3. Evaluating the applicability of new technologies for generating data (e.g. "omics", informatics) and for integrating new types of data (e.g. systems approaches) for risk-based safety assessment

Our research efforts are focussed in the priority areas of skin allergy, cancer and general toxicity (including inhaled toxicity). In all of these areas, a long-term investment is essential to increase the scientific understanding of the underlying biology and molecular mechanisms that we believe will ultimately form a sound basis for novel risk assessment approaches.

Our research programme in these priority areas consists of in-house research as well as Unilever-sponsored academic research, involvement in EU-funded projects (e.g. Sens-it-iv, Carcinogenomics), participation in cross-industry collaborative research (e.g. Colipa, EPAA) and ongoing involvement with other scientific initiatives on non-animal approaches to risk assessment (e.g. UK NC3Rs, US "Human Toxicology Project" consortium).

Keywords: risk assessment, skin allergy, cancer, Colipa, NC3Rs, European Commission, EPAA

1 Introduction

Unilever's commitment to eliminating animal testing has been underpinned by our scientific research programme since the 1980's in developing and using alternatives to animal tests. Since 2004, we have invested an additional € 3 million a year towards the development of novel non-animal approaches to assure consumer safety. Unilever's conceptual framework for safety assurance is risk-based rather then hazard-based, meaning all available data on a new ingredient (including predicted levels of consumer exposure during product use) are analysed using a weight-of-evidence approach to judge the level of risk prior to generating any new data to further characterise the hazard.

Our research efforts are currently focussed on the development and evaluation of new risk assessment approaches for assuring consumer safety in the areas of skin allergy (Maxwell et al., 2008) and cancer, and consideration of how risk-based approaches and the application of new models and technologies could be applied in the area of general toxicity (where we have used inhaled toxicity as a specific case study).

The scientific and technical challenges associated with assuring consumer safety without any animal testing in the areas outlined above are enormous and it is clear that no single research group or company will achieve these goals alone. For this reason, in order to develop alternative approaches for consumer safety, Unilever has its own in-house research programme and in addition works in partnership with a number of external groups. These partnerships include sponsoring research with academic institutions, investigating new approaches with contract research organisations, initiating bespoke research with biotechnology companies, and consultancies with key experts. In addition to our internal research programme we are also involved in EU-funded projects, e.g. Sens-it-iv, Carcinogenomics, we participate in cross-industry collaborative research, e.g. The European Cosmetics Association (Colipa) and the European Partnership on Alternative Approaches to Animal Testing (EPAA), and have ongoing involvement with other scientific initiatives on non-animal approaches to risk assessment, e.g. UK National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), US "Human Toxicology Project" consortium looking to follow up the 2007 report from the US National Research Council (NRC, 2007).

2 Skin allergy

Following the principles of the conceptual framework (Fentem et al., 2004), we have been exploring the feasibility of delivering consumer safety risk assessment decisions for skin allergy that do not rely on data generated in animals (Maxwell et al., 2008). A key element of this work is to improve our current estimates of dermal exposure to ingredients for which we have detailed information on both formulation effects and consumer habits and use patterns. A new experimental approach based on *ex vivo* human skin has been investigated to determine skin compartmental concentrations and the delivery kinetics of a chemical *in vitro* (Pendlington et al., 2008). Techniques like this aim to establish a relationship between the exposure on the skin and the chemical bioavailability at the target site assumed to be the epidermis and dermis. However, the true bioavailability of free chemical in the skin tissue is also influenced by other parameters (e.g. skin metabolism, tissue adsorption and clearance mechanisms) and consequently ongoing research in these areas should ultimately provide more valuable information for novel risk assessments.

To determine whether a chemical has the potential to induce skin sensitisation many groups are currently working on non-animal predictive models to encompass the events which are considered to be key to the induction of skin sensitisation (including large programmes of work at Colipa (Aeby et al., 2008) and within Sens-it-iv). To test and explore the relative contributions of individual biological pathways thought to be key to the induction of skin sensitisation, we developed an *in silico* mathematical model (the "Skin Sensitisation PhysioLab®" [SSP] platform; Maxwell and MacKay, 2008) in collaboration with Entelos® Inc. The aim of this project was to determine the key biological pathways that drive the sensitisation response by mechanistic modelling of the biology that has been reported as relevant in the scientific literature. The technique provided a systematic approach for the identification of key pathways as well as the identification of knowledge gaps. This information has been used to focus our *in vitro* assay development on the pathways of importance (e.g. inflammatory signals in the epidermis) and has motivated some of our fundamental research studies, in particular, investigation of functional proteomics of skin proteins modified by sensitising chemicals.

Jowsey et al. (2006) hypothesised that no single non-animal approach could be envisaged to generate sensitiser potency information and proposed that multiple forms of non-animal data would be required for this purpose. Based on this hypothesis and our evaluation of the published skin sensitisation literature (by using the SSP platform), our current hypothesis is that the integration of some or all of the following categories of non-animal information, in the context of human exposure, should yield a new measure of skin sensitiser potency: Chemical reactivity; epidermal disposition (or bioavailability); epidermal inflammation; dendritic cell activation; T-cell proliferation. For example, it is generally understood that any chemical (or metabolite derived from it) must form a stable (covalent) adduct with protein in the skin in order to stimulate an immune response (Divkovic et al., 2005). Consequently, the covalent modification of a protein by a reactive chemical (haptenation) is considered to be a key step in the induction of skin sensitisation. Several *in chemico* assays for measuring the extent and nature of chemical reactions with model peptides are being developed, underpinned by this hypothesis (e.g. Aeby et al., 2008; Gerberick et al., 2007; Natsch et al., 2007; Kato et al., 2003). For the purposes of deriving the maximum qualitative and quantitative information on the reactivity of a chemical with peptides for a non-animal skin allergy RA framework, we have developed an *in chemico* peptide reactivity profiling assay, which uses a panel of six single-nucleophile peptides (generic sequence AcFAAXAA, where X = Cys, Lys, Tyr, His or Arg, with H_2N-FAAAAA representing the N-terminal nucleophile) with the aim of determining the reactivity profile of a chemical with a high level of confidence. We have now standardised this approach and tested 36 chemicals; varied patterns of reactivity reaction mechanisms which are not always theoretically foreseen have been observed (Aleksic et al., 2009).

Due to the increasing complexity of datasets from this and other *in vitro* approaches, statistical tools (including principle component analysis, partial least squares and generalised linear models) are being used to analyse data and establish the predictive capacity of each assay alone and in combination. A network approach is also being used to integrate the data in a probabilistic and biologically relevant manner by drawing on the pathway structure modelled in the SSP. The aim is to determine the feasibility of such approaches for providing hazard data for risk assessment. Currently animal data such as local lymph node assay data is used to evaluate non-animal approaches. However, such approaches will need to be evaluated within a risk assessment and consequently more emphasis will need to be placed on human clinical experience of skin sensitisation to ensure that

any extrapolation of data from non-animal predictive models is appropriate. By maintaining the emphasis of new approaches on the human relevance we can move beyond animal replacement and towards better risk assessment methods in skin allergy.

3 Cancer

Like skin allergy, the prevention of cancer from the use of consumer products represents an extremely important safety endpoint. Past strategies have relied heavily on results from *in vitro* tests (genetic toxicology tests) being confirmed by "definitive" animal studies (genetic toxicology and carcinogenicity tests). We believe that a new non-animal strategy can be developed that is more informative and ensures safety to the consumer.

Since March 2009, as a consequence of the 7th Amendment to the EU Cosmetics Directive (EC, 2003) there has been a ban on the genetic toxicology-based animal testing of chemical ingredients intended for use in cosmetic products in Europe. In practical terms this has meant the cessation of several genetic toxicology tests, including the widely used bone marrow micronucleus assay in rodents.

In vitro-only genetic toxicology assay strategies have a high irrelevant positive rate (i.e. positive results will be obtained for chemicals that are not carcinogenic (Kirkland et al., 2005)), and many common food-based biochemicals can be erroneously rejected (e.g. flavonoids) if *in vitro* regulatory tests are employed alone. This is because of the inherent nature of the current assays, and much on-going research is focussed on identifying approaches to increase the specificity of currently available *in vitro* genetic toxicology tests (e.g. a large research programme at Colipa). Because these tests are used purely in a hazard identification mode – the label of "genotoxicity" indicated by the current tests necessitates the rejection of that chemical if no follow-up testing is conducted. However, greater evidence and wider acceptance of the existence of thresholds for genotoxic events, determined in in vitro systems, is emerging (Carmichael et al., 2009). We believe this may provide a way forward for the risk assessment of new chemicals and we are utilising high-throughput methodologies, such as automated micronucleus scoring (Diaz et al., 2007), to provide the data necessary for low dose determinations of thresholds, in standard and newly engineered cell lines.

The characterisation of a material as "low-dose thresholded" will, however, require adequate understanding of the molecular mechanism of action of carcinogens. "Omics" technologies offer real hope in this regard. Successes with transcriptomics and metabolomics have shown discrimination between chemicals with probable thresholded characteristics, based on mechanistic understandings (e.g. the activation of DNA repair pathways, changes in the cell cycle and oxidative/metabolic stress). Major programmes of research, in the EU and US (e.g. Carcinogenomics and ILSI-HESI programme) will continue this trend.

Novel insights are being generated that will be capable of informing a risk-based approach and, through collaborative work, we are investigating several other new technologies to increase our understanding of the complex interactions that occur in biological systems in response to carcinogenic ingredients. For example, technology from the field of biophysics (i.e. infra-red micro-spectroscopy is proving to be valuable in mapping and understanding the transformation of Syrian Hamster Embryo cells in culture, in response to chemical carcinogen exposure (Walsh et al., 2009)). Furthermore, work at MIT, Boston is providing new ways to interpret the complex pathways and interactions involved in eukaryotic responses to carcinogenic chemicals using specific gene-deleted libraries teamed with systems biology tools such as Cytoscape.

The challenge ahead will be to integrate these data to allow risk assessment to be performed for new chemicals in consumer products under the conditions of use. The application of systems biology approaches to anchor the *in vitro* measurements to relevant biomarkers and pathology pathways will be key in this regard and we are conducting research with partners at Barts and The London School of Medicine, UK in order to provide the much needed, greater molecular understanding of the processes that lead to human skin tumours.

4 General toxicity

The assurance of consumer safety for novel ingredients without the generation of new animal data still remains a considerable scientific challenge, but in light of the progress described above towards new risk assessment frameworks for skin allergy and cancer we remain convinced that this is ultimately achievable. A major challenge for the future is how risk assessments for systemic toxicity may ultimately be performed in the absence of animal testing. We are currently working on understanding the work necessary in each of the following areas:

- Developing new, exposure-driven risk assessment approaches
- Developing new biological (*in vitro*) and computer-based (*in silico*) predictive models
- Evaluating the applicability of new technologies for generating data (e.g. "omics", informatics) and for integrating new types of data (e.g. systems approaches) for risk-based safety assessment.

Key to progressing the development of new risk assessment strategies is the identification of the adverse health effects (and underlying mechanistic understanding of these health effects) that we are aiming to prevent in our consumers (a fundamental reason for the current progress that is being made in the areas of skin allergy and cancer). We have used a case study of inhalation toxicology to begin to understand how non-animal based approaches may be integrated for risk assessment purposes. This research includes (i) the development of a new exposure-based waiving approach for certain chemical classes (Carthew et al., 2009), (ii) exploration of the molecular understanding of mechanistic divergence between adverse and non-adverse effects (e.g. Carthew et al., 2006) and (iii) development of cellular models that will allow adverse lung effects to be predicted (e.g. Grainger et al., 2009)

New technologies are rapidly emerging that could offer the potential for ground-breaking opportunities in developing novel

ways to assess consumer safety (Fentem and Westmoreland, 2007). Advances in new technologies, particularly during the past decade, have opened up new avenues to the possibility that non-animal approaches could be achievable at some point in the future for human adverse effects more complex than local skin and eye irritation. For example, we have begun to explore potentially relevant new technologies in the areas of human tissue engineering (e.g. artificial lymph nodes), relevant cell-based approaches, "omics" technologies (transcriptomics, proteomics and metabolomics (which we have investigated in the context of skin inflammation), bioinformatics, advanced analytical methods, computer modelling (including systems biology) and new data interpretation/integration algorithms. Combined application of these tools and technologies in complementary and integrated ways should provide an enhanced scientific and increasingly more mechanistic basis for consumer safety assessment as well as enabling us to move away from animal testing to more human relevant analyses. The management and analysis of the vast amounts of data generated from "omics" experiments represents a major logistical and technical informatics challenge. Analysis can be extremely time-consuming and requires specialist bioinformatics capabilities. We developed a new informatics platform to support the analysis and interpretation of these experimental data in an integrated manner. Working with the European Bioinformatics Institute, in-house databases have been built and federated to Web-based databases for adding further information about the biomolecules identified in our experiments. Working with the University of California San Diego, the open-source software *Cytoscape* (Shannon et al., 2003) has been applied to integrate the data generated with human biological network and pathway data.

A wider vision and a shared strategic view of incorporating data from new models and technologies into potential novel frameworks for human safety testing has been articulated in a 2007 report from the US National Research Council (NRC), commissioned by the US Environmental Protection Agency. In its summary, the report states that: "Advances in toxicogenomics, bioinformatics, systems biology, epigenetics, and computational toxicology could transform toxicity testing from a system based on whole-animal testing to one founded primarily on *in vitro* methods that evaluate changes in biologic processes using cells, cell lines, or cellular components, preferably of human origin" (NRC, 2007). Recently, the US "Human Toxicology Project" Consortium has been established to facilitate the global implementation of this NRC vision on toxicity testing for the 21st century.

Likewise within Europe, the EPAA have considered what approaches may be needed to address repeat-dose toxicity without animals, and a recent report in 2008 concluded "The time is right to harness more effectively the very substantial achievements that have been witnessed in biology and chemistry during the last 10 years. Many seminal discoveries and technological advances have the potential to impact substantially on the development of alternative approaches. Funding at the nexus of the disciplines of toxicology, biology chemistry and mathematics was recommended" (Anon, 2008). In 2009, the European Commission launched a Call for Proposals including funding from Colipa for the development of a strategy towards alternatives to safety tests using animals in the area of repeat dose systemic toxicity. This includes research in the areas of advanced organ-simulating devices, novel methods to achieve functional differentiation of human-based target cells *in vitro*, optimisation of computational modelling and estimation techniques and integrated data analysis (EC, 2009).

A key area of new technology beginning to be investigated in the context of human safety is biological tissue engineered models derived from human primary cells, cell lines and stem cells (Westmoreland and Holmes, 2009). Organotypic models are very much in their infancy of development and, before any study of more complex human adverse effects in relation to chemical insult can begin, we must understand for what steps of human biological adversity we are trying to build the models. It may be possible to deconstruct and model some aspects of key multi-component pathways of complex adverse effects, e.g. using similar principles most commonly found in other areas of engineering. The analogy of mechanical engineering, however, is built upon an existing understanding of the component parts of the machine and its mechanism. However, we have a conundrum in human safety in that the first complex problem is identifying what the key pathways or components are for any interpretable adverse effect in order to be able to model it in a systems approach. We know the organs of the human body, but we do not often know the mechanisms of adversity in response to chemical insult. We cannot use randomly selected *in vitro* models to begin generating data as they may not be at all relevant to human effect. This aligns well with the NRC vision, which calls for a shift to a toxicity pathway-based paradigm for chemical risk assessment that holds great promise to be quicker and more predictive of human outcomes, including dose response modelling utilising computational systems biology models of the circuitry underlying each toxicity pathway (Andersen and Krewski, 2009).

The results from our research programme to date confirm our belief that an essential aspect of future success will be to involve multidisciplinary teams from all aspects of relevant new technology early in defining the strategy for addressing the best practical ways forward for exploring novel ways to assure consumer safety for complex safety endpoints.

References

Aeby, P., Ashikaga, T., Diembeck, W. et al. (2008). The COLIPA strategy for the development of in vitro alternatives: Skin sensitisation. *AATEX 14*, 375-379.

Aleksic, M., Thain, E., Roger, D. et al. (2009). Reactivity profiling: Covalent modification of single nucleophile peptides for skin sensitization risk assessment. *Tox. Sci. 108*, 401-411.

Andersen, M. E. and Krewski, D. B. (2009). Toxicity testing in the 21st Century; bringing the vision to life. *Toxicol. Sci. 107*, 324-330

Anon (2008). New perspectives on safety: Workshop report. http://ec.europa.eu/enterprise/epaa/3_activities/3_3_research/new_perspectives.pdf

Carmichael, P., Kirsch-Volders, M. and Vrijhof, H. (2009). Thresh-

olds for genotoxins and their application in risk assessment. *Mutation Research, 678(2)*, 71-158.

Carthew, P., Clapp, C. and Gutsell, S. J. (2009). Exposure based waiving: The application of the toxicological threshold of concern (TTC) to inhalation exposure for aerosol ingredients in consumer products. *Food and Chemical Toxicology 47*, 1287-1295.

Carthew, P., Fletcher, S., White, A. et al. (2006). Transcriptomic and histopathology changes in rat lung after intratracheal instillation of polymers. *Inhalation Toxicology 18*, 227-245

Diaz, D., Scott, A., Carmichael, P. et al. (2007). Evaluation of an automated in vitro micronucleus assay in CHO-K1 cells. *Mutation Research 630*, 1-13.

Divkovic, M., Pease, C. K., Gerberick, G. F. and Basketter, D. A. (2005). Hapten protein binding: from theory to practical application in the in vitro prediction of skin sensitisation. *Contact Dermatitis 54*, 73-74.

EC (2009) Call for proposals under the 2009 and 2010 work programmes of the 7th EC Framework Programme for Research, Technological Development and Demonstration Activities 2009/C 177/06: FP7-HEALTH-2010-Alternative-Testing. *Official Journal of the European Union 52*, 16.

EC (2003). Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products'. European Commission. *Official Journal of the European Union L66*, 26-35.

Fentem, J. and Westmoreland, C. (2007). Risk assessment + new technologies: Opportunities to assure safety without animal testing and better protect public health? http://www.AltTox.org

Fentem, J., Chamberlain, M. and Sangster, B. (2004). The feasibility of replacing animal testing for assessing consumer safety: a suggested future direction. *ATLA 32*, 617-623.

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M. et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicological Sciences 97*, 417-427.

Grainger, C. I., Greenwell, L. L., Martin, G. P. and Forbes, B. (2009). The permeability of large molecular weight solutes following particle delivery to air-interfaced cells that model the respiratory mucosa. *European Journal of Pharmaceutics and Biopharmaceutics 71*, 318-324.

Jowsey, I. R., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). A future approach to measuring relative skin sensitising potency: a proposal. *Journal of Applied Toxicology 26*, 341-350.

Kato, H., Okamoto, M., Yamashita, K. et al. (2003). Peptide-binding assessment using mass spectrometry as a new screening method for skin sensitization. *Journal of Toxicological Sciences 28*, 19-24.

Kirkland, D., Aardema, M., Henderson, L. and Muller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutation Research 584*, 1-256.

Maxwell, G. and MacKay, C. (2008). Application of a systems biology approach for skin sensitization risk assessment. *ATLA 36*, 521-556.

Maxwell, G., Aleksic, M., Aptula, A. et al. (2008). Assuring consumer safety without animal testing: A feasibility case study for skin sensitisation, *ATLA 36*, 557-568.

Natsch, A., Gfeller, H., Rothaupt, M. and Ellis, G. (2007). Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicology in Vitro 21*, 1220-1226.

NRC (2007). Toxicity testing in the twenty-first century: A vision and a strategy, 146pp., Committee on Toxicity and Assessment of Environmental Agents, National Research Council, Washington, DC

Pendlington, R. U., Minter, H. J., Stupart, L. et al. (2008). Development of a modified in vitro skin absorption method to study the Epidermal/Dermal disposition of a contact allergen in human skin. *Cutaneous and Ocular Toxicology 27*, 283-294.

Shannon, P., Markiel, A., Ozier, O. et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 2498-2504.

Walsh, M. J., Bruce, S. W., Pant, K. et al. (2009). Discrimination of a transformation phenotype in Syrian golden hamster embryo (SHE) cells using ATR-FTIR spectroscopy. *Toxicology 258*, 33-38.

Westmoreland, C. and Holmes, A. M. (2009). Assuring consumer safety without animals: Applications for tissue engineering. *Organogenesis 5*, 67-72.

Acknowledgements

We would like to acknowledge the input of all of the scientists both at Unilever and within the following organisations for the roles that they continue to play in Unilever's ongoing research to deliver novel ways of assuring consumer safety.

4-Front Research, Altogen Biosystems, Barts and the London School of Medicine, BioFocus DPI, BioReliance, Cardiff University, CeeTox, Charles River, Covance, Cranfield University, CXR Biosciences, Cytoscape, DuPont, Entelos, Fera, Gentronix, Hall Analytical, HSL, Hunter Systems, Imperial College, Intertek, IVDK, Kings College London, Lancaster University, Leadscope, LGC, Massachusetts Institute of Technology, MatTek, M-Scan, Münster University, Oroxcell, ProBioGen, Qsar Foundation, Sheffield Hallam University, Syngenta, UCSD, University College London, University of Birmingham, University of Cambridge, University of Glasgow, Universidade Nova de Lisboa, University of Liverpool, University of Manchester, University of Nottingham, University of Pittsburgh, University of Southampton School of Medicine, University of Tennessee, University of Wales Swansea, University of Warwick and Xceleron.

Correspondence to

Carl Westmoreland
Safety & Environmental Assurance Centre (SEAC), Unilever
Colworth Science Park, Sharnbrook
Bedford, MK44 1LQ, UK
e-mail: carl.westmoreland@unilever.com

The Economics of Animal Testing

Annamaria A. Bottini [1] and Thomas Hartung [2]

[1]Johns Hopkins University Medicine International, Baltimore, USA, and European School of Economics, Milan, Italy;
[2]Johns Hopkins University, CAAT, and CAAT-EU, University of Konstanz, Germany

Summary

The area of animal testing and its alternatives attracts considerable attention in the media and public opinion, and it involves a spectrum of aspects including the sciences, policy, legislation, health and safety, ethics, and philosophy. Yet, in remarkable contrast, analyses of its economic aspects are rare. We investigate and critically review animal testing with special focus on the economics. The economic implications of animal testing are colossal with sales value of regulated products in Europe alone reaching 1.7 trillion € per year (5.6 t€ worldwide). Within this context alone the reader should be able to grasp the enormity and massive effects on manpower, investments, and animal numbers related to animal testing. Indeed, the classical toxicology of chemical substances on animals costs 620 m€ in the EU (2-2.5 b€ worldwide), all animal experimentation directly employs 15,000 people (worldwide 73,000), and involves about 60,000 experimenters (300,000 worldwide). In terms of animals, 23.3% of the 12.1 million animals used in the EU 2005 were for regulatory tests plus 31% for industrial R&D!

For the sake of brevity, and also because of the limited availability of reliable data resources, the analysis discussed here focuses on Europe; where possible and necessary, a global perspective is given as well, since the effects of globalization cannot be ignored. We also tackle some of the more technical aspects, such as the role of animal testing in the product life cycle and the economic drivers that work for or against an improvement (radical or incremental) in animal testing. The economic consequences of the shortcomings of animal tests are considered, and an attempt is made to realize the true societal consequences of such tests. As globalization advances, it will have an impact on the whole concept of consumer product safety, as well as on animal testing, its alternatives, and related industries.

Keywords: animal experimentation, toxicology, econometry, statistics

1 Introduction

In industrialized countries, few topics are more controversial than the use of animals in science and industry: on one side, militant animal activists threaten the lives of researchers and release animals from laboratories; on the other side, industry and regulators insist that consumer safety and product quality depend on progress made with animal tests. The discussions center on the ethical justification, the safety of citizens, the scientific feasibility of alternative approaches, and the efficiency of regulation (Fig. 1). An important component is largely absent from these discussions, however: economics. The discussion immediately takes on a whole new spin when viewed in an economic context (Fig. 2).

Most animal testing for regulatory purposes occurs for safety assessment, i.e. toxicology and vaccine testing. Toxicology is in a crisis – especially regulatory toxicology; modern, mechanistic toxicology is indistinguishable from other sciences in its methodology and scope. Major limitations in the science of risk assessment have been summarized elsewhere (Hartung and Leist, 2008; Griesinger et al., 2009).

The economies relying on animal testing can be analyzed on a national, European, or global level. The national level is highly dependent upon the respective European legislation. For implementation of European directives, however, it is worth noting the particular national attitude toward the subject of interest. For example, Italians are very passionate about animals: the number of indoor-housed cats and dogs has increased an estimated 200% over the last decade in Italy. On a scale of 1 to 10, Italians rate the importance of animal welfare at 7.8 (Eurobarometer, 2007). Independent of socio-demographic background, 51% of Italians consider animal welfare when buying food (EU only 43%). About 900,000 animals are used in Italian laboratories per year. Only 40% of Italians agree to animal experiments to advance human health (EU 45%), and 34% disagree; in the age range 15-24, only 35% agree while 46% disagree. More than

half (53%) of Italians feel that animal welfare does not receive enough political attention (Eurobarometer, 2007; with 29,000 European citizens).

As the home of some of the world's biggest names in fashion, Italy has built a reputation for high-end and luxury cosmetics and fragrances. Clothing designers such as Gucci, Dolce & Cabanna, Giorgio Armani, and Benetton all have successfully expanded their branding into the international arena. The annual turnover in the Italian cosmetic industry is about 8 billion €, with current growth rates around 2.5% per year (personal communication). In the last 20 years, the Italian chemical industry (130,000 directly employed, 275,000 indirectly employed, 57 billion € production, 2% annual growth) has changed greatly. The larger chemical industries have disappeared, breaking up into numerous smaller enterprises with fewer than 100 employees, according to the Italian trade association; only 40 companies have more than 200 million € turnover. As part of this shift from the large-scale production of chemical products, these smaller Italian chemical enterprises have concentrated more on the specialization of "niche products"; the latter refer to specialized products that target a specific market segment where there is very little competition (Kim and Maurborgne, 2002). This strategy has enabled the Italian chemical industry to emerge, over the years, as one of the world leaders in chemical production. Nonetheless, it is precisely this small and medium-sized structure that could expose the Italian chemical industry to the greatest risks in the process of conforming to the REACH provisions. Currently, the Italian pharmaceutical industry (324 companies) directly employs 67,000 people and makes 22 billion € turnover. Export amounts to 12 billion €.

These few numbers characterize one of the 27 EU countries, but it shows that EU politics in this area is deeply anchored in the feelings of the citizens and the importance of some industries regulated by animal testing.

2 Methodology

This study uses open source research based on the internet, also employing various scientific databases (MedLine/PubMed, Knowledgefinder, Google-Scholar), as well as Internet-based research and direct contact with peers in the field. For details see Bottini and Hartung (2009). The authors attended a series of conferences to complete the picture and also drew upon their professional experiences and work contacts at the European Centre for the Validation of Alternative Methods (ECVAM, 2004-2007 and 2002-2008, respectively).

3 Analysis of the economics of animal testing

3.1 The "regulated communities" relying upon animal testing

The first step in understanding the economic mechanisms of animal testing is to analyze which industrial sectors use animal testing and to what extent. This excludes animal experimentation for basic research, or 33% of all animal use (European Commission, 2007). Basic research, however, will follow industrial needs and priorities to some extent, primarily via research funding. We have discussed elsewhere some opportunities to improve this (Gruber and Hartung, 2004). Thus, we will focus mainly on regulatory testing, which accounts for 23.3% of all animal use (European Commission, 2007). Of these animals, 8% are used for toxicology and safety assessments and 15.3% for veterinary and human medical product safety. This represents the requested testing of substances and products to comply with legal restrictions promulgated by government authorities, along with animal-based research for innovation of industrial products (31% of all animal use; European Commission, 2007). Table 1 summarizes the size of the different indus-

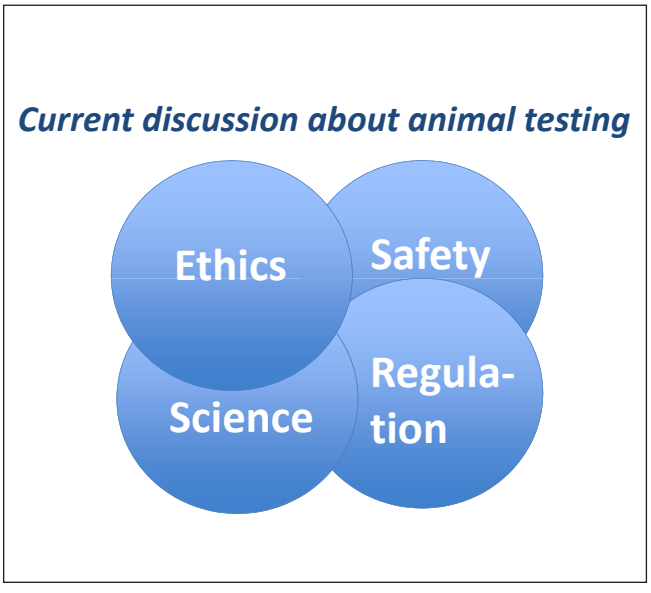


Fig. 1: The four major current areas of the animal testing debate

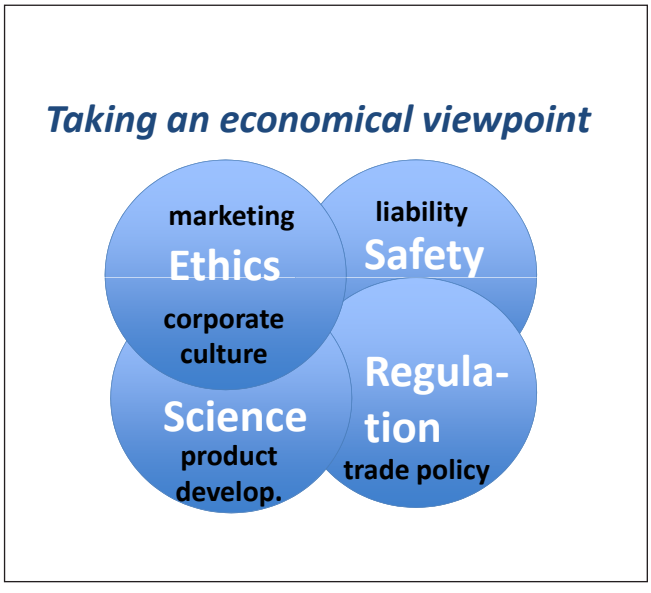


Fig. 2: The economic perspective on animal testing

tries in Europe, their share of the world market, and estimated new substances introduced per year, as well as animal use.

Figure 3 (below) shows the relative markets in Europe, while Figure 4 shows the situation worldwide. Details and sources can be found in Bottini and Hartung (2009).

3.2 Globalization

Taylor and coworkers recently completed a very interesting analysis of animal use in different countries (Taylor et al., 2008). GDP, the gross domestic product, is "the market value of all the goods and services produced by labor and property located in a given region, usually a country." When plotting their estimates per country against gross domestic product, i.e. the most common measure of national income and output for a given country's economy, we see (Bottini and Hartung, 2009) an almost perfect correlation ($r^2 = 0.99$), which, however, is strongly influenced by the extreme values, still above $r^2 = 0.82$ for the raw values. The message is simple: economy needs animals – quite exactly one animal per million $ GDP – and there are no major deviations for the major economies. Europe has about 30% of world GDP, and we can use this to translate the European data to other economies.

What does regulatory safety testing of synthetic substances cost? Fleischer (Fleischer, 2007) has carried out a survey on costs and capacities of laboratories running safety assessments in nine countries. Since this is a competitive market with a stable situation of demand and offer of test capacities (which might change soon in the context of REACH, see below), the average prices also can be used as a solid estimate of test costs for in-house tests in some companies. Applying these costs and the

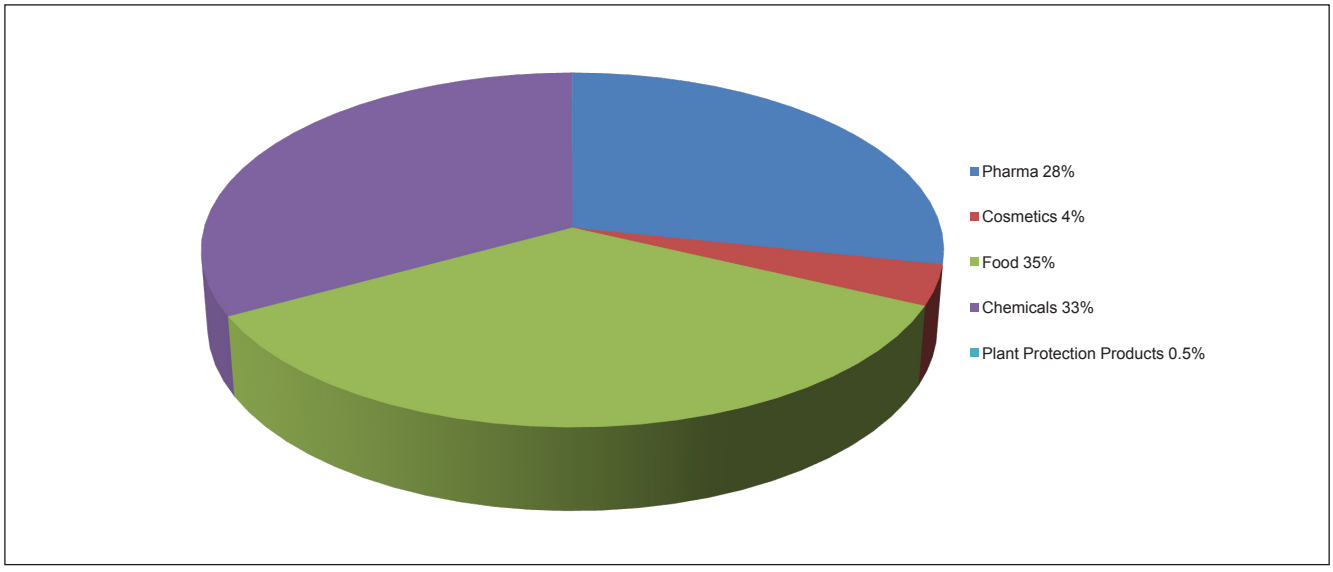


Fig. 3: Markets of animal testing regulated industries in Europe (Total = 1.7 trillion €)

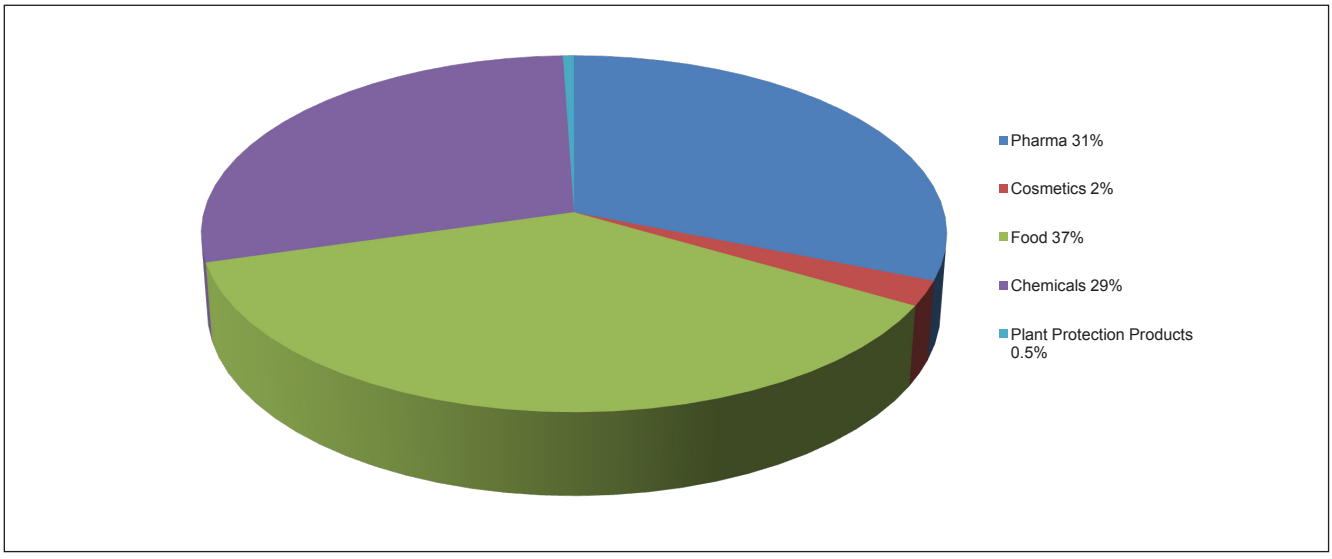


Fig. 4: World markets of animal testing regulated indutries (Total = 5.6 trillion €)

animal numbers per test to the EU statistics on animal use in 2005 allows us to estimate testing costs (Bottini and Hartung, 2009), shown graphically in Figure 5, i.e. the contribution of different types of tests to the costs of toxicity testing in Europe. This rough estimate would suggest that toxicity and safety assessments in Europe cost 620 m€ per year. Interestingly, this estimate would suggest that most resources are spent on (sub-) chronic testing (332 m€), (sub-)acute testing (127 m€), and carcinogenicity testing (82 m€). Extrapolating these data to a global scale, a factor 3-4 based on overall animal numbers and

Tab. 1: summarizes the size of the different industries in Europe and their share of the world market, estimated new substances introduced per year as well as animal use

EU	Pharma	Chemicals	Cosmetics	Food
Companies	> 2,200	29,000	2,000	310,000
Sales	484 b€	563 b€	63 b€	600 b€ 3.3 b€ food additives 8.6 b€ pesticides
% World market	28%	35%	50%	29%
Animal use	490,000 R&D: 3.8 million	90,000 strong increase due to REACH to come	5,600 Banned 11 March 2009	500,000 mainly shellfish toxin testing
New substances per year	12	ca. 300 (> 100kg)	400 (5,500 new products)	Not applicable 10 food additives 8 pesticides
Employee	635,000	1.26 million	150,000 + 350,000 retail	4 million

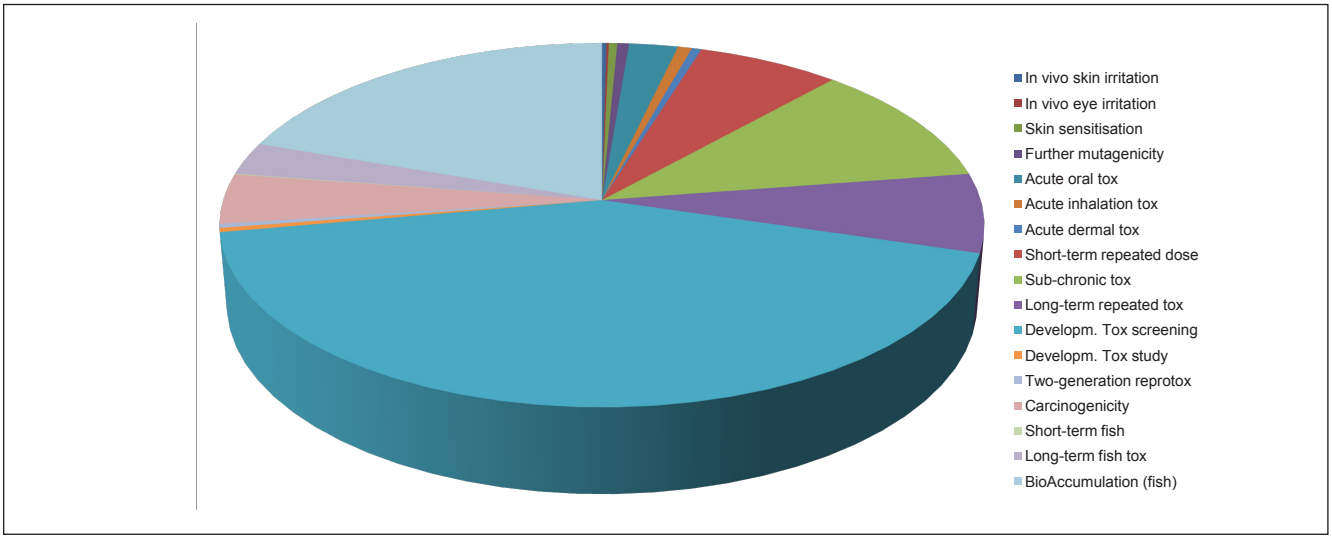


Fig. 5: Contribution of different toxicological hazards to testing costs in Europe

Animal use number for 2005 (European Commission, 2007) were interpreted by applying costs and animal numbers of guideline tests. Some tests are grouped in the 2005 animal use statistics; the following rough assumptions were used to split: acute and subacute tests (Acute oral tox 55%, Acute inhalation tox 5%, Acute dermal tox 10%, Short-term repeated dose 30%); subchronic and chronic tests (Sub-chronic tox 50%, Long-term repeated tox 50%), developmental toxicity (Developm. Tox screening 20%, Developm. Tox study 80%) and toxicity to aquatic vertebrates (Short-term fish 80%, Long-term fish tox 5% and BioAccumulation (fish) 5%).

share of world market of industries seems realistic, suggesting 2-2.5 b€ spent for toxicological testing per year.

The number of people working with animals or in the industry serving testing is difficult to determine. A UK report (Lantra, 2006) indicates that 14,000 personal license holders for animal experimentation (mainly scientists) carried out experiments on 2.72 million animals in 2003. There is no reason to assume that in other countries significantly fewer or more personnel are required for the conduct of animal experimentation. Thus, more than 60,000 people might be involved in animal experimentation in Europe, or about 300,000 worldwide. Similarly, the 3,400 professionals working in the UK on animals for experimental purposes can be extrapolated to 14,600 in the EU and 73,000 worldwide. Contract research laboratories, which represent an industry with 4.1 b€ total sales, form an important part of the animal testing market. In conclusion, animal experimentation represents a relevant industrial branch on its own.

3.3 The extent of animal-test-based regulation differs for different industrial sectors

In 2005, with 25 member states of the EU (EU-25) at that time, 12.1 million animals were used in experiments (European Commission, 2007). In EU-15, the total number of animals used increased in 2005 by 339,279 compared to the previous report from 2002, an increase of 3.1%. The number of animals used for toxicological and other safety evaluation dropped from about 9.9% (2002 data) to 8% of all animal use in Europe in 2005. The decrease, however, is minor in total numbers, dropping from 1,066,047 to 1,026,286 animals, and this is due mainly to some increase in total numbers with the inclusion of the 10 new Member States.

Table 1 allows comparisons of the different industries with regard to the extent of animal use for regulation of the sector, also based on the regulatory testing per new substance or per billion of turnover of the industrial sector (ranging from 90 for cosmetics to 10,000 for PPP). In conclusion, the search for pharmaceuticals is most animal-intensive (40,000 per new substance for safety and 350,000 for R&D), obviously because of the many substances tested and abandoned along the way. Chemicals require little testing; most receive just the base-set of testing ("six-pack": skin and eye irritation, skin corrosion, mutagenicity, sensitization, acute toxicity testing). Food additives and PPP require considerably more animals, because of both higher testing demands and the many substances that never make it to the market (Hartung and Koëter, 2008). Hartung (2008b) explains why little animal testing is carried out for cosmetics: This European industry represents 2,000 relatively profitable companies with 60 billion € turnover. The sector is characterized by quick product exchange (5,000 new products in Europe and 22,000 world-wide per year, 25% of turnover with products released within the last 6 months). Market leader L'Oréal, for example, releases 3,000 new products per year and, out of 500 patents, about 100 are patents on substances. It can be assumed that several hundred new substances are introduced into cosmetics every year. Given a total of about 8,000 cosmetic ingredients in use, this number represents a reasonable assumption of turnover. This might be compared with only eight, on average, new active substances entering the pharmaceutical world-market per year. Although safety must be assured for products we put on our skin, into our eyes, and into our mouth, testing cosmetic products on animals is now rare to negligible. This is possible because the cosmetic industry is not producing its chemical ingredients; ingredients are tested as chemicals, and many food ingredients, as well as natural products, are used.

3.4 Different functions of animal testing in the product life cycle

Agent discovery refers mainly to the biologically active agents, i.e. drugs and PPP. Animal testing has lost a lot of its importance for agent discovery. Search strategies with higher through-put typically are employed, which makes good sense given the figures on the number of substances that must be tested to identify a lead compound. The overall reduction by probably two-thirds in the number of animals used over the last three decades largely reflects this change (along with the reduction of duplicate testing). Today, molecular understanding of pathophysiology and desired mode of action, in general, allow non-animal method use – with late stage confirmatory testing in a disease model in the case of drug development.

Safety evaluations typically are defined by the regulatory requirements. This is somewhat telling: Shouldn't every product have a certain profile of reasonable test requirements that address, e.g. the "excess pharmacology" of the agent and chemical class-specific concerns? Shouldn't the company spending a billion on the development of a drug know very well to choose the right set of tests? As in other areas, however, a tick-box approach is still followed wher a standard "laundry list" of tests is run. This has several advantages for the notifier as well as the regulator:

- a standard set of information means fewer difficulties in interpretation (comfort zone of the regulator);
- omitting a certain test might imply responsibility if later effects are found;
- the production of the dataset can be easily timed because a standard battery is run;
- defense against liability claims is easy, since the state of the art has been applied.

Animal use might be considerably reduced, however, if information requirements must be justified towards the regulator and liability stays with the producer. REACH is heading in exactly this direction by requesting a testing plan to be approved by the agency for higher tonnage-level chemicals and turning the burden of proof (liability) to the industry. It will be most interesting to see whether this new approach works out and can be adapted to other sectors.

The success of Body Shop with their aggressive "not tested on animals" strategy is the key example for marketing with animal welfare (Wikipedia website, Bodyshop). Signage posted in Body Shop locations reads, "Our products are not tested on animals, never have been and never will be." However, the Body Shop website expands this to acknowledge, "the fact is that

almost all cosmetics ingredients have been animal tested by somebody at some time for someone. So no cosmetics company can claim that its ingredients have never been animal tested." It is not clear what the actual value of this statement is (McSpotlight website): According to market leader L'Oréal, no cosmetic has been tested on animals in Europe since 1989, but all (most probably including Body Shop) make use of ingredients tested on animals by others, especially the supply industry. The Body Shop represents a role model for a marketing strategy, however, which has put competitors under pressure. The Body Shop experienced rapid growth, expanding at a rate of 50% per year. The success of the brand (which has been part of L'Oréal since 2006) demonstrates the value consumers place on animal welfare, a finding in line with results of recent European surveys showing the extremely high importance citizens accord to animal welfare issues. A 2005 EUROBAROMETER survey (Eurobarometer, 2007) reported that 82% of EU citizens believe we have a duty "to protect the rights of animals whatever the cost." The EU Commission survey showed that 93% of respondents believe that more should be done to improve the welfare/protection of experimental animals. It might be appropriate to think about a protected "not tested on animals" label with the respective standards.

Different safety testing standards also are barriers to free trade. We discussed earlier the important role of OECD test guidelines in harmonizing test requirements (Bottini et al., 2007). However, we must not forget that there are important markets outside OECD. Brazil, China, Russia, and other such countries with an increasing upper class and high-tech industry represent not only exporters but also importers of European products. Thus, reliance on traditional methods there presents an obstacle for the introduction of any novel approach – even if accepted in the OECD. The only possible solution is the insistence on the novel method (e.g. by deleting the traditional one from the catalogue of test guidelines, which so far has been done only for the classical LD_{50} test, OECD TG 401), and the international collaboration and harmonization with all major (emerging) markets. Since we represent for them far more interesting export markets, chances are good to actually export our standards (see below).

3.5 Alternatives methods as a market

We demonstrated earlier that animal testing is a multi-billion € market. Alternatives evidently have a similar potential to become a very profitable market. In fact, some alternative methods, such as pregnancy tests or pyrogenicity testing (Limulus test), have reached three-digit m€ sales. REACH raises hopes that certain novel methodologies will find a market, allowing their standardization and commercialization. This, in turn, improves their availability and international use. To some extent, the alternatives market is also expanding due to the silent substitution of services by contract research laboratories (mutagenicity, skin and eye corrosion/irritation, fish egg test etc.). Still, development, validation, and regulatory acceptance take too long (more than 10 years) to suit the needs of biotech companies for return of investment. There is room for improvement.

3.6 The limitations of animal tests have economic consequences

Using a technology with limitations carries a price. There are mechanisms at work to deny the limitations as well as its price. We shall offer here some estimates as to the latter.

The impact of prevalence on toxicity testing has been discussed earlier (Hoffmann and Hartung, 2005). The concept is easy: when looking for a hazard, we need to consider whether it is a frequent or a rare one. If it is frequent, even methods with a large uncertainty give a fair bet, but if we are looking for something rare, the error rate becomes relevant even for reasonably good methods. This precautionary approach might be acceptable for new substances without established economic value, but when REACH is now applying the same strategy to the most valuable substances, the sacrifice might be more than substantial (Hartung, 2009). The precautionary principle goes back to the German concept of a "*Vorsorge-Prinzip*" (cautionary principle). In general, it means, in case of uncertainty, assume the worst and make decisions on that basis. In toxicology, the concept is inherent in the creation of over-sensitive test models, which minimize false-negatives (missed toxicants) by accepting false-positives (wrong allegations of toxicity). High doses, most sensitive species, testing in several species, and sensitive endpoints (low thresholds for toxicity categories) are the typical tools employed to render a test sensitive. This follows the same pre-cautionary logic.

If taken to an extreme, the precautionary principle may indeed inhibit economic development (Gollier and Treich, 2003). One key problem is the "opportunity cost of precautionary measures" (Majone, 2002). The attempt to control poorly understood, low-level risks necessarily uses up resources that in many cases could be directed more effectively to the reduction of well-known, large-scale risks. But there is also a further type of opportunity cost to consider: abandoning a certain substance or restricting a technology impairs opportunities for business and for serving societal needs. An evident example is a therapy that does not make it to the market/patient because the precautionary tests suggest a toxicity problem. It is difficult to estimate this effect. We have tried elsewhere to estimate the false-positive rate of prominent toxicity tests (Hartung, 2009). To identify estimated 5% carcinogenic chemicals, for example, the cancer bioassay is employed, which finds 53% of all substances test positive; to identify 2-3% reproductive toxicants, the two-generation test in two species is employed, which finds about 60% of substances positive. If these were independent tests, the test battery of both tests would let only 19% of substances pass without assigning them a carcinogenic or reproductive toxicant label, a sure k.o. for further development. But the toxicological toolbox is much larger…

Lichtenberg (Lichtenberg, 2005a) has calculated the impact of the launch of New Chemical Entities (NCE), i.e. structurally new drugs, on the steady increase in longevity. He showed that launches of NCEs have a strong positive impact on the probability of survival. Between 1986 and 2000, the overall population gained 1.96 years of life expectancy and, according to these estimates, NCE launches account for 40% or 0.79

years. He calculated that the average annual increase of the entire population resulting from one NCE launch is 0.056 years (2.93 weeks). It is tempting to speculate now what the impact will be of five times more substances moving ahead in clinical development, because the false-positive results of the cancer bioassay and the two-generation reproductive toxicity study could be avoided.

Lichtenberg has confirmed his data with a study of differences in drug launches in different federal states of the US (Lichtenberg, 2007), where he found even a 63% contribution to increase in life expectancy of new drugs (2.43 years from 1982 to 2004). In another study (Lichtenberg, 2005b), he showed the impact of prescription of old (pre-1970) to new drugs on mortality in 800,000 people of the Medicaid program in the first half of 2000. He calculated that the actual mortality rate of 3.5% for the following two years would have been 4.4% if all received pre-1970 drugs.

This argument shows that a precautionary test approach resulting in an unnecessary reduction of drug candidates or pressure to change to lead compounds with a less favorable pharmacological profile will directly affect society, not just the prosperity of the individual company. Similar calculations for other areas of regulation are not available, but the true costs of precautionary chemicals regulation have been predicted (Durodie, 2003).

The question we have to address is, can products of the early 21st century be regulated with methods of the early 20th century? Sure, some things never change because they are basic and meet the needs, such as measuring length with a meter. But even this required negotiation, such as the international agreement on a gold standard. The purpose of animal tests, however, is complex and changeable, and we are far from international gold standards. The closest we come are international test guidelines, with known and unknown limitations (Hartung, 2008a).

We commented earlier on the problem of freezing these international guidelines (Bottini et al., 2007): it takes a decade not only to generate one but also to make significant changes. At the same time, regulated products are changing at an ever accelerated pace. Recent examples include cellular therapies, gene technologies, and nanotechnologies. These force us to regularly review the way we are doing things, to allow the evolution of toxicology (Hartung and Leist, 2008; Leist et al., 2008). However, we lack the mechanism for change. We might argue that the validation of alternative approaches represents exactly this, a means to introduce innovative methods to replace the current ones. The key problem, however, is that as long as we consider the status quo as a gold standard without limitations, we cannot really move ahead and improve; the best will be an approximation with a perceived compromise, because 100% identical results are usually not achieved. We have therefore proposed (Hoffmann and Hartung, 2006) a mechanism to systematically review the *status quo* on the basis of the best evidence available at the time with the most objective and transparent processes. Borrowing from the clinical medicine field and their evidence-based medicine movement, the suggestion was put forward to create an evidence-based toxicology. It is remarkable that, in less than three years, the idea has flourished, with the proceedings of the first international forum from 2007 now available (Griesinger et al., 2009; Evidence-based toxicology website), a symposium at the last EuroTox in 2008, and a special issue of *Toxicology*, in preparation. It is worth noting that the first chair for evidence-based toxicology was created at the beginning of 2009 (Doerenkamp-Zbinden chair for evidence-based toxicology in the Bloomberg School of Public Health at Johns Hopkins University, Baltimore, US). Since this university also hosts the Cochrane center for evidence-based medicine for the US, an optimal synergy with the far more established EBM (Evidence Based Medicine) movement can be hoped for.

3.7 Economic forces to improve the situation

The common European market was the driver for the progressive European animal welfare legislation. A regulation is a legislative act of the European Union, which becomes immediately enforceable as law in all member states simultaneously. Regulations can be distinguished from directives, which, at least in principle, need to be transposed into national law. Under the European Constitution regulations would have become known as "European laws," but this proposal has since been dropped.

EU regulation has a general scope and is obligatory in all its elements; it is directly applicable in all Member States of the European Union. Any local laws contrary to the regulation are overruled, as EU Law has supremacy over the laws of Member States. New legislation enacted by Member States must be consistent with the requirements of EU regulations. Thus, regulations constitute the most powerful or influential of the EU legislative acts.

In principle, the EU had no mandate for animal welfare in 1986. It is remarkable that the Directive 86/609/EEC on the welfare of laboratory animals was only created by considering different animal welfare standards unfair conditions to enter the common market in different Member States. In principle, the legislation does not cover basic research, but most Member States, when translating it into national law, have updated this in one act.

But it is not political will and the forces of the common market alone that have helped raise awareness of animal welfare. We are seeing a more general increase in attention to the ethical aspects of business. Ethical treatment of employees, sensitivity to different cultures, marketing with ethical arguments, and also social responsibility as an asset increasingly form an integral part of the culture of individual companies.

In an earlier article (Bottini et al., 2007), we discussed the probable impact of globalization on the international spread of alternative approaches. Here, we would like to reflect only on the similarity between our arguments and those made by David Vogel with regard to environmental standards (Vogel, 1995). Vogel argues that, although trade liberalization has undermined national regulatory sovereignty, it also globalizes regulatory policy-making by exporting standards as well as goods. The primary challenge concerns whether a country should be al-

lowed to restrict the sale of a product on the basis of how it was produced outside its legal jurisdiction. In the case of environmental legislation, the classic dispute concerned the US ban on tuna from Mexico because it did not fulfill the US standards for dolphin protection. In 1991, the GATT (General Agreement on Tariffs and Trade from 1947), the predecessor of the World Trade Association (WTO), created in 1995, ruled against the higher protective standards of the US. It will be extremely important to see whether the animal protection standards of the 7th amendment will lead to a similar WTO ruling. This could have happened already; since September 2004, end-product testing of cosmetics should lead to a marketing ban in Europe, as should the use of ingredients tested on animals after alternatives became available. Products that have conflicting regulations continue to be made available in world markets, however, without any known legal challenge. It will be most interesting to see whether this changes after the deadline of March 2009. This is likely to depend more on animal welfare NGOs challenging companies (as we have seen in the field of environment legislation) than on national prosecution.

Independent of any cost/benefit analysis, the attractive example of the European cosmetics market has led many countries into a discussion of the convergence of legislation. If Europe can demonstrate the feasibility of its novel approach, others are likely to follow. This would follow the argument of Vogel, who does not see free trade impairing higher national standards but instead favoring their export, for the following reasons:

– producers who operate in many markets have a strong interest in making national product standards more similar in order to reduce their production costs;
– the compatibility between trade expansion and protective regulation has to do with the structure and authority of international institutions;
– the increase in regulation has not been more disruptive to trade (due to the link) with increasing international treaties and agreements.

We will see whether this is wishful thinking or, in fact, will lead to the globalization of animal protection standards.

The life sciences are one of the fastest growing disciplines: knowledge is said to double every five to seven years. The key technologies of alternative approaches, i.e. cell culture and computer-based models, have undergone especially dramatic developments, a trend often referred to as the informatics and biotech revolution. Science aims to stay cutting-edge, with such mechanisms as peer-review forcing researchers to stay up to date. The backlog of renewal of methods in regulatory toxicology is most remarkable, since hardly any scientific field is continuing to use experimental set-ups developed 40 to 60 years ago. This reflects the absence of scientific control mechanisms such as publication, peer-review, repeat experiments, and competitive funding.

However, the development of (bio-)informatics and biotechnologies also creates momentum due to the commercialization of these technologies. Companies trying to develop their markets challenge established approaches. They aim to take a share of the agent discovery area (most open to innovation) and basic research, as well as the regulatory testing market (most resistant to change). If our earlier estimates hold true, we are talking about markets of several billion € per year in Europe. Many start-ups have targeted new therapies first but then discovered that alternative testing may be quicker or generate additional revenue. Prominent examples include artificial skin, originally developed for burn patients, stem cell technologies that originally targeted transplant indications, or novel liver cells often developed for liver failure patients. With fresh ideas, professional solutions, and the punch to bring their methods to success, small/medium enterprises have been especially effective in changing the field of alternative approaches over the last few years.

Animal welfare standards are continuously rising – and not because of animal welfare considerations alone. Increasingly, researchers recognize that maintaining high standards of animal welfare also improves the quality of results. This starts with a clear definition of the experimental animal (defined inbred and outbred strains), exclusion of confounding factors, ranging from latent infections (specified pathogen-free, SPF) to sufficiently large and enriched cage environments. The positive effect on experimental design and outcome has been documented on numerous occasions. Various professional organizations, funding bodies, and journals now request that certain standards be explicitly referred to in publications, which has greatly strengthened the effort to implement and further develop these standards. Last but not least, the animal welfare legislation 86/609/EEC has been instrumental in improving and harmonizing animal welfare standards in the EU; expectations are high that the current revision will further augment this.

The animal welfare field, like the environmental field, is characterized by large activist groups that are aligned internationally and are increasingly accepted as stakeholders in legislative and regulatory processes. Such groups play an important role in promoting animal welfare standards and legislation, not only by creating awareness and shaping public opinion but also by lobbying politicians and serving as watch-dogs for the implementation of these standards. It is likely that these organizations will further influence the interpretation and application of legislative standards via court cases, especially the European Court of Justice, as was the case in the environmental field. This is especially important since the EU has no executive forces for the implementation and monitoring of legal provisions.

The treaty of Amsterdam, which went into effect May 1, 1999, establishes new ground rules for the actions of the European Union (EU) regarding animal welfare in a special "Protocol on the Protection and Welfare of Animals." In this sense, it recognizes that animals are sentient beings, and it obliges European institutions to pay full attention to the welfare requirements of animals when formulating and implementing Community legislation. This has been reinforced in the Lisbon treaty of December 2009.

Europe has taken the role of pacemaker for legislative standards aimed at protecting the environment. Recent initiatives give evidence of a similar role in the chemicals, food, and plant protection product area. For animal welfare standards, the lead-

ing role of EU legislation is obvious from the 1986 legislation binding all Member States and as a horizontal legislation with an impact on various other political frameworks. In contrast, the US animal welfare legislation, which dates back to 1967, is in desperate need of an update.

The two key legislations affecting the development of alternative methods, i.e. REACH and the 7th amendment of the cosmetics directive, have already been discussed here extensively. The credibility of such legislation depends strongly on the accompanying measures, such as the provision of research funding and the creation of institutions and agencies to carry out the validation and implementation of novel approaches. European funding (with about 25 m€ by the EU and 19 m€ by member state organizations) is not matched in other economic regions. Seeing the market opportunities, as well as the regulatory needs summarized earlier, this appears to be a good investment – a rather small investment, actually, given turn-over in this field and the enormous potential for improved regulation.

Porter (1990), most prominently, has put forward the hypothesis that regulations stimulate innovation (Porter, 1990). The concept has been challenged for chemical industry and REACH (Frohwein and Hansjurgens, 2005), but only with regard to the substitution of substances. A principal goal of REACH (article 1), however, is the development of new alternative methods as well; here, the Porter effect might come into action, as we have seen the positive impact of the 7th amendment to the cosmetics directive (Hartung, 2008a). REACH focuses its information demands on exactly the complementary toxicological endpoints to those of the 7th amendment, or more precisely those required only for the later deadline in 2013, which are – with the notable exception of repeated dose toxicity – less often required for cosmetic ingredients. Together, the two legislations thus create a pressure toward innovation over the entire spectrum of toxicological endpoints.

3.8 PESTEL model approach

The PEST/PESTEL analysis approach will be used to summarize the findings. PEST stands for "Political, Economic, Social, and Technological analysis," describing a framework of macro-environmental factors used in the environmental scanning component of strategic management. PESTEL extends this to environmental and legal factors.

Political factors are represented mainly by European legislations, i.e. the currently revised animal welfare legislation, the 7th amendment of the cosmetics legislation, the chemical legislation, REACH, the common food legislation, and the ongoing revision of the PPP regulation. Among these, the horizontal animal welfare legislation, which for 24 years now has established the highest welfare standards worldwide, as well as the cosmetics and chemicals legislation, are most remarkable. Implementation of the latter two will require methods not yet available; they also affect global industries, since they ask that these standards be met by trade partners. These political decisions greatly influenced the health, education/science, and infrastructure of Europe.

Economic factors are represented first of all by the costs of the regulation, market access opportunities, which affect the costs of exporting goods, and the supply and price of imported goods. However, a very complex picture of the animal testing industry emerges on its own.

Social factors play a key role here: both the increasing risk avoidance and animal welfare attitudes have an impact on the political process and, via marketing, on companies' strategies and images. These trends in social factors affect the demand for a company's products and how that company operates.

Technological factors play a major role in a science-based area. They have been addressed only briefly, but it is evident that the lack of translation of technological progress into current regulatory testing procedures is an important characteristic of this field. The market opportunities for biotech and informatics products in this area also represent key drivers. These technological shifts affect costs and quality of testing, and they lead to innovation.

Environmental factors do not apply here, with the notable exception that environmental hazards represent an animal testing need, creating a sub-market.

Legal factors are represented by the legislations cited above, which were the result of highly political discussion processes. These factors now determine a company's operation, its costs, and the demand for its products.

4 Conclusions

Coordination and harmonization of legislation is an important step towards reducing the number of regulatory animal tests. Companies would benefit from clarity and uniformity of the regulatory requirements for the registration and release of substances and products in the different market segments.

Governments would benefit from uniform regulations that would not require business to comply with a different set of rules for every market or market.

The concept of the 4Cs (Bottini et al., 2008), developed in a workshop on "Optimization of the Post-Validation Process," will be a useful approach to improving communication among the various stakeholders.

4Cs of Communication:
Consultation
Collaboration
Coordination
Convergence

Enhanced communication at the European level, and even at a broader level, including, for example, the other member countries of the OECD, could facilitate the harmonization of legislation and regulations as a key requirement for reducing regulatory animal testing.

It is necessary to identify the regulatory key players and responsible persons in each toxicological and substance/product area. This will serve as a basis for networking and will improve communication among the stakeholders, including industry and animal welfare groups.

The effort to view animal testing in terms not of scientific or ethical concerns but of business considerations demonstrates the degree to which many stakeholders with significant commercial interests are affected by the ethical and scientific discussions. Although it was difficult to produce a precise representation, since many figures are not freely available, an interesting picture has already emerged on the basis of estimates and extra- and interpolations. This picture shows that major industries, trade, and workforces are intimately linked to regulation based on animal testing. Clearly, this is more than simply production costs or barriers imposed; rather, the problematical quality of current testing also has an impact on decision making and the regulation of products. This understanding might help stakeholders to consider more critical current procedures, the first step toward opening up to the prospect of change. Businesses are bringing animal welfare standards to the fore at this point because the delays and costs involved in animal testing impair the economy to a sometimes surprising extent. The somewhat premature calculation that precautionary testing may slow down medical progress and thereby reduce gain in life expectancy suggests that we are talking about relevant societal effects, not just impaired business opportunities for individual companies.

It appears that the field merits a more in-depth analysis of its economic drivers. With access to the privileged information of regulators and private companies and business intelligence, a more detailed picture of the multi-faceted economics of animal use could be generated. This promises to inform the political decision- making process and might allow real cost/benefit analysis of the current use of animals. Some of the larger political programs (REACH, PPP revision, cosmetics amendment), especially, might require this assessment. The analysis is also very much restricted by the European focus (due to the availability of animal use data), while the global character of the described phenomena is well acknowledged (Bottini et al., 2007). In some areas, the discussion on animal use currently is entirely scientific, ethical, or legal; it might be worthwhile to add the economic dimension to broaden perspectives.

To close, this study represents only a first sketch of an analysis of the economic aspects of animal testing. Some trends perceived might need to be revised, but it is comforting that economists also see their approaches, in general, as somewhat limited. To cite Laurence J. Peter (1919-1988): "*An economist is an expert who will know tomorrow why the things he predicted yesterday didn't happen today*."

References

Bottini, A. A. and Hartung, T. (2009). Food for thought … on economics of animal testing. *ALTEX 26*, 3-16.

Bottini, A. A., Alepee, N., De Silva, O. et al. (2008). Optimization of the post-validation process. The report and recommendations of ECVAM workshop 67, *ATLA 36*, 353-366.

Bottini, A. A., Amcoff, P. and Hartung, T. (2007). Food for thought … on globalization of alternative methods, *ALTEX 24*, 255-261.

Durodie, B. (2003). The true cost of precautionary chemicals regulation. *Risk Analysis 23*, 389-398.

Eurobarometer (2007). Attitudes of EU citizens towards Animal Welfare. http://ec.europa.eu/food/animal/welfare/survey/sp_barometer_aw_en.pdf

European Commission (2007). Fifth Report from the Commission to the Council and the European Parliament on the Statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union COM/2007/675 final. http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm

Evidence-based toxicology website. www.ebtox.org

Fleischer, B. (2007). Testing costs and testing capacity according to the REACH requirements – results of a survey of independent and corporate GLP laboratories in the EU and Switzerland. *Journal of Business Chemistry 4*, 96-114.

Frohwein, T. and Hansjurgens, B. (2005). Chemicals regulation and the Porter hypothesis – a critical review of the new European chemicals regulation. *Journal of Business Chemistry 2*, 19-36.

Griesinger, C., Hoffmann, S., Kinsner-Ovaskainen, A. et al. (2009). Proceedings of the First International Forum Towards Evidence-Based Toxicology. Conference Centre Spazio Villa Erba, Como, Italy. 15-18 October 2007. *Human Exp. Toxicol. 2008, Spec. Issue: Evidence-Based Toxicology (EBT) 28*, 81-163.

Gruber, F. P. and Hartung, T. (2004). Alternatives to animal experimentation in basic research. *ALTEX 21, Suppl. 1*, 3-31.

Gollier, C. and Treich, N. (2003). Decision-making under scientific uncertainty: the economics of the precautionary principle. *Journal of Risk and Uncertainty 27*, 77-103.

Hartung, T. (2009a). Toxicology for the twenty-first century. *Nature 460*, 208-212.

Hartung, T. (2008a). Food for thought … on animal tests. *ALTEX 25*, 3-9.

Hartung, T. (2008b). Food for thought … on alternative methods for cosmetics safety testing. *ALTEX 25*, 147-162.

Hartung, T. and Koëter, H. (2008). Food for thought … on alternative methods for food safety testing, *ALTEX 25*, 259-264.

Hartung, T. and Leist, M. (2008). Food for thought … on the evolution of toxicology and phasing out of animal testing, *ALTEX 25*, 91-96.

Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human and Experimental Toxicology 25*, 497-513.

Hoffmann, S. and Hartung, T. (2005). Diagnosis: Toxic! – Trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Tox. Sci. 85*, 422-428.

Kim, W. C. and Maurborgne, R. (2002). *Blue Ocean Strategy*. Cambridge: Harvard Business. School Press.

Lantra (2006). The sector skills council for the environmental and land-based sector, sector skills agreement Animal Technology, Annex, www.lantra.co.uk

Leist, M., Hartung, T. and Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX 25*, 103-114.

Lichtenberg, F. (2007). Why longevity increased more in some

state than others? The role of medical innovation and other factors. *Medical Progress Report No. 4, Center for Medical Progress at the Manhattan Institute*, 1-31.

Lichtenberg, F. (2005a). The impact of new drug launches on longevity: evidence from longitudinal disease-level data from 52 countries, 1982-2001. *International Journal of Health Care Finance and Economics 5*, 47-73.

Lichtenberg, F. (2005b). The effect of access restrictions on the vintage drugs used by Medicaid enrollees. *American Journal of Managed Care 11, Spec. Issue*, SP7-SP13.

Majone, G. (2002). The precautionary principle and its policy implications. *Journal of Common Market Studies 40*, 89-109.

McSpotlight. http://www.mcspotlight.org/beyond/companies/bodyshop.html

Porter, M. (1990). The competitive advantage of nations. *Harvard Business Reviews (March/April),* 73-93.

Taylor, K., Gordon, N., Langley, G. and Higgins, W. (2008). Estimates for worldwide laboratory animal use in 2005. *ATLA 36*, 327-342.

Vogel, D. (1995). *Trading up – consumer and environmental regulation in a global economy*. Cambridge: Harvard University Press.

Wikipedia. Bodyshop. http://en.wikipedia.org/wiki/The_Body_Shop

Acknowledgements

This paper is based on dissertation of the degree of Bachelor of Arts (Honours) in International Business of A. A. B. The supervision by Elisabeth Farmaki and Prof. David Word is gratefully appreciated. A variant of this paper has been published: Bottini A. A. and Hartung T., Food for thought … on economics of animal testing, ALTEX, 2009.

Correspondence to

Thomas Hartung, MD PhD
Johns Hopkins University
Bloomberg School of Public Health
Doerenkamp-Zbinden Chair for Evidence-based Toxicology
615 N. Wolfe St.
Baltimore, MD, 21231
USA
e-mail: THartung@jhsph.edu

Theme 1:
Innovative technologies, concepts and approaches

Breakout Sessions

Session BS6: Non-vertebrate models

Medium- and High-Throughput Toxicity Screens Using *C. Elegans*

Windy A. Boyd and Jonathan H. Freedman

National Toxicology Program and National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, NC, USA

Summary

The use of the nematode Caenorhabditis elegans as an alternative model organism in medium-throughput chemical screening is being assessed. A high degree of evolutionary conservation exists between C. elegans and higher organisms for many signal transduction and stress-response pathways. In addition, homologues for many of the genes induced in response to toxicant exposure in vertebrates have been identified in C. elegans. For these reasons, it is likely that responses elicited in C. elegans will be applicable to understanding similar processes in humans. Methods have been developed to rapidly measure sub-lethal toxicity endpoints including growth, reproduction, feeding, and movement. These assays utilize COPAS Biosort flow cytometry and automated microscopic observation. In addition, new mathematical and statistical models have been developed to quantitatively measure the effects of chemicals on the nematode. Using medium-throughput technologies and statistical modeling, several chemical libraries have been tested including the NTP 1408 and the ToxCast 320.

Keywords: C. elegans, toxicity testing, growth, feeding, reproduction

1 Introduction

The Environmental Protection Agency (EPA) estimates that at least 10,000 chemicals are in need of prioritization (Dix et al., 2007). Due to the large number of chemicals that require toxicity testing, high- and medium-throughput assays are being developed as one option to prioritize testing for mammalian organisms. The National Toxicology Program (NTP) along with the EPA and the National Institutes of Health Chemical Genomics Center (NCGC) are currently investigating methods to rapidly screen the large number of chemicals using cell-based and biochemical assays. In addition, *in vivo* toxicological assays are being developed that could reduce the number of higher animals used by utilizing simpler organisms (e.g. invertebrates or microorganisms) (Collins et al., 2008). The results from these alternative assays will ultimately be used to inform the prioritization of further testing in traditional bioassays. While it is not possible to completely eliminate the use of higher organisms, using replacement strategies to prioritize chemicals prior to mammalian toxicological testing can be used to decrease the number of animals used, thus reducing the expense and time associated with toxicological testing.

The nematode *Caenorhabditis elegans* is a simple model organism that is being assessed for toxicological testing. *C. elegans* are small (1.5 mm) and easily cultured in the laboratory; thousands of individuals can be maintained on a single

agar-filled petri dish. Under standard culturing conditions, *C. elegans* populations consist mainly of self-fertilizing hermaphrodites, which allows for the maintenance of genetically-identical populations. An adult hermaphrodite is made up of 959 somatic cells, and the complete cell lineage has been mapped from the oocyte to adult. *C. elegans* develop from fertilized embryo to gravid adult through four larval stages, termed L1-L4 (Byerly et al., 1976). At 20°C, nematodes mature from embryos to adults in approximately 3 days, but can be cultured from 15-25°C to manipulate the length of the life cycle (Wood, 1988). *C. elegans* has been extensively used in biological research; it is one of the most thoroughly characterized multi-cellular organisms. For example, its genome has been fully sequenced (The *C. elegans* Sequencing Consortium, 1998) and databases describing gene function, cell lineage, and neuronal connectivities are readily available (Harris et al., 2004).

Several characteristics of *C. elegans* biology have shown that it can serve as a model organism in studies of human disease and toxicology. First, there is a high degree of evolutionary conservation between *C. elegans* and higher organisms. Homologues for many stress response proteins and regulatory pathways have been identified in *C. elegans* (Weston et al., 1989; Heschl and Baillie, 1990; Stringham et al., 1992; Freedman et al., 1993; Giglio et al., 1994; Wolf et al., 2008). In addition, homologues for many of the genes induced in response to toxicant exposure in vertebrates have been identified in *C. elegans*. Because of this homology, it is likely that responses elicited in *C. elegans* will be applicable to understanding similar processes in humans. As a result of the technical advantages associated with using *C. elegans* as a test organism and its conserved biology, several medium-throughput assays have been developed.

2 Assays

The rate-limiting step for most *C. elegans* toxicological studies has been the time required to accurately dispense exact numbers of *C. elegans* at specific developmental stages to exposure plates and then to quantitatively measure the responses of those nematodes after exposure. Automation of nematode handling and analyses provides the opportunity to rapidly screen large numbers of chemicals and endpoints. The COPAS Biosort flow cytometer system allows users to dispense specific numbers and stages of nematodes into individual wells of a 96 well plate while collecting data on the size and fluorescence characteristics of individual nematodes (Pulak, 2006). The Biosort detectors record two size characteristics of each animal: time of flight (TOF), which is the length of time a nematode takes to pass in front of a laser and is related to the length of the nematode; and extinction (EXT), which is the total amount of light blocked over the TOF and is related to the optical density of a nematode.

Figure 1 presents plots of typical *C. elegans* data acquired using the Biosort. The upper panel presents a scatter-plot of size data (TOF versus EXT) collected every 24 h for 72 h of an untreated population of *C. elegans*. At the start of the experiment (t = 0 h), first larval stage (L1) nematodes (*black dots*), which have relatively low TOF and EXT values, were added to media containing a sufficient supply of food. As the nematodes developed, length (TOF) and optical density (EXT) increased until 72 h, at which time gravid adults (high TOF/EXT) and offspring (low TOF/EXT) were observed (*blue dots*). An alternative method of presenting the same data is with frequency histograms (*lower panel*). In this case, a single measurement (i.e., EXT) is plotted versus the number of nematodes with specific EXT values.

Assays for three Biosort-based, medium-throughput endpoints – growth, reproduction, and feeding – have been developed. In the medium-throughput assays, each endpoint was designed to measure toxicity at specific developmental stages and after various exposure times and was previously shown to be affected by toxicants in low-throughput experiments (Anderson et al., 2001; Boyd et al., 2003). Using the Biosort, nematodes are dispensed into 96-well plates containing food (*E. coli*) and varying concentrations of the chemical of interest. Following

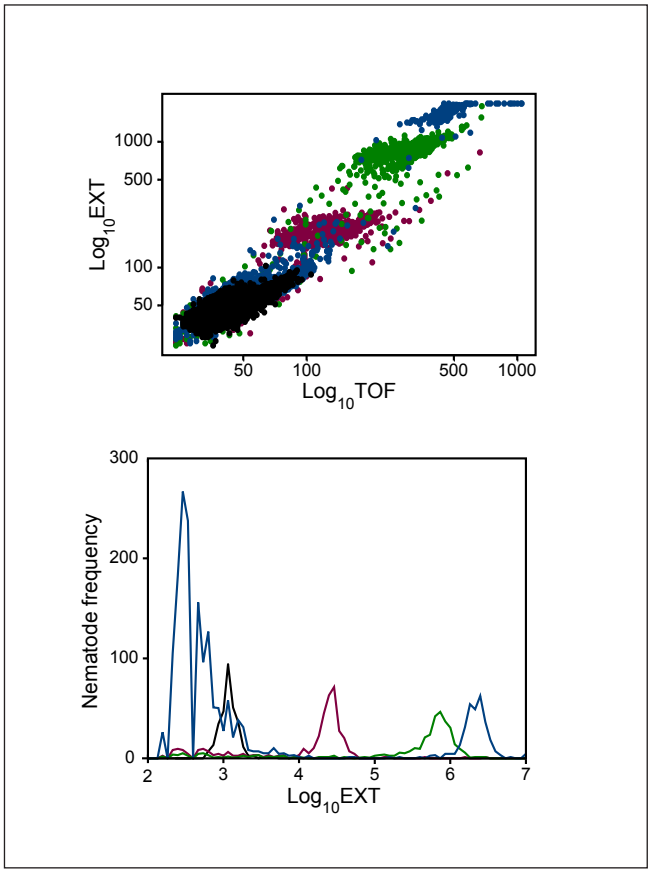


Fig. 1: *C. elegans* growth from L1 to adult
Untreated L1 nematodes were incubated at 20°C and sampled at 0 h (*black*), 24 h (*red*), 48 h (*green*), and 72 h (*blue*). *Upper panel*, scatter-plot of optical density (log(EXT)) versus length (log(TOF)). Each point corresponds to an individual nematode. *Lower panel*, frequency distributions of log(EXT) versus numbers of nematodes. At 72 h, adult nematodes (high EXT, TOF) and their offspring (low EXT, TOF) were observed.

the appropriate incubation times, nematodes are aspirated from the wells using the Biosort, which simultaneously records TOF, EXT, and fluorescence levels of individual nematodes. Brief descriptions of the assays are presented below:

Feeding

C. elegans feed via coordinated contractions of two pharyngeal bulbs that push bacterial suspension to the back of the pharynx, concentrate the bacteria and pump it into the intestine, and then expel excess liquid out through the mouth (Avery and Shtonda, 2003). These rhythmic contractions are controlled by a self-contained nervous system that can be affected by environmental conditions such as food availability and toxicant exposure (Avery and Horvitz, 1990; Boyd et al., 2003). Although pharyngeal pumping rates of nematodes can easily be observed with a dissecting microscope, each animal must be monitored individually (Avery, 1993). A feeding assay that uses the Biosort to quantify food ingestion by thousands of nematodes in minutes has been developed (Boyd et al., 2007). Adult nematodes are exposed to toxicants for 24 h and then to red fluorescent microspheres for 15 min. Nematode size-corrected red fluorescence as measured by the Biosort is used to calculate feeding activity: the greater the fluorescence the faster the feeding rate. Figure 2 illustrates how varying concentrations of cocaine base decrease feeding activity.

Reproduction

C. elegans egg-laying rates and embryonic survival are commonly observed phenotypes that may be affected by environmental conditions including salt concentration and chemical exposures (Horvitz et al., 1982; Dhawan et al., 1999; Kim et al., 2001). Sexual development in *C. elegans* hermaphrodites begins with sperm production in L4 larvae followed by oocyte production and fertilization in adults (Wood, 1988). Egg-laying occurs via the vulva and the surrounding neuromuscular system (Trent et al., 1983). In low-throughput fashion, the number of laid embryos and their survival from a few adults can be measured on agar plates using a dissecting microscope. This process requires several hours of microscope observation to complete a few plates. In addition, it can be difficult to control chemical exposures using solid media. The medium-throughput reproduction assay uses the Biosort to load L4 animals into 96-well plates. The nematodes are then exposed to toxicants for 48 h, which yields a population of adults, embryos, and L1-L2 larvae. The Biosort aspirates and measures the number of adults and their offspring in several minutes; total nematode counts are plotted against chemical concentration. The effects of methyl mercury on reproduction are presented in Figure 3.

Growth

C. elegans develop from fertilized embryo through four distinct larval stages, growing in burst between stages, before maturing to adult in about 3 days. Thus it is possible to test for the effects of chemicals on growth and development in a relatively short time period. In the medium-throughput assay, L1 larvae are loaded into 96-well plates and exposed to varying concentrations of chemical for 48 h. At the end of this incubation time, untreated animals have reached the L4 stage (Boyd et al., 2009). Exposure time is limited to 48 h to avoid offspring, which have

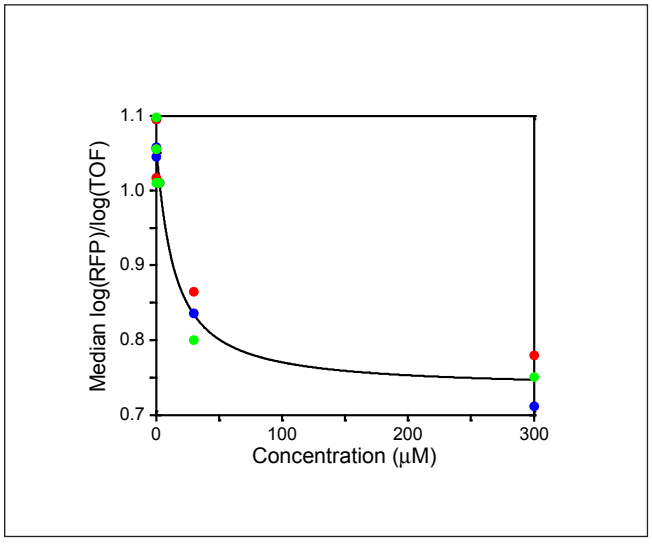


Fig. 2: Effects of cocaine base on *C. elegans* feeding
For each of three replicate experiments (shown as blue, red, and green), groups of 25 adult *C. elegans* were exposed to cocaine (0, 3, 10, 30, 100, 300 µM) for 24 h. Medians of log(fluorescence)/log(TOF) values are plotted and the data fit to the Hill equation.

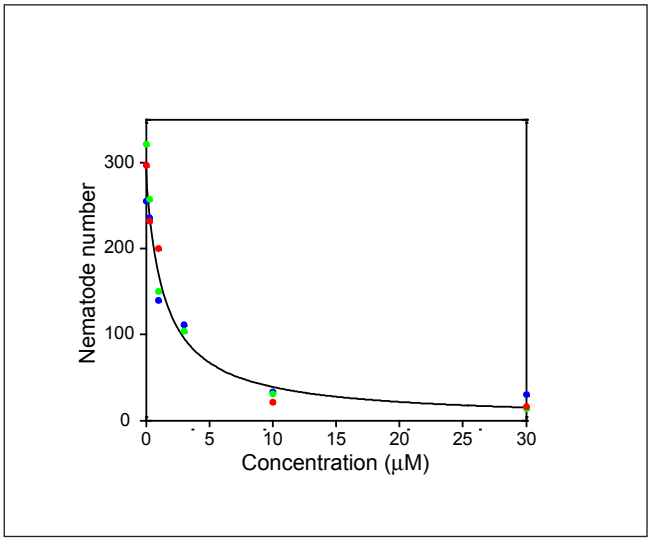


Fig. 3: Effects of methyl mercury on *C. elegans* reproduction
For each of three replicate experiments (shown as blue, red, and green), groups of five L4 nematodes were exposed to methyl mercury (0, 0.3, 1, 3, 10, 30 µM) for 48 h. The numbers of observations (larvae and embryos) were fit to the Hill equation.

been observed after incubation times of 60 h or longer. After exposure, the size distribution of the nematodes in the sample is measured using the Biosort. Log(EXT) or log (TOF) values are then used to calculate growth rates of the nematode size distributions (Smith et al., 2009). The concentration-dependent effect of chlorpyrifos on *C. elegans* growth and development is presented in Figure 4.

The 48-h growth assay is a promising method for high-throughput screening of chemical libraries using *C. elegans*. Through collaboration with the U. S. EPA's ToxCast program (http://www.epa.gov/NCCT/dsstox/sdf_toxcst.html), a library of over 300 compounds has been screened using the growth assay. Statistical analysis tools have been developed to summarize the results of these experiments. Preliminary results indicate that a large number of the compounds, which are mainly pesticide active ingredients, significantly affect *C. elegans* growth and development. In the future, the results from this screen will be compared to the results from other alternative tests including *in vitro* systems and zebrafish.

Motion tracking

C. elegans move in sinusoidal waves across solid surfaces, or thrash from side-to-side in liquid culture (Bargmann, 1993; Miller et al., 1996). Both phenotypes were shown to be affected by exposure to anthelminthic chemicals known to cause decreased feeding and paralysis (Bull et al., 2007). Several systems exist for tracking individual *C. elegans* on agar surfaces (Feng et al., 2004; Cronin et al., 2005; Hoshi and Shingai, 2006) or in liquid (Buckingham and Sattelle, 2009). Using lower magnification, the centroids of hundreds of nematodes can be tracked real-time using a video camera interfaced with a computer tracking system (Williams and Dusenbery, 1990).

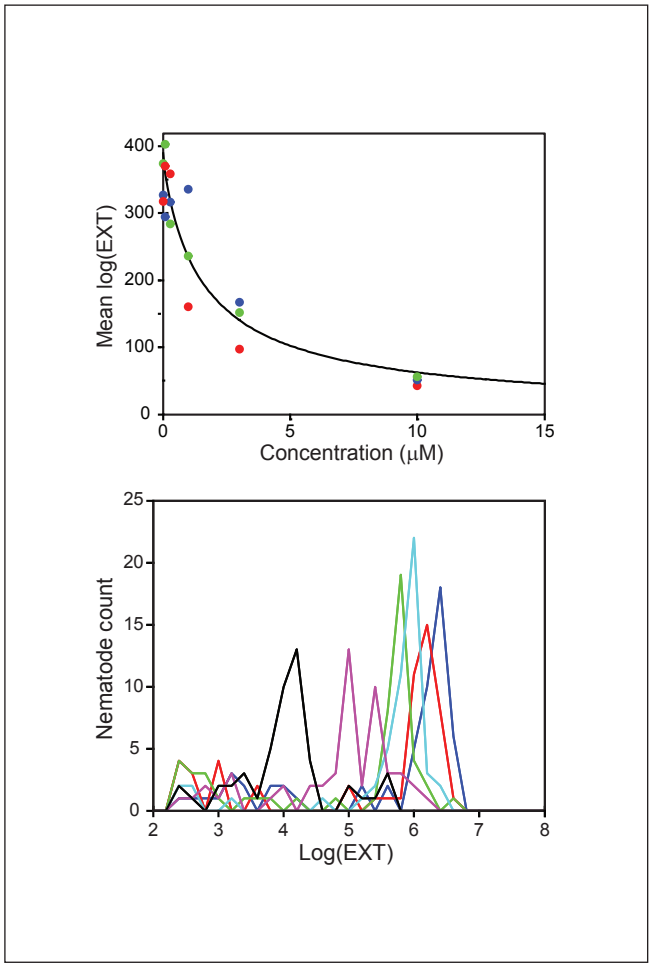


Fig. 4: Effects of chlorpyrifos on *C. elegans* growth
For each of three replicate experiments (shown as blue, red, and green), groups of 50 L1 nematodes were exposed to chlorpyrifos (0, 0.1, 0.3, 1.0, 3.0, and 10 µM) for 48 h. Upper panel, observations measured and estimated means fit to Hill equation. *Lower panel*, modeled frequency histograms showing the effects of different concentrations of chlorpyrifos on nematode size distributions (log(EXT)) for one of the replicates.

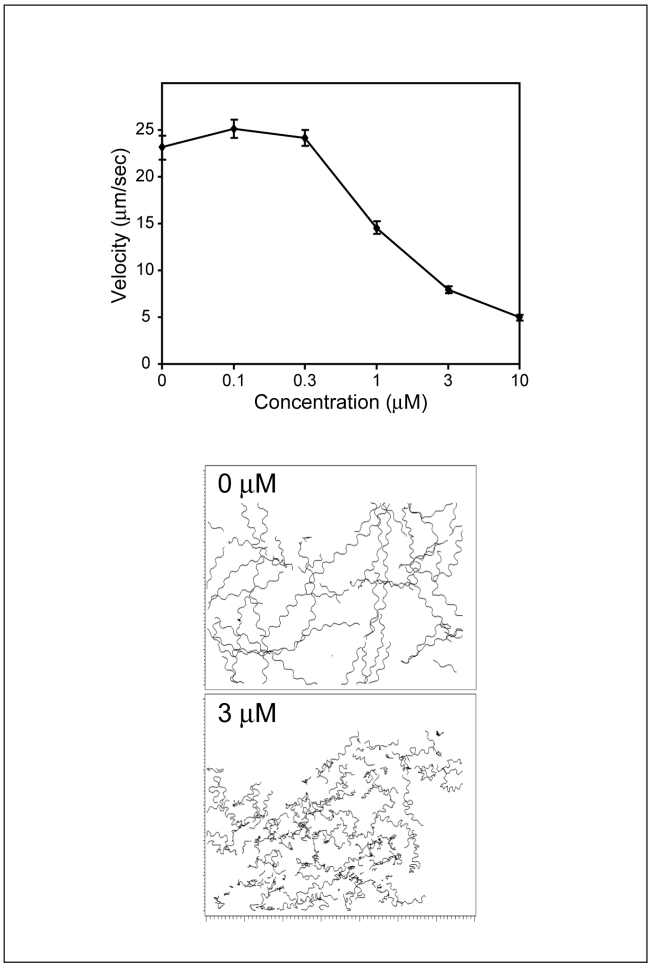


Fig. 5: Effects of chlorpyrifos on *C. elegans* movement
For each of three replicate experiments, groups of 40 L4 nematodes were tracked after exposure to chlorpyrifos (0.1, 0.3, 1, 3, 10 µM) for 4 h. *Upper panel*, average velocity of *C. elegans* motion, mean ± s.e. *Lower panel*, computer generated tracks for control nematodes (*top*) and those exposed to 3 µM chlorpyrifos (*bottom*).

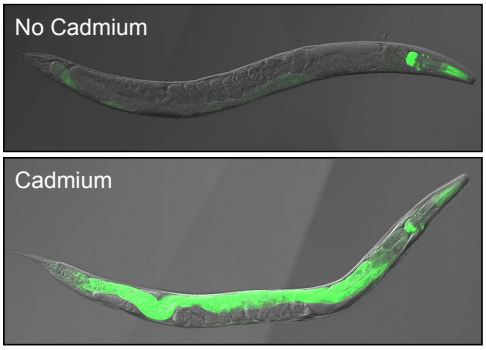


Fig. 6: Effect of cadmium-exposure on *mtl-1*: GFP transgenic *C. elegans*
Transgenic nematode expressing GFP under the control of the *C. elegans* metallothionein promoter (*mtl-1*) were grown in the absence (*upper panel*) or presence (*lower panel*) of 100 µM cadmium for 24 h. Constitutive *mtl-1* transcription is observed in the pharynx of the nematodes, whole metal-inducible transcription occurs in the nematode intestine.

Our motion tracking assay uses a similar computer tracking system, which consists of an inverted fluorescence microscope, a CCD video camera, and an automated stage. Using a transgenic reporter strain in which the pharynx is fluorescently-labeled (Fay et al., 2003), head movements of approximately 40 L4s are simultaneously tracked following 4-h chemical exposures. The effects of chemicals on *C. elegans* motion are quantitatively assessed using motion tracking software originally developed to study sperm motility. Concentration-dependent effects were observed on velocity and track shapes after exposure to chlorpyrifos (Fig. 5).

Reporter gene expression
Currently the ability to visually monitor changes in transcription in *C. elegans* using transgenic nematodes is being explored as a method to rapidly measure toxicant-induced changes in gene expression. Microinjection of DNA into the *C. elegans* gonad allows for germline transformation generating transgenic lines that carry fluorescent reporter genes fused to the promoter region of a gene of interest (Chalfie et al., 1994; Mello and Fire, 1995; Boulin et al., 2006). Because *C. elegans* are transparent, spatial and temporal expression of fluorescent reporter transgenes can be observed *in vivo*. The Profiler II, a module of the COPAS Biosort, measures three channels of fluorescence and EXT in optical slices along the length of individual nematodes (Dupuy et al., 2007). Profiles are generated that depict the level of fluorescence and EXT across the length of nematodes.

To develop assays that can be used to rapidly and quantitatively measure the effects of toxicants on transcription, green fluorescent protein (GFP) based transgenic *C. elegans* strains are being generated. The transgenic strains fall into two categories. The first includes promoter fusions of genes that have been shown to be stress responsive in previous studies. By focusing on well-conserved signaling pathways (National Research Council, 2000), the responses of nematode genes to chemicals can be investigated *in vivo*. Figure 6 illustrates the effects of metal exposure on metallothionein transcription in transgenic *C. elegans*. The second category includes transgenic strains in which specific groups of neurons are labeled with GFP. Currently, two strains that label the dopaminergic (*dat-1::GFP*) (Nass, 2002) and the GABAergic (*unc-25::GFP*) (Eastman et al., 1999) neurons are being tested. For both groups of transgenic nematodes, green fluorescence profiles generated using the Biosort will be used to monitor the effects of chemicals in specific tissues and at various developmental stages.

3 Conclusion

Several medium- and high-throughput assays have been developed using *C. elegans* as an alternative toxicological test organism. However, the need to define the relationship between the results obtained from any high-throughput screen to human health and safety remains. The US EPA's ToxCast program is investigating this link by screening a defined set of three hundred compounds with available whole animal toxicity data, using mainly *in vitro* cell systems and, to a lesser extent, alternative whole organism models. Comparison of the *C. elegans* growth assay results with these systems is a critical next step in the evaluation of the usefulness of *C. elegans* in the prioritization of further chemical testing.

References

Anderson, G. L., Boyd, W. A. and Williams, P. L. (2001). Assessment of sublethal endpoints for toxicity testing with the nematode Caenorhabditis elegans. *Environ. Toxicol. Chem. 20*, 833-838.

Avery, L. and Shtonda, B. B. (2003). Food transport in the C. elegans pharynx. *J. Exp. Biol. 206*, 2441-2457.

Avery, L. (1993). The genetics of feeding in Caenorhabditis elegans. *Genetics 133*, 897-917.

Avery, L. and Horvitz, H. R. (1990). Effects of starvation and neuroactive drugs on feeding in Caenorhabditis elegans. *J. Exp. Zool. 253*, 263-270.

Bargmann, C. I. (1993). Genetic and cellular analysis of behavior in C. elegans. *Annual review of neuroscience 16*, 47-71.

Boulin, T., Etchberger, J. F. and Hobert, O. (2006). Reporter gene fusions. *WormBook*, 1-23.

Boyd, W. A., Smith, M. V., Kissling, G. E. et al. (2009). Application of a mathematical model to describe the effects of chlorpyrifos on Caenorhabditis elegans development. *PLoS-ONE. 4*, e7024.

Boyd, W. A., McBride, S. J. and Freedman, J. H. (2007). Effects of genetic mutations and chemical exposures on

Caenorhabditis elegans feeding: evaluation of a novel, high-throughput screening assay. *PLoS ONE. 2*, e1259.

Boyd, W. A., Cole, R. D., Anderson, G. L. et al. (2003). The effects of metals and food availability on the behavior of Caenorhabditis elegans. *Environ. Toxico.l Chem. 22*, 3049-3055.

Buckingham, S. D. and Sattelle, D. B. (2009). Fast, automated measurement of nematode swimming (thrashing) without morphometry. *BMC Neurosci. 10*, 84.

Bull, K., Cook, A., Hopper, N. A. et al. (2007). Effects of the novel anthelmintic emodepside on the locomotion, egg-laying behaviour and development of Caenorhabditis elegans. *Int. J. Parasitol. 37*, 627-636.

Byerly, L., Cassada, R. C. and Russell, R. L. (1976). The life cycle of the nematode Caenorhabditis elegans. I. Wild-type growth and reproduction. *Dev. Biol. 51*, 23-33.

Chalfie, M., Tu, Y., Euskirchen, G. et al. (1994). Green fluorescent protein as a marker for gene expression. *Science 263*, 802-805.

Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science 319*, 906-907.

Cronin, C. J., Mendel, J. E., Mukhtar, S. et al. (2005). An automated system for measuring parameters of nematode sinusoidal movement. *BMC Genet. 6*, 5.

Dhawan, R., Dusenbery, D. B. and Williams, P. L. (1999). Comparison of lethality, reproduction, and behavior as toxicological endpoints in the nematode Caenorhabditis elegans. *J. Toxicol. Environ. Health A. 58*, 451-462.

Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci. 95*, 5-12.

Dupuy, D., Bertin, N., Hidalgo, C. A. et al. (2007). Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. *Nat Biotechnol. 25*, 663-668.

Eastman, C., Horvitz, H. R. and Jin, Y. (1999). Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein. *J. Neurosci. 19*, 6225-6234.

Fay, D. S., Large, E., Han, M. et al. (2003). lin-35/Rb and ubc-18, an E2 ubiquitin-conjugating enzyme, function redundantly to control pharyngeal morphogenesis in C. elegans. *Development. 130*, 3319-3330.

Feng, Z., Cronin, C. J., Wittig, J. H., Jr. et al. (2004). An imaging system for standardized quantitative analysis of C. elegans behavior. *BMC Bioinformatics 5*, 115.

Freedman, J. H., Slice, L. W., Dixon, D. et al. (1993). The novel metallothionein genes of Caenorhabditis elegans. Structural organization and inducible, cell-specific expression. *J. Biol. Chem. 268*, 2554-2564.

Giglio, A. M., Hunter, T., Bannister, J. V. et al. (1994). The copper/zinc superoxide dismutase gene of Caenorhabditis elegans. *Biochem. Mol. Biol. Int. 33*, 41-44.

Harris, T. W., Chen, N., Cunningham, F. et al. (2004). WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Res. 32 Database issue*, D411-417.

Heschl, M. F. and Baillie, D. L. (1990). The HSP70 multigene family of Caenorhabditis elegans. *Comp. Biochem. Physiol. [B]. 96*, 633-637.

Horvitz, H. R., Chalfie, M., Trent, C. et al. (1982). Serotonin and octopamine in the nematode Caenorhabditis elegans. *Science 216*, 1012-1014.

Hoshi, K. and Shingai, R. (2006). Computer-driven automatic identification of locomotion states in Caenorhabditis elegans. *J. Neurosci. Methods 157*, 355-363.

Kim, J., Poole, D. S., Waggoner, L. E. et al. (2001). Genes affecting the activity of nicotinic receptors involved in Caenorhabditis elegans egg-laying behavior. *Genetics 157*, 1599-1610.

Mello, C. and Fire, A. (1995). DNA transformation. *Methods Cell Biol. 48*, 451-482.

Miller, K. G., Alfonso, A., Nguyen, M. et al. (1996). A genetic selection for Caenorhabditis elegans synaptic transmission mutants. *Proc. Natl. Acad. Sci. USA 93*, 12593-12598.

Nass, R., Hall, D. H., Miller, D. M., 3rd et al. (2002). Neurotoxin-induced degeneration of dopamine neurons in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. USA 99*, 3264-3269.

National Research Council (2000). Appendix C: Signaling Pathways. In BoESaT Committee on Developmental Toxicology, Commission on Life Sciences, and National Research Council, eds., *Scientific Frontiers in Developmental Toxicology and Risk Assessment* (296-308). Washington, DC: National Academy Press.

Pulak, R. (2006). Techniques for analysis, sorting, and dispensing of C. elegans on the COPAS flow-sorting system. *Methods Mol. Biol. 351*, 275-286.

Smith, M. V., Boyd, W. A., Kissling, G. E. et al. (2009). A discrete time model for the analysis of medium-throughput C. elegans growth data. *PLoSONE. 4*, e7018.

Stringham, E. G., Jones, D. and Candido, E. P. (1992). Expression of the polyubiquitin-encoding gene (ubq-1) in transgenic Caenorhabditis elegans. *Gene 113*, 165-173.

The C. elegans Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science 282*, 2012-2028.

Trent, C., Tsuing, N. and Horvitz, H. R. (1983). Egg-laying defective mutants of the nematode Caenorhabditis elegans. *Genetics 104*, 619-647.

Weston, K., Yochem, J. and Greenwald, I. (1989). A Caenorhabditis elegans cDNA that encodes a product resembling the rat glutathione S-transferase P subunit. *Nucleic Acids Res. 17*, 2138.

Williams, P. L. and Dusenbery, D. B. (1990). A promising indicator of neurobehavioral toxicity using the nematode Caenorhabditis elegans and computer tracking. *Toxicol. Ind. Health 6*, 425-440.

Wolf, M., Nunes, F., Henkel, A. et al. (2008). The MAP kinase

JNK-1 of Caenorhabditis elegans: location, activation, and influences over temperature-dependent insulin-like signaling, stress responses, and fitness. *J. Cell Physiol. 214,* 721-729.

Wood, W. B. (1988). Introduction to C. elegans Biology. In W. B. Wood (ed.), *The Nematode Caenorhabditis elegans*. (1-16). N.Y.: Cold Spring Harbor Laboratory.

Acknowledgments

The authors would like to thank Marjolein Smith and Grace Kissling for analysis of the Biosort and motion tracking experiments, respectively. We thank Julie Rice, Daniel Snyder and Brooke Tvermoes for technical assistance in generating the data presented in this manuscript. This work was supported, in part, by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01ES102045), and the National Toxicology Program (Z01ES102046).

Correspondence to

Jonathan H. Freedman
Laboratory of Toxicology, and Pharmacology
National Institute of Environmental Health Sciences,
Mail Drop E1-05, P.O. Box 12233,
111 T.W. Alexander Drive
Research Triangle Park,
NC, 27709 USA
e-mail: freedma1@niehs.nih.gov

Validation of Innovative Technologies and Strategies for Regulatory Safety Assessment Methods: Challenges and Opportunities

William S. Stokes[1] and Marilyn Wind[2]

[1]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA; [2]US Consumer Product Safety Commission, Bethesda, MD, USA NC, USA

Summary

Advances in science and innovative technologies are providing new opportunities to develop test methods and strategies that may improve safety assessments and reduce animal use for safety testing. These include high throughput screening and other approaches that can rapidly measure or predict various molecular, genetic, and cellular perturbations caused by test substances. Integrated testing and decision strategies that consider multiple types of information and data are also being developed. Prior to their use for regulatory decision-making, new methods and strategies must undergo appropriate validation studies to determine the extent that their use can provide equivalent or improved protection compared to existing methods and to determine the extent that reproducible results can be obtained in different laboratories. Comprehensive and optimal validation study designs are expected to expedite the validation and regulatory acceptance of new test methods and strategies that will support improved safety assessments and reduced animal use for regulatory testing.

Keywords: validation, safety testing, integrated testing strategies, integrated decision strategies

1 Introduction

Safety assessment methods are necessary to determine if new chemicals and products are safe or if they may adversely affect the health of people, animals, and the environment. Advances in science and innovative technologies are providing new opportunities to develop test methods and strategies that may improve safety assessments and reduce animal use for safety testing. Research continues to improve our understanding of the molecular and cellular alterations by which chemical exposures can cause or contribute to injury or disease. High throughput screening, toxicogenomics, and other approaches can now be used to rapidly measure many of the molecular, genetic, and cellular perturbations caused by chemicals. Robot operated laboratories can rapidly generate vast amounts of *in vitro* data for thousands of chemicals (Michael et al., 2008). Analysis of this data is expected to help identify panels of *in vitro* biomarkers that can be used to help assess chemical toxicity. Integrated testing strategies that consider information and data from such assays and various test methods are also being developed (Stokes, 2007).

Prior to their use for regulatory decision-making, new methods and strategies must undergo appropriate validation studies to determine if they can provide equivalent or improved protection compared to existing methods and to determine if reproducible results can be obtained in different laboratories (ICCVAM, 1997, 2003; OECD, 2005). Validation studies must be carefully designed to optimize test methods and to ensure that they generate adequate data for decisions on their regulatory acceptability (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Adequate validation will expedite the acceptance and use of new test methods and strategies that support improved safety

assessments and contribute to reduced animal use for regulatory testing. This paper will discuss emerging innovative technologies, concepts, and approaches applicable to regulatory safety assessments, and opportunities and challenges for their scientific validation.

2 Changing the paradigm of toxicity testing

Two recent reports have proposed using advances in science and technology to change the current paradigm of toxicity testing. These include the 2004 National Toxicology Program Roadmap, and the 2007 National Research Council (NCR) publication, *Toxicology in the 21st Century, A Vision and a Strategy* (NTP, 2004; NRC, 2007a). The NTP Roadmap envisions moving from toxicology studies that depend on observing the actual adverse outcome from chemical exposures, such as cancer and birth defects in animal models, to one based on understanding and detecting cellular and molecular perturbations in simpler models such as cell cultures and lower organisms that are predictive of these eventual adverse outcomes. To implement this vision, the NTP plan is to develop and validate improved testing methods and to ensure, where feasible, that such methods provide for the reduction, refinement, and replacement of animals.

The NTP report emphasizes that activities and assays developed under the NTP Roadmap will be done in cooperation and consultation with the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to maximize their value to regulatory agencies (NTP, 2004). ICCVAM is a U.S. interagency committee composed of 15 research and regulatory agencies that is charged with evaluating the scientific validity of new, revised, and alternative test methods proposed for regulatory testing (ICCVAM, 2003; Stokes and Schechtman, 2007).

The 2007 NRC report similarly envisions future testing based on an understanding of key toxicity pathways at the cellular and molecular levels and using predictive high throughput assays to detect the potential for chemicals to sufficiently alter these pathways to cause injuries or disease. The report states that the use of a comprehensive array of *in vitro* tests to identify relevant biological perturbations based on human biology could eventually eliminate the need for whole-animal testing and provide a stronger mechanistically based approach for environmental decision-making. However, a 2009 NRC report states that the realization of the promise of this vision is at least a decade away (NRC, 2009).

3 Emerging science and technology

New scientific advances and innovative technologies are now available to help develop future testing methods and strategies outlined in the NRC and NTP reports. These include high throughput screening, toxicogenomics, and computational modeling approaches.

High throughput screening involves the use of computerized robots to conduct the laboratory procedures necessary to study hundreds of compounds per day in multiple *in vitro* assays. The National Chemical Genomics Center at the National Human Genome Research Institute has a laboratory where such studies are conducted (Michael et al., 2008). In collaboration with the NTP and EPA, the lab is now conducting quantitative HTS using fifteen concentrations of each chemical (Collins et al., 2008). The lab uses 1536-well plates, which have a net testing capability of 1504 individual chemicals. Over 100,000 concentration response profiles can be generated per week. These profiles are then evaluated to determine if *in vitro* biomarker alterations are associated with known adverse health effects. Bioinformatics techniques will be used to identify complex relationships between different types of biological responses that may provide insights into critical toxic pathways (Schmidt, 2009).

Another NRC report published in 2007 addressed the application of toxicogenomic technologies to predictive toxicology and risk assessment (NRC, 2007b). Toxicogenomics is defined as the application of genomic technologies to study the adverse effects of environmental and pharmaceutical chemicals on human health and the environment. These technologies include genetics, genome sequence analysis, gene expression profiling, proteomics, metabolomics, and other related approaches. These are used to measure chemical-specific perturbations on expression patterns of genes, proteins, and metabolites in cells, tissues, and organisms. Such technologies are being investigated for their potential to improve the prediction of safety or potential hazards of chemicals to human health.

Computational modeling is being applied to estimate the absorption, distribution, metabolism, and excretion of chemicals (ADME) (NRC, 2007a). These models seek to estimate the relationship between the dose or amount of chemical exposure via oral, dermal, or inhalation routes, and the concentration of chemical that reaches individual cell types in various critical organs and tissues. These estimates will be essential for non-animal estimates of exposure levels that are safe and those that are likely to be associated with toxic effects. It is also important that data used to construct computational models is of high quality and derived from adequately designed studies.

4 Application of new science and technology to regulatory decision-making

As emerging scientific advances provide insights into the pathways and mechanisms of chemical toxicity, the National Toxicology Program and other public health agencies seek to apply this information so that it can be used to improve public health. Several recent and planned activities and initiatives have and will continue to investigate potential applications for public health decision-making. For example, at the request of the National Institute of Environmental Health Sciences, the National Academies recently formed a Standing Committee on the Use of Emerging Science for Environmental Health Decisions (NAS, 2009). The committee is charged with facilitating communication among government, industry, environmental groups, and the academic community about scientific advances that may be used in the identification, quantification, and

control of environmental impacts on human health. The topics covered will build on recent NRC reports on toxicity testing and toxicogenomics and will explore new developments in toxicology, molecular biology, bioinformatics, and related fields (NRC, 2007a, 2007b). Three workshops have been or will be held in the near future. (Fig. 1)

Mechanistic toxicity data from animal studies and humans are necessary to link *in vitro* pathway data to adverse health effects. To address this need, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Toxicological Methods (ICCVAM) recently convened an *International Workshop on Acute Chemical Safety Testing – Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations* (NICEATM, 2008). The primary goals of the workshop were to identify approaches for collecting additional mechanistic data from current *in vivo* testing that would support the development of predictive mechanism-based *in vitro* alternative models and that could also be used to identify earlier more humane endpoints.

5 Validation and acceptance of test methods based on new science and technology

In the United States, Federal laws require that new safety assessment methods proposed for regulatory safety assessment decisions must be determined to be sufficiently valid and acceptable for their proposed use (USC, 2000). National and internationally harmonized principles for validation and regulatory acceptance are available (ICCVAM, 1997; OECD, 2005). Determination of validity involves assessing the accuracy and reliability of the test method for a specific proposed purpose (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Accuracy assessments typically characterize sensitivity, specificity, and false positive and negative rates compared to existing reference data. Regulatory acceptance decisions involve reviewing the validation database to determine if the proposed use of the method for decision-making will provide equivalent or improved protection compared to existing methods (USC, 2000). Reliability assessments determine if reproducible results can be obtained in different laboratories when using the proposed standardized test method protocol.

National and International authorities have agreed on validation and regulatory acceptance criteria for new, revised, and alternative test methods (ICCVAM, 1997; OECD, 2005). These are general criteria that should be appropriately addressed when considering the validity of test methods. The published criteria emphasize that flexibility is essential in interpreting and applying the criteria and that the extent that they will need to be addressed will depend on the intended purpose and nature of the proposed test (ICCVAM, 1997; OECD, 2005).

6 Validation of new science and technologies: challenges

New test methods based on scientific advances and technologies are likely to initially have limitations. Early definition of a test method's limitations can contribute to more efficient validation for the initial proposed uses and aid in identifying directed research to discover ways to address defined limitations. In some cases, test methods may be limited in terms of the physical and chemical properties of substances that can be tested. For example, the current NCGC HTS protocol is only capable of testing substances soluble in DMSO, so those that are not soluble cannot be adequately evaluated in this test system. The highest concentration that can be achieved in a test system may be limited by solubility in the required vehicle, which may not be sufficient for regulatory testing purposes. A significant limitation of most current *in vitro* testing methods is their inability to determine if there is metabolic activation of the substance to a more toxic or less toxic moiety. Additionally, there are still challenges in accurately estimating the toxicokinetics associated with specific exposures by various routes and the concentrations that will result in various critical target tissues. These limitations present challenges that will need to be

2009 Workshops:
National Academies' Standing Committee on Use of Emerging Science for Environmental Health Decisions

- Use of Emerging Science and Technologies to Explore Epigenetic Mechanisms Underlying the Developmental Basis for Disease, July 30-31, 2009. http://dels.nas.edu/envirohealth/epigenetic.shtml
- Computational Toxicology: From Data to Analyses to Applications, September 21-22, 2009. http://dels.nas.edu/envirohealth/comptox.shtml
- The Exposome: A Powerful Approach for Evaluating Environmental Effects on Chronic Diseases, February 25-26. 2010. http://dels-old.nas.edu/envirohealth/exposome.shtml

Fig. 1: 2009 Workshops: National academies' standing committee on use of emerging science for environmental health decisions

addressed in order to fully move away from the use of intact living organisms for safety assessments.

Another significant challenge for evaluating the validity of new testing methods and strategies for human health safety assessments is the availability of high quality reference data from humans. For ethical reasons, most existing reference data is from animal studies. However, for some toxicity endpoints such as allergic contact dermatitis (ACD), there is considerable human testing data and experience from occupational and consumer exposures (ICCVAM, 1999; Basketter et al., 2007). These human data supported the validity of a new animal model for ACD testing that has many scientific and animal welfare advantages compared to the traditional animal tests for ACD. Improved ways of obtaining data regarding the health effects from human exposures and ways to more accurately extrapolate exposures and effects from animal models to humans are needed to help validate new test methods.

7 Validation of new science and technologies: opportunities

Early consideration of the potential application of new technologies for regulatory testing during research and development stages provides an important opportunity to incorporate efforts that will support the validation of eventual test methods. Early standardization and use of harmonized technology platforms for approaches such as toxicogenomics and HTS will allow for data from different studies to be compared and combined for data analyses. This will also help minimize experimental variables, aid in achieving more reproducible results across labs, and contribute to achieving a high signal to noise ratio. For example, a recent workshop developed recommendations for the standardization and validation of toxicogenomic-based platforms that will be evaluated for their potential use for safety assessments (Corvi et al., 2006).

There is also an opportunity to develop data during research and development that may contribute to the validation database supporting the validity of proposed test methods and approaches. Several critical factors should be considered during research, development, translation, and validation stages for new technologies. These include selection of reference substances, dose/concentration selection procedures, defining the test method purpose and potential regulatory use, and phased validation studies to develop an optimized test method protocol.

Reference Substances: Reference substances selected for evaluation of the new technology should have high quality data available from existing reference test methods or the species of interest for the toxicity endpoint under evaluation (ICCVAM, 2003; Stokes and Schechtman, 2007). Selection of reference data should generally address established selection criteria for reference substances (Stokes and Schechtman, 2007), which include:

- Represent the dynamic range of responses possible for the toxicity endpoint of interest and the range of potential responses that can be measured in the test system
- Represent the range of physical and chemical properties of substances for which the test system is proposed to be capable of testing (e.g., physical form, water solubility, pH, volatility)
- Represent the range of relevant biologic properties, as appropriate (e.g. peptide reactivity, mutagenicity)
- Represent the range of chemistry of substances proposed for evaluation in the new test method (i.e., chemical classes)
- Represent the range of known or suspected modes or mechanisms of action for the toxicity measured or predicted by the test method
- Supported by existing high quality data from the currently accepted test method, and where possible, data and/or experience in the species of interest (e.g. for humans, ethical test data or accidental exposures information)
- Readily available from commercial sources
- Avoidance of chemicals with excessive occupational or environmental hazard, if feasible.

Dose or concentration-setting procedures: The basis and procedures for determining the highest dose or concentration that will be tested should be clearly stated. For animal-based tests this is normally based on the highest minimally toxic dose (MTD) or a defined upper limit dose. For *in vitro* tests, this is normally the highest soluble concentration, the highest non-cytotoxic concentration, or a defined upper concentration based on the highest potential exposure that might occur (Stokes, 2006).

Test method purpose and regulatory applicability: The specific proposed purpose of the test method and the proposed or potential use for regulatory decision-making in the context of current or anticipated regulatory requirements should be clearly defined (ICCVAM, 1997; OECD, 2005; Stokes and Schechtman, 2007). Proposed uses may range from serving as a complete replacement for a current existing test method to providing adjunct mechanistic data for weight-of-evidence decisions. For test methods proposed for use in chemical screening, the specific decisions that can be made with each possible test result must be clearly defined. For example, a positive result in a screening method might be used as the basis for hazard classification and labeling, while negative results associated with sufficient uncertainty may require further testing. Screening tests may also be proposed for prioritization decisions on whether further testing will or should be conducted. In such casest the uncertainty of the prediction of potential hazard or safety for a specific toxic endpoint should be characterized and transparent for the prioritization decision.

Phased validation studies: optimizing the test method protocol: Recent *in vitro* validation studies managed by NICEATM have shown that a validation study design consisting of several sequential progressive phases with coded chemicals was an efficient means of optimizing the test method protocol and minimizing intra- and inter-laboratory variation (Fig. 2) (ICCVAM, 1997; OECD, 2005; Paris et al., 2006; Stokes et al., 2007; Stokes and Schechtman, 2007). The initial laboratory evaluation phase involves a series of multiple testing with positive and vehicle

controls, with cycles of protocol modifications until all labs are able to obtain sufficiently reproducible results. Two stages of the second phase each test a small number of chemicals representative of the potential range of responses and vehicle solubility. After each phase, excessive experimental variation and discordance are stage and appropriate modifications made to the protocol. Retesting is conducted where substantive modifications are deemed necessary to confirm the effectiveness of these changes for obtaining consistent results. The last phase uses the final optimized protocol to generate data to assess accuracy and reliability.

8 Validation of integrated testing strategies

Integrated testing strategies involve considering all available information and data to determine if decisions can be made about the safety or hazard of substances in a stepwise or tiered manner. These are usually designed to minimize or avoid the use of animals. If there is not sufficient information and data for a decision at the initial level or tier, then testing proceeds to the next tier where a decision is made as to what is the most appropriate additional testing to conduct that might provide sufficient information for hazard classification decisions. Generally the stepwise testing proceeds from existing information and data to *in vitro* tests, followed by limited *in vivo* testing, and then to a full traditional *in vivo* test as the final tier, if necessary.

Normally, validation of testing strategies can be made using existing data, provided that there is sufficient data on the same substances for all of the test methods proposed for the test strategy. In designing prospective studies for testing strategies, it is important to ensure that all test substances are tested in all of the proposed test components proposed for the testing strategy. Each test method is assessed individually to determine which results can be useful for a hazard classification decision either alone or in combination with the various potential outcomes of each of the other test methods in the strategy. This involves determining the sensitivity and specificity for each of these possible combinations of test outcomes and assessing which ones can provide equivalent or improved predictions compared to the current existing test method.

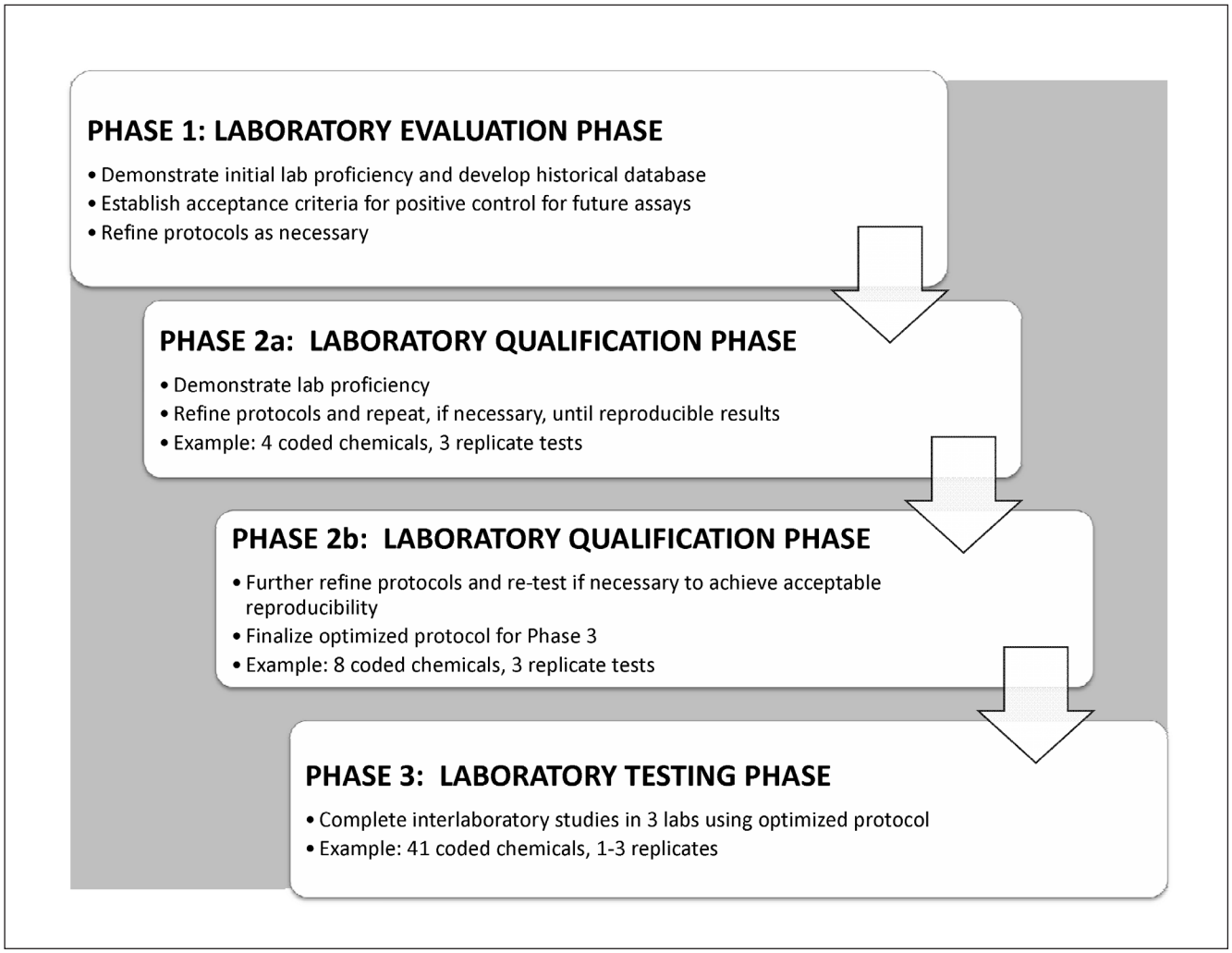


Fig. 2: Validation study approach using sequential progressive phases

9 Integrated decision strategies

With some test methods, initially proposed single decision criteria may not provide sufficient certainty with regard to the predicted outcome for some specific results, while the remaining results may have sufficient certainty in terms of sensitivity and/or specificity. For example, a test method may have a false negative rate for a certain range of responses that is not considered adequately protective compared to the reference test method. Conversely, a test method may have a false positive rate for a certain range of responses that is sufficiently high so as to not be considered acceptable. In these situations, multiple decision criteria may be necessary, where each individual decision criteria provide sufficient certainty for responses within a specified range of test results. There may also be one or more decision criteria that identify a range of responses that are associated with an unacceptable level of uncertainty, and therefore should not be used for hazard or safety decisions. In this later

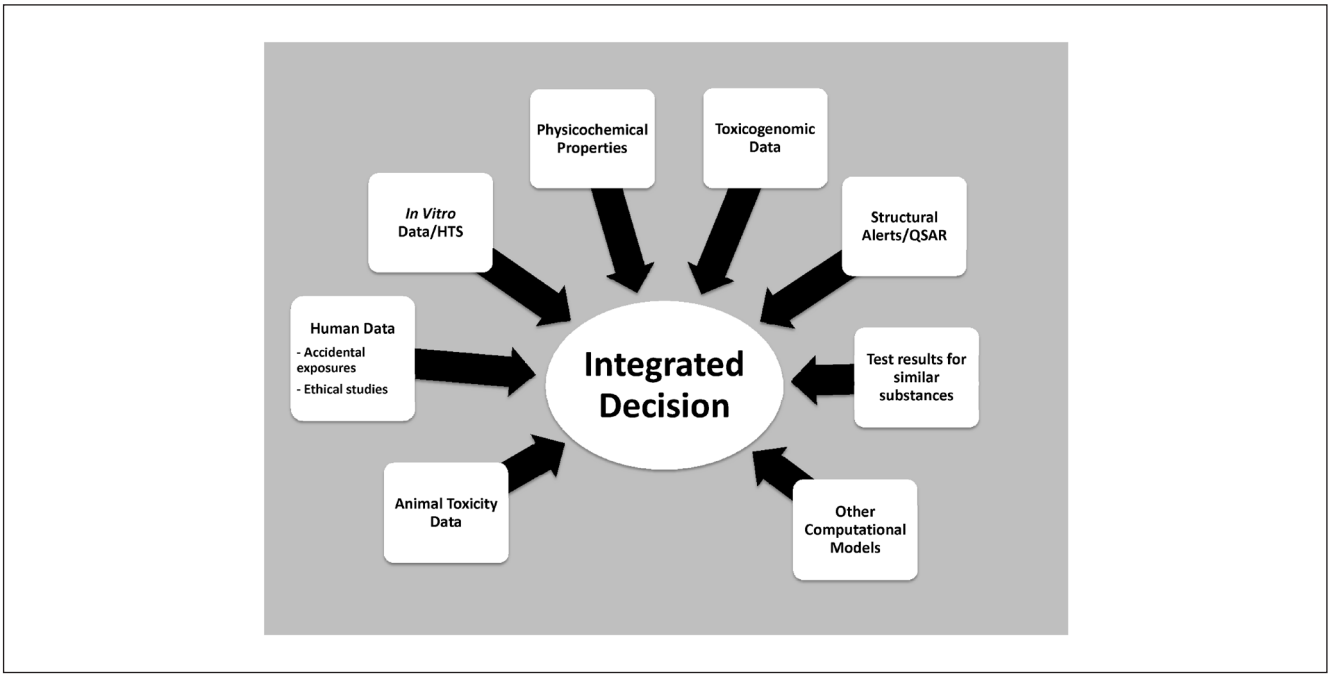


Fig. 3: Potential sources of data and information for integrated decision strategies

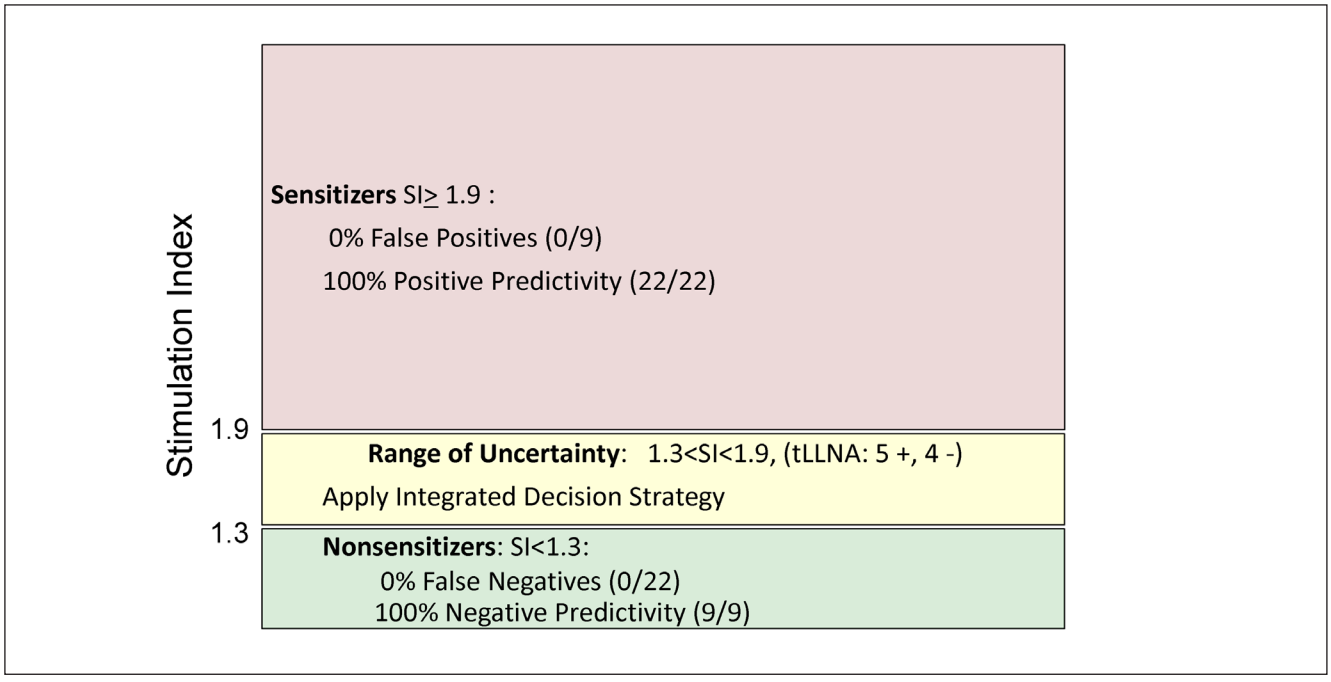


Fig. 4: Integrated decision strategy – LLNA: BrdU-ELISA

situation, additional information or data could be used to reduce the uncertainty associated with these results using an integrated decision strategy to reach a hazard or safety decision. Integrated decision strategies using multiple sources of data and information can increase the certainty of hazard and safety predictions beyond the certainty associated with only a single source of data or information (Fig. 3).

Two examples of the need for integrated decision strategies are provided by the results of a recent ICCVAM test method peer review evaluation of two non-radioactive versions of the LLNA, the LLNA: DA and LLNA: BrdU-ELISA (ICCVAM, 2009a, 2009b, 2009c). For both test methods, a single decision criteria for whether a substance was a sensitizer or non-sensitizer could not be identified that would provide the same sensitivity and specificity as the traditional LLNA for the chemicals evaluated. However, in the LLNA: BrdU-ELISA, a decision criterion using a stimulation index (SI) ≥ 1.9 to classify substances as sensitizers was found to produce a false positive rate compared to the traditional LLNA of 0% [0/9] and a positive predictivity of 100% (22/22), which was obviously considered acceptable (Fig. 4)[1]. A second decision criterion of SI ≤ 1.3 to classify substances as non-sensitizers was found to produce a false negative rate compared to the traditional LLNA of 0% [0/22] and a negative predictivity of 100% (9/9) which also was considered acceptable (Fig. 4).

However, there were five sensitizers and four non-sensitizers among the reference test substances that produced an SI greater than 1.3 and less than 1.9 in the LLNA: BrdU-ELISA. Accordingly, SI results of greater than 1.3 and less than 1.9 were not considered sufficiently predictive to be used for hazard or safety decisions. To reduce the uncertainty associated with SI results in this range, additional information and data were considered necessary for evaluation in an integrated decision strategy to determine if this combined data would support a hazard decision. Additional information that could be considered included dose-response information, statistical comparison of treated vs. vehicle control groups, peptide reactivity, molecular weight, results from related substances, presence or absence of structural alerts, and *in vitro* and other testing data (Fig. 5).

The sensitivity and specificity associated with each of these other types of data or information must be available. For the LLNA: BrdU-ELISA, when an SI result occurs in the range of uncertainty, negative results for peptide reactivity and negative results in one or more *in vitro* assays for ACD were found to provide sufficient additional data to support a hazard decision as a non-sensitizer without resulting false negatives. This approach allowed an overall specificity of 100% for the validation database.

NICEATM and ICCVAM are currently assessing the other types of test method data and information that might be used in an integrated decision strategy for these two test methods. The sensitivity and specificity associate with the outcomes in each of these other types of data will need to be carefully assessed and incorporated into classification decisions. Successful applica-

[1] These results were obtained using the most prevalent outcome for substances that were tested multiple times.

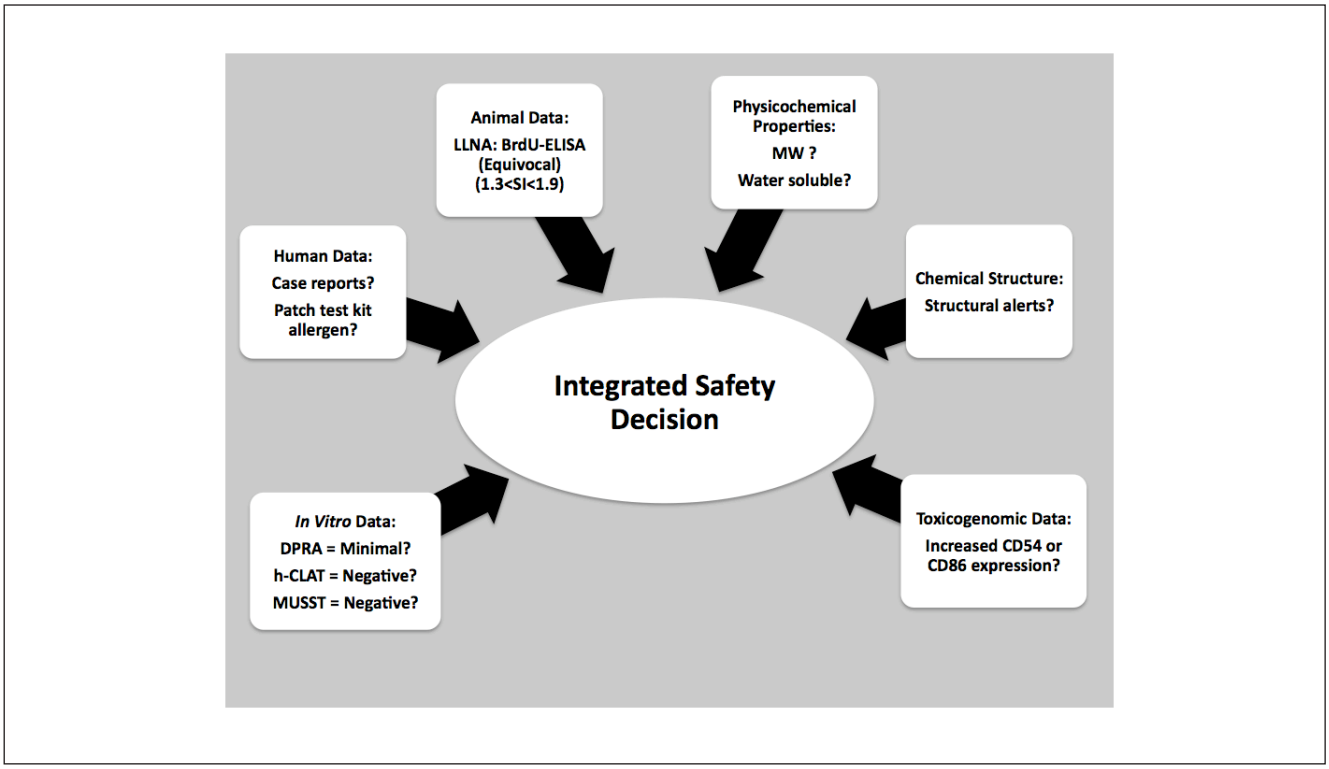


Fig. 5: Sources of potentially relevant data and information for an integrated decision strategy for uncertain results in the LLNA: BrdU-ELISA

tion of the integrated decision strategy approach is expected to produce acceptable classification decisions and avoid the need for additional testing.

10 Conclusions

Advances in science and innovative technologies are providing new opportunities to develop improved safety testing methods and strategies. Consideration of validation principles and potential application to regulatory decision-making during early stages of research, development, and validation will help expedite the scientific validation of these new methods and strategies. Validation databases will need to adequately characterize the usefulness and limitations of new proposed test methods and strategies, and support determinations of whether the new method or approach can provide equivalent or improved protection compared to existing test methods. New methods and integrated strategies should be developed and validated in consultation with relevant stakeholders and national validation centers in order to ensure adequate and appropriate studies. Comprehensive and optimal validation study designs are expected to expedite the validation and regulatory acceptance of new test methods and strategies that support improved safety assessments and contribute to reduced animal use for regulatory testing.

References

Basketter, D. A., Gerberick, G. F. et al. (2007). The local lymph node assay: current position in the regulatory classification of skin sensitizing chemicals. *Cutan. Ocul. Toxicol. 26(4),* 293-301.

Collins, F. S., Gray, G. M. and Bucher, J. (2008). Toxicology. Transforming environmental health protection. *Science 319(5865),* 906-907.

Corvi, R., Ahr, H. J. et al. (2006). Meeting report: Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ. Health Perspect. 114(3),* 420-429.

ICCVAM (2009a). Independent Scientific Peer Review Panel Report: Updated Validation Status of New Versions and Applications of the Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemical and Products. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf (Accessed: September 17, 2009).

ICCVAM (2009b). Revised Draft Background Review Document: Non-Radioactive Murine Local Lymph Node Assay: Modified by Daicel Chemical Industries, Ltd. Based on ATP Content (LLNA: DA). Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/immunotox/llna-DA/BRDcomplete.pdf (Accessed: September 17, 2009).

ICCVAM (2009c). Revised Draft Background Review Document: Non-radioactive Murine Local Lymph Node Assay: BrdU-ELISA Test Method Protocol. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/immunotox/llna-ELISA/BRDcomplete.pdf (Accessed: September 17, 2009).

ICCVAM (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods. NIH Publication Number 03-4508. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.pdf

ICCVAM (1999). The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4494. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf (Accessed: September 17, 2009).

ICCVAM (1997). Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication Number 97-3981. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf (Accessed: September 17, 2009).

Michael, S., Auld, D. et al. (2008). A Robotic Platform for Quantitative High-Throughput Screening. *ASSAY and Drug Development Technologies 6(5),* 637-638.

NAS (2009). Emerging Science for Environmental Health Decisions. Available at: http://dels.nas.edu/envirohealth/index.shtml (Accessed September 17, 2009).

NICEATM (2008). Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Available at: http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm (Accessed September 17, 2009).

NRC (2009). *Science and Decisions: Advancing Risk Assessment. Washington*, DC: National Academy Press.

NRC (2007a). Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: National Academies Press.

NRC (2007b). *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*. Washington, DC: National Academies Press.

NTP (2004). Toxicology in the 21st Century: The Role of the National Toxicology Program. Available at: http://ntp.niehs.nih.gov/ntp/main_pages/NTPVision.pdf (Accessed September 17, 2009).

OECD (2005). OECD Series on Testing and Assessment. Number 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: http://appli1.oecd.org/

olis/2005doc.nsf/linkto/env-jm-mono%282005%2914 (Accessed: September 17, 2009).

Paris, M., Strickland, J. et al. (2006). Reproducibility analyses for in vitro neutral red uptake methods from a validation study to evaluate in vitro cytotoxicity assays for estimating rodent acute systemic toxicity. *The Toxicologist 90, Abstract 1970*. Available at: http://iccvam.niehs.nih.gov/meetings/SOT06/sot06paris.pdf (Accessed September 17, 2009).

Schmidt, C. W. (2009). TOX 21: new dimensions of toxicity testing. *Environ Health Perspect 117(8),* A348-353. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19672388

Stokes, W., Bremer, S. et al. (2007). NICEATM/ECVAM/JaCVAM Multi-phased International Validation Study of a Stably-Transfected Estrogen Receptor (ER) Transcriptional Activation (TA) Test Method. *The Toxicologist*. Available at: http://iccvam.niehs.nih.gov/meetings/SOT07/sot07stokes_pos.pdf

Stokes, W. S. (2007). Alternative and integrated approaches to skin sensitization: recent progress and the way forward (Abstract, 46th Annual Meeting of the Society of Toxicology). *Toxicologist 96*. Available at: Available: http://iccvam.niehs.nih.gov/meetings/SOT07/sot07stokes_pre.pdf. (Accessed September 17, 2009).

Stokes, W. S. and Schechtman, L. M. (2007). Validation and regulatory acceptance of new, revised, and alternative toxicological methods. *Principles and Methods of Toxicology* (1103-1128). A. W. Hayes. Philadelphia, Pennsylvania: Taylor and Francis.

Stokes, W. S. (2006). Validation of In Vitro Methods for Toxicology Studies. *Biological Concepts and Techniques in Toxicology: An Integrated Approach* (161-176). J. E. Rieviere. New York, NY: Taylor & Francis Group.

USC (2000). ICCVAM Authorization Act of 2000. Public Law 106-545. Available at: http://frwebgate.access.gpo.gov/cgi-bin/useftp.cgi?IPaddress=162.140.64.183&filename=publ545.pdf&directory=/diska/wais/data/106_cong_public_laws (Accessed: September 17, 2009).

Acknowledgements

The authors acknowledge the expert assistance of David Allen and Steven Morefield from ILS, Inc., the NICEATM Support Contractor, and Debbie McCarley, NICEATM/NIEHS, in the preparation of this manuscript. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

William S. Stokes, D.V.M., DACLAM
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park
NC 27709, USA
e-mail: stokes@niehs.nih.gov

Developing Performance Standards to Expedite Validation of Innovative and Improved Test Methods

Marilyn Wind[1] and William S. Stokes[2]

[1]US Consumer Product Safety Commission, Bethesda, MD, USA; [2]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA

Summary

Regulatory acceptance of scientifically valid new test methods is often followed by improved versions that incorporate innovations and enhancements to provide better performance or other advantages. The use of performance standards evolved to support the use of proprietary test methods for regulatory testing and to expedite validation of new and revised test methods that are structurally and functionally similar to accepted test methods. As new innovative technologies become available that can be used to improve existing test methods or to develop similar test methods, the availability of performance standards will facilitate more rapid evaluation of these methods. Performance standards are based on the validated reference test method and consist of essential test method components, a minimum list of reference substances, and standards for accuracy/reliability. The routine development and availability of scientifically sound performance standards is expected to expedite the efficient validation of innovative and improved test methods and testing strategies that provide for improved hazard assessments and the reduction, refinement, and replacement of animal use.

Keywords: performance standards, validation, innovative methods

1 Introduction

Performance standards communicate the basis by which new proprietary (i.e., copyrighted, trademarked, registered) and nonproprietary test methods are determined to be scientifically valid (i.e., have sufficient accuracy and reliability) for specific testing purposes. These performance standards, following acceptance by regulatory agencies, can then be used to evaluate the acceptability in terms of reliability and accuracy of other test methods that are based on similar scientific principles and measure or predict the same biological or toxic effect (Stokes et al., 2006). The development and availability of performance standards allow regulatory agencies to endorse proprietary test methods and include them in test guidelines. Performance standards also provide the basis for evaluating the acceptability of proposed test methods that are mechanistically and functionally similar to an adequately validated and accepted reference test method (Stokes and Schechtman, 2007). This paper describes the concept of performance standards and examples of recently developed performance standards for regulatory safety testing methods.

2 Elements of performance standards

Performance standards consist of three critical elements:

- *Essential test method components*
  These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a proposed test method is based on the same concepts as the corresponding validated test method.
- *Minimum list of reference substances*
  These are used to assess the accuracy and reliability of a proposed, mechanistically and functionally similar test method. These substances are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method, but they should not be used to develop the decision criteria for the proposed test method. They are the *minimum* number that should be used to evaluate the performance of a proposed, mechanistically and functionally similar test method.
- *Accuracy and reliability values*
  These values provide the standards that a proposed alternative test method should meet or exceed when evaluated using the minimum list of reference substances.

3 Process for developing performance standards

ICCVAM uses a detailed, transparent process for developing performance standards for new test methods that emphasizes independent peer review and the opportunity for stakeholder involvement (ICCVAM, 2003). First, NICEATM and the

appropriate ICCVAM working group develop proposed performance standards based on available validation study data. This initial step includes input from working group liaisons designated by ECVAM, JaCVAM, and Health Canada. Alternatively, if a proposed test method sponsor proposes performance standards, these will be considered by ICCVAM at this stage.

The next step involves an international independent peer review of the proposed performance standards. A peer review panel comprised of experts from across the world evaluates the performance standards for completeness and appropriateness during its evaluation of the validation status of the proposed test method. At this stage, the proposed performance standards are also made available with the test method submission to the public for comment prior to and during the peer review panel meeting. This public review process also allows for the performance standards to be considered by other national and international advisory committees such as the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) and the ECVAM Scientific Advisory Committee (ESAC).

The appropriate ICCVAM working group, with the assistance of NICEATM, prepares the final performance standards for ICCVAM approval, taking into consideration all of the recommendations of the peer review panel, advisory committees, and public comments. Performance standards recommended by ICCVAM are provided to Federal agencies and made available to the public so that they then can be referenced as adopted by regulatory authorities in guidelines issued for new test methods.

4 Critical step: developing the reference substance list

The process for developing a list of performance standards references substances should follow a step-wise process and consider several selection criteria during this process. An initial list of candidate reference substances can be generated simply by identifying those within the validation database that are 1) commercially available and 2) have high quality reference data demonstrating consistent results in the validated reference test method and in the *in vivo* or other reference test method. Substances that do not meet these criteria should not normally be considered as a reference substance. Once a qualifying list has been developed based on these essential criteria, additional selection criteria can be applied to prioritize the reference list. To the extent possible, the final list of performance standards substances should:

- Represent the full range of responses that the validated test method is capable of measuring or predicting, from strong to moderate to weak effects, as well as negative effects
- Represent the relevant range of chemistry and chemical classes
- Representative the relevant range of physical properties (e.g., solids/liquids, molecular weight, solubility)
- Have data or experience (e.g. accidental exposures) available from the species of interest; i.e., human data if proposed for predicting human effects
- Reflect the accuracy of the validated test method
- Have well-defined chemical structures
- Not be associated with excessive occupational or environmental hazard or prohibitive disposal costs
- Be readily available from commercial sources
- Have high quality data available from the validated reference test method and the *in vivo* or other reference test method

5 Development of performance standards for *in vitro* corrosivity test methods

Following the concepts detailed above, ICCVAM developed performance standards for three proprietary dermal corrosivity test methods previously recommended by ICCVAM: Corrositex®, EPISKIN™, and EpiDerm™ (ICCVAM, 1999, 2002). ICCVAM also developed performance standards for the one non-proprietary test method (the rat skin transcutaneous electrical resistance [TER] assay) that was also recommended by ICCVAM (ICCVAM, 2002). Due to the structural and functional differences of these test methods, three different sets of performance standards were developed (ICCVAM, 2004).

One set of performance standards was based on the reconstructed human skin model systems (i.e., EPISKIN™, and EpiDerm™). Because the validation database was larger for EPISKIN™ at the time, the standards were based on that method. In addition to the essential test method components, a minimum list of 24 substances was selected from the 60 substances used for validation of EPISKIN™; this included 12 corrosives and 12 noncorrosives. The decision criteria for these assays (as well as the rat skin TER) do not allow detection of all United Nations corrosivity packing groups and instead distinguish between Category I, a combined Category II/III, and not corrosive. As a result, fewer substances were required for a balanced design and the final minimum substances included 12 corrosive substances and 12 noncorrosive substances. The 24 substances were selected based on commercial availability, representation of the full severity range of dermal corrosivity, and representation of relevant chemical classes.

Another set of performance standards were developed for the rat skin TER, in which a minimum list of 24 substances was selected from those in the total validation database of 60 chemicals. Again, commercial availability, representation of the full severity range of dermal corrosivity, and representation of relevant chemical classes were used as criteria for selection. NICEATM and ICCVAM recently submitted proposals to OECD to update the TGs for these test methods (TG 430 [rat skin TER] and TG 431 [human skin model systems]) with performance standards (OECD, 2009a, 2009b).

The third set of performance standards were developed for membrane barrier systems like Corrositex®. The candidate list used to select the proposed minimum reference substances for the membrane barrier was initially generated from the original validation database of 163 substances. This was reduced

to 40 substances after considering the commercial availability of substances, representation of severity range of dermal corrosivity (UN Packing Groups 1, 2, and 3); and representation from relevant chemical classes. In order to allow detection of all 4 United Nations corrosivity packing group categories (I, II, III, and not corrosive), the final list needed sufficient numbers of a balanced design. This resulted in the final selection of 12 Noncorrosive substances, 9 Packing Group I substances, 9 Packing Group II substances, and 10 Packing Group III substances. OECD Test Guideline (TG) 435 (Membrane Barrier Systems) is the first OECD TG to include test method performance standards (OECD, 2006).

6 Putting concepts into practice: performance standards for the murine Local Lymph Node Assay (LLNA)

Internationally harmonized performance standards for the LLNA were recently developed through collaboration between NICEATM-ICCVAM, ECVAM, and JaCVAM. These performance standards will allow rapid assessment of the validity of modified versions of the traditional LLNA, such as those using non-radioisotopic methods.

The candidate list used to select the proposed minimum reference substances for the LLNA performance standards was initially generated from the database submitted to ICCVAM for the 1998 evaluation of the LLNA. This database of 209 substances was reduced to 127 candidate substances by identifying those substances for which comparative guinea pig maximization (GPMT) or Buehler test (BT) data were available. The availability of such data is important because any accuracy comparisons of new or revised methods must include the currently accepted regulatory test methods (i.e., in this case, the LLNA, and the GPMT and/or BT), as well as comparison to available human data and/or experience. Limiting the list to substances with GPMT and BT data that were collected using a standard protocol (e.g. EPA, 1998) and those with unequivocal LLNA results reduced the set from 127 to 97. Substances must also be readily available from commercial sources. Further limiting the list of substances to those that are commercially available reduced the list from 97 to 81 candidate substances.

The candidate list was then reduced to a candidate list of 40 substances taking into consideration, where feasible, the following criteria:

– Maintaining similar accuracy statistics to those achieved in the original LLNA validation report
– Availability of human testing data or experience

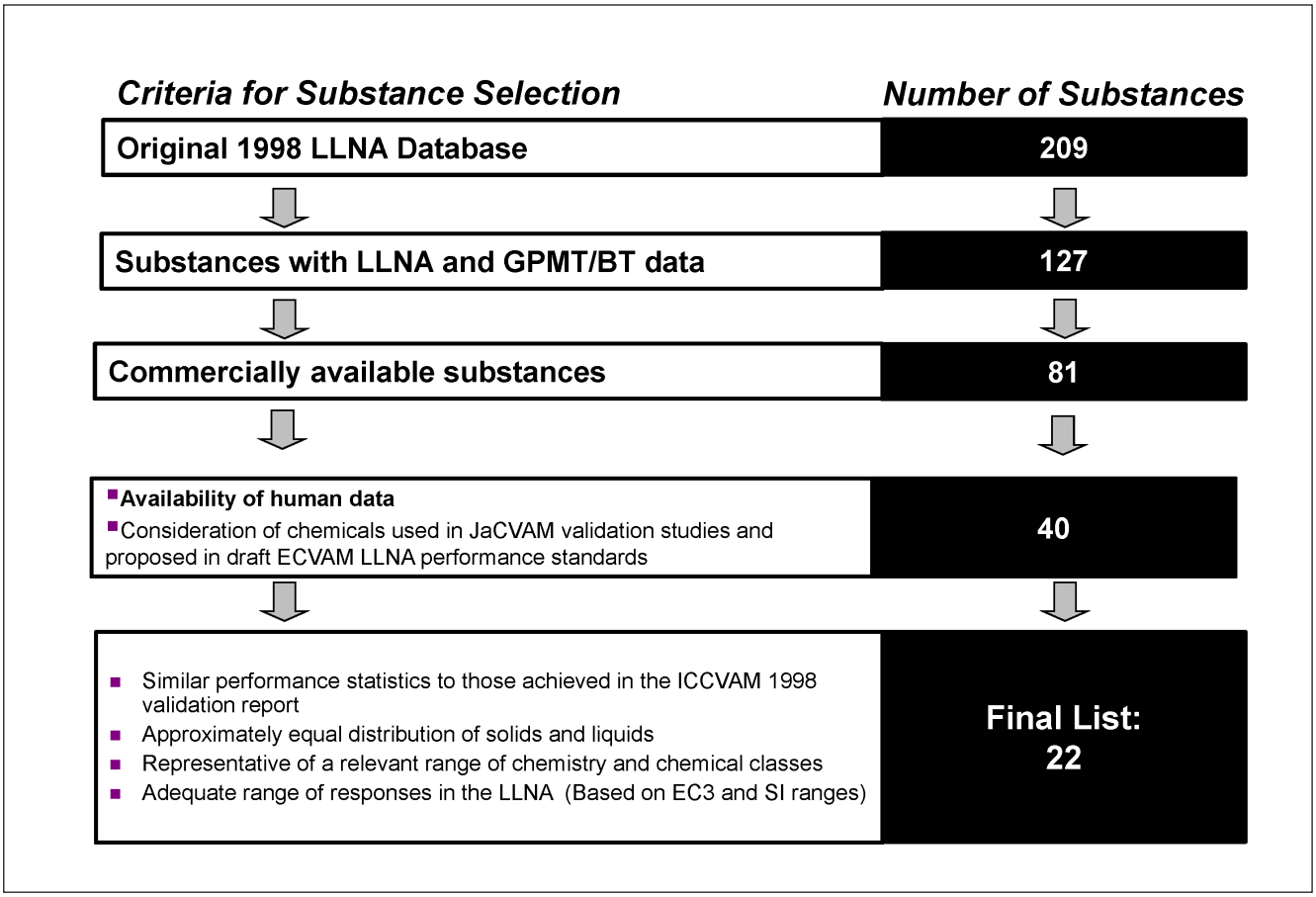


Fig. 1: Process for developing a minimum list of LLNA performance standards substances.

Tab. 1: Timeline for the Development of ICCVAM Performance Standards for the LLNA

Date	Event
Jan 10, 2007	ICCVAM nomination from the CPSC
May 17, 2007	ICCVAM public request for comments, Panel nominations, and data (NIEHS, 2007a)
Sep 12, 2007	Draft ICCVAM LLNA Performance Standards (PS) released for public comment (NIEHS, 2007b)
Sep 25-27, 2007	ECVAM Workshop on an evaluation of performance standards and alternative endpoints for the LLNA (Basketter et al., 2008).
Jan 8, 2008	Revised draft PS published for public comment (NIEHS, 2008)
Mar 4-6, 2008	ICCVAM Independent Peer Review Panel Meeting, CPSC Headquarters, Bethesda, MD; public meeting with opportunity for oral public comments.
Sep 23-24, 2008	ECVAM/JaCVAM/ICCVAM meeting on internationally harmonized LLNA PS
Oct-Nov 2008	ICCVAM (Oct 29) and ESAC (Nov 5) endorse harmonized LLNA PS
Jun 2009	Circulation of revised OECD TG 429 with LLNA PS

– Maintaining approximately the same proportion of solids and liquids
– Representing a relevant range of chemistry and chemical classes
– Providing an adequate range of responses in the LLNA, from strong to weak to negative
– Consideration of substances used in the Japanese Center for the Validation of Alternative Methods (JaCVAM) validation studies and in draft performance standards proposed by ECVAM

A final list of 22 proposed reference substances was then selected from the list of 40 candidate substances based on the selection criteria (ICCVAM, 2009). Figure 1 provides a breakdown of the impact of specific criteria on the list of candidate substances.

7 Development of internationally harmonized LLNA performance standards

ICCVAM released draft performance standards to the public for comment on September 12, 2007 (NIEHS, 2007b). Concurrently, two other international validation organizations, ECVAM and JaCVAM, were also independently developing LLNA performance standards. ECVAM was independently drafting LLNA performance standards that could be used to evaluate a non-radioactive LLNA test method submitted from a European developer (Ehling et al., 2005), and JaCVAM was drafting performance standards that could be applied to two non-radioactive LLNA methods for which validation studies were underway (Takeyoshi et al., 2001; Idehara et al., 2008).

Harmonized performance standards were viewed as critical for the success of efforts to reduce, refine, and replace the use of animals in regulatory testing for allergic contact dermatitis testing. Therefore, NICEATM and ICCVAM invited ECVAM and JaCVAM to designate liaisons to the ICCVAM Immunotoxicity Working Group (IWG) in order to work closely together to develop internationally harmonized performance standards. Input was also obtained from the ECVAM Task Force on Skin Sensitization (Basketter et al., 2008).

After consideration of these comments, a revised version was made available to an ICCVAM Independent Expert Peer Review Panel (Panel) for consideration at a public meeting in March 2008 (ICCVAM, 2009). The revised draft performance standards were also made available to the public for comment in advance of the Panel meeting, and all comments received were provided to the Panel for their consideration. The Panel's conclusions and recommendations were made available to the public and to SACATM and ESAC for comment. The Panel Report and all comments by the public, ESAC, and SACATM were considered by ICCVAM in preparing final LLNA performance standards recommendations for public release and submittal to U.S. Federal agencies (ICCVAM, 2009). Table 1 provides a summary of the timeline associated with the development of harmonized LLNA performance standards, which recently culminated with a proposal to update OECD TG 429 (OECD, 2002) with these performance standards.

8 Conclusions

Performance standards allow for expedited validation of innovative and improved versions of validated and accepted reference test methods that may provide advantages such as greater accuracy and efficiency. An appropriate set of high quality reference substances that can be readily obtained is critical for adequate performance standards. These reference substances should be selected from a robust database of commercially available candidate substances. Therefore, it is essential that stakeholders facilitate the collection and public availability of high quality reference substance data to support robust performance standards. Performance standards are now routinely

developed and incorporated in national and international guidelines and are consistent with international guidance on validation and regulatory acceptance as outlined in OECD Guidance Document 34 (OECD, 2005).

References

Basketter, D., Cockshott, A. et al. (2008). An evaluation of performance standards and non-radioactive endpoints for the local lymph node assay. *ATLA 36(2),* 243-257. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-45549100795&partnerID=40&rel=R8.0.0

Ehling, G., Hecht, M. et al. (2005). An European inter-laboratory validation of alternative endpoints of the murine local lymph node assay 2nd round. *Toxicology 212*, 69-79.

EPA, U. S. (1998). Health Effects Test Guidelines: OPPTS 870.1200 – Acute Dermal Toxicity. EPA 712-C-98-192. Washington, DC, U.S. Environmental Protection Agency. Available at: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-1200.pdf

ICCVAM (2009). Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf (Accessed: September 17, 2009).

ICCVAM (2004). Recommended Performance Standards for In Vitro Test Methods for Skin Corrosion: Prepared by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 04-4510. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf (Accessed: September 17, 2009).

ICCVAM (2003). ICCVAM Guidelines for the Nomination and Submission of New, Revised and Alternative Test Methods NIH Publication Number 03-4508. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/SuppDocs/SubGuidelines/SD_subg034508.pdf

ICCVAM (2002). Evaluation of EPISKIN™ , EpiDerm™ , and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: In Vitro Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication Number 02-4502. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/cwgfinal02/cwgfinal.pdf (Accessed: September 17, 2009).

ICCVAM (1999). Corrositex®: An In Vitro Test Method for Assessing Dermal Corrosivity Potential of Chemcials - An Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4495. Research Trinagle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf (Accessed: September 17, 2009).

Idehara, K., Yamagishi, G. et al. (2008). Characterization and evaluation of a modified local lymph node assay using ATP content as a non-radio isotopic endpoint. Journal of Pharmacological and Toxicological Methods 58(1), 1-10. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-42649119427&partnerID=40&rel=R8.0.0 (Accessed September 17, 2009).

NIEHS (2008). Announcement of an Independent Scientific Peer Review Panel Meeting on the Murine Local Lymph Node Assay; Availability of Draft Background Review Documents; Request for Comments. Federal Register 73, 1360-1361. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_25553.pdf (Accessed September 17, 2009).

NIEHS (2007a). The Murine Local Lymph Node Assay: Request for Comments, Nominations of Scientific Experts, and Submission of Data. Federal Register 72, 27815-27817. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_9544.pdf. (Accessed September 17, 2009).

NIEHS (2007b). Draft Performance Standards for the Murine Local Lymph Node Assay: Request for Comments. Federal Register 72, 52130-52131. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_E7_18011.pdf (Accessed September 17, 2009).

OECD (2009a). Draft Proposal for an Update to Test Guideline 430: In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). Available at: http://www.oecd.org/dataoecd/14/2/43295353.pdf (Accessed September 17, 2009).

OECD (2009b). Draft Proposal for an Update to Test Guideline 431: In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method. Available at: http://www.oecd.org/dataoecd/63/6/43302385.pdf (Accessed September 17, 2009).

OECD (2006). OECD Guidelines for the Testing of Chemicals. Test No. 435: In Vitro Membrane Barrier Test Method for Skin Corrosion. Adopted July 19, 2006. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s33/p1 (Accessed: September 17, 2009).

OECD (2005). OECD Series on Testing and Assessment. Number 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: http://appli1.oecd.org/olis/2005doc.nsf/linkto/env-jm-mono%282005%2914 (Accessed: September 17, 2009).

OECD (2002). OECD Guidelines for the Testing of Chemicals. Test No. 429: Skin Sensitisation: Local Lymph Node Assay. Adopted April 24, 2002. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s29/p1 (Accessed: September 17, 2009).

Stokes, W. S. and Schechtman, L. M. (2007). *Validation and regulatory acceptance of new, revised, and alternative toxicological methods*. *Principles and Methods of Toxicology*. A. W. Hayes. Philadelphia, Pennsylvania: Taylor and Francis: 1103-1128.

Stokes, W. S., Schechtman, L. M. et al. (2006). The use of test method performance standards to streamline the validation process. *ALTEX 23, Suppl*., 342-345. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17492210

Takeyoshi, M., Yamasaki, K. et al. (2001). Development of non-radio isotopic endpoint of murine local lymph node assay based on 5-bromo-2?-deoxyuridine (BrdU) incorporation. *Toxicology Letters 119(3),* 203-208. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035826075&partnerID=40&rel=R6.0.0 (Accessed September 17, 2009).

Acknowledgements

The authors acknowledge the expert assistance of David Allen and Steven Morefield from ILS, Inc., the NICEATM Support Contractor, and Debbie McCarley, NICEATM/NIEHS, in the preparation of this manuscript. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

William S. Stokes, D.V.M.
Director, NICEATM
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park
NC 27709, USA
e-mail: stokes@niehs.nih.gov

Breakout Sessions

EB1: Status report on Predict-IV

An Integrated Modelling Approach for *In Vitro* to *In Vivo* Extrapolations

Frédéric Y. Bois[1,2], Dany Habka[2] and Céline Brochot[2]

[1]UTC – Technological University of Compiegne, Chair of Mathematical Modelling for Systems Toxicology, Royallieu Research Centre, Compiegne Cedex, France; [2]INERIS, Parc Technologique ALATA, Verneuil en Halatte, France

Summary

Predicting in vivo drug toxicity from in vitro testing requires modelling those processes that are not reproduced in vitro. The most obvious difference between the two settings is the lack of integrated absorption, distribution, metabolism and excretion (ADME) that govern target tissue exposure in vivo. For equal input doses, the concentrations to which in vitro systems are exposed may not correspond to those found in vivo. The partners of the European project PREDICT-IV in charge of modelling develop prediction models for ADME processes and integrate them into generic physiologically based pharmacokinetic (PBPK) models. Examples of simulations of concentration-time profiles for diazepam and ciclosporin A, in various human or rat organs and tissues, are presented here. The in vivo toxicity of drugs will be forecasted by coupling the predicted target tissue concentration profiles to dose-response relationships observed in vitro. Human variability will be evaluated for each component of the approach.

Keywords: ciclosporin A, diazepam, PBPK models, PK/PD models, toxicity prediction

1 Introduction

It is now recognised that replacing animals by *in vitro* systems for assessment of medium- to long-term toxicity of drugs or general chemicals requires an integration of pharmacokinetics (Bouvier d'Yvoire et al., 2007). Predicting *in vivo* drug toxicity from *in vitro* testing requires modelling the absorption, distribution, metabolism and excretion (ADME) processes that are not reproduced *in vitro*. For equal input doses, the concentrations to which *in vitro* systems are exposed may not correspond to the target tissue exposure experienced *in vivo*. INERIS, Simcyp Limited, the German Cancer Research Centre, Emergentec Biodevelopment GmbH and ECVAM, which constitute the modelling team of the European project PREDICT-IV, will integrate prediction models for ADME processes into global generic physiologically based pharmacokinetic (PBPK) models able to simulate concentration-time profiles in human or rat blood and tissues. *In vivo* drug toxicity will be predicted by coupling the predicted target tissue concentration profiles to dose-response relationships observed *in vitro*. Human variability will be evaluated for each component of the approach.

While the establishment of predictive dose-response relationships for neurotoxicity, liver and kidney toxicity is a very important part of the project, this article focuses on the progress made so far on PBPK modelling by PREDICT-IV.

PBPK models are increasingly used in drug development and regulatory toxicology to predict the kinetics and metabolism of substances in the body (Barton et al., 2007; Bouvier d'Yvoire et al., 2007; Loizou et al., 2008). In these models, the body is represented by a set compartments corresponding to specific organs or tissues, and the transfers or transports of drugs are dictated by various physiological flows (blood, bile, pulmonary ventilation, etc.) (Bois and Paxman, 1992). A system of differential equations can be written, with parameters representing blood

flow rates, organ volumes etc., for which information is available in the published scientific literature or may be obtainable *in vitro* (Woodruff and Bois, 1993; Parrott et al., 2005). Numerical integration of that differential system computes the quantity and concentration of the drug considered in each compartment as a function of time and exposure dose. Indeed, such a description of the body is approximate, if not rough, but a balance has to be found between precision (and therefore complexity) and simplicity (ease of use).

We will briefly describe the global approach taken by PREDICT-IV to integrate experimental evidence for a predictive approach to toxicology. We will then describe Simcyp's generic PBPK model and the specific parameter setting adopted for ciclosporin A and diazepam in rats and humans. Modelling results for those two substances are presented and discussed.

2 Methods

General modelling framework

Figure 1 presents graphically the integrated prediction and validation framework of PREDICT-IV. *In vitro* studies will provide the necessary data for the development of a set of ADME QSARs and multivariate dose-response models for cellular and organ-level effects. Those models will be coupled to a generic PBPK model and will provide an *in silico* pharmacokinetic-pharmacodynamic (PK/PD) tool predictive of medium- to long-term toxicity. In order to validate that tool, its predictions for plasma and tissue drug levels will be confronted with pharmacokinetic data obtained in humans or rats. We do not expect to get enough data relating plasma to pathological effects in humans. Therefore the pharmacodynamic component will be

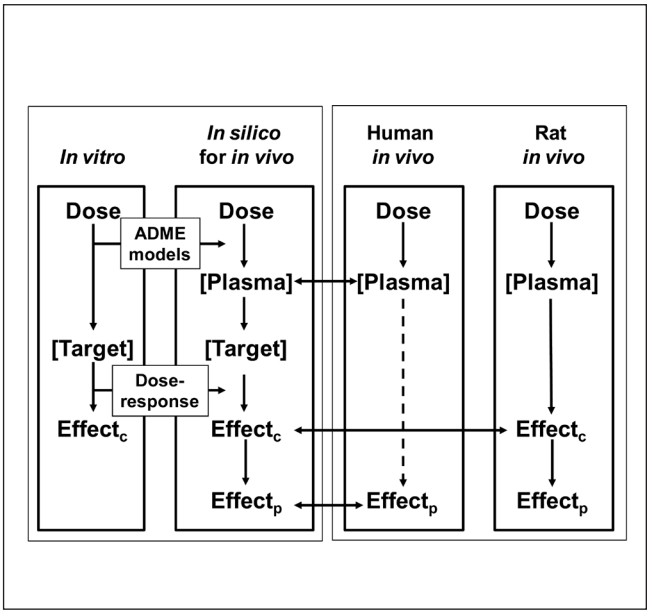


Fig. 1: Graphic representation of the integrated prediction and validation framework of PREDICT-IV.
Left panel: Predictive component of project *in vitro* studies will provide the necessary data for the development of a set of ADME QSARs and multivariate dose-response models for cellular and organ-level effects. Those models will be coupled to a generic PBPK model for humans and rats and will provide an *in silico* pharmacokinetic-pharmacodynamic (PK/PD) tool predictive of medium- to long-term toxicity. Right panel: In order to validate that tool, its predictions for plasma and tissue drug levels will be confronted to pharmacokinetic data obtained in humans or rats. To relate plasma to pathological effects, the pharmacodynamic component will be validated on the basis of published animal (rat) data.

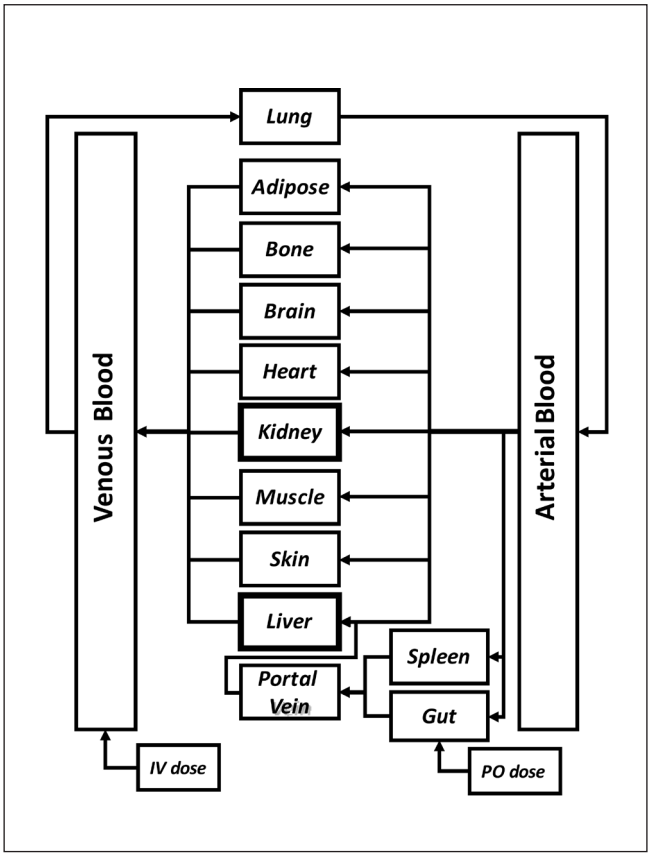


Fig. 2: Graphic representation of the physiologically based pharmacokinetic (PBPK) model of the human or rat body.
The model structure is the same for rats and humans, but parameter values and corresponding absorption, distribution, metabolism and excretion (ADME) sub-models differ between species. Dosing can be by intravenous injection (*i.v.*) or *per os (p.o.)* (the dermal and inhalation routes are being developed for humans). The liver and gut are sites of metabolism. Renal excretion takes place in the kidney. Solving numerically the corresponding set of differential equations yields the time-course of the quantity or concentration of a given drug in any of the organs or tissues included in the model.

validated on the basis of already published rat data. We do not expect perfect predictions of pharmacokinetics or toxicity data, and meta-analysis models will be developed to analyse and correct for the discrepancies.

Simcyp's generic PBPK models
The structure of the PBPK model developed by Simcyp for rats and humans is shown in Figure 2. A set of ADME sub-models, able to predict substance-specific parameter values for a large range of molecules, is already coupled to this model (Howgate et al., 2006; Jamei et al., 2009a,b,c). Obviously, while the model structure remains the same, parameter values and corresponding absorption, distribution, metabolism and excretion (ADME) sub-models differ between species. Dosing can be by intravenous injection (i.v.) or *per os (p.o.)* (the dermal and inhalation routes are also developed for humans). The liver and gut are sites of metabolism. Renal excretion takes place in the kidney. Solving numerically the corresponding set of differential equations yields the time-course of the quantity or concentration of a given drug in any of the organs or tissues included in the model.

Specific parameters for diazepam in rats
Simcyp also provides default parameter values for diazepam with its rat model. A total *in vivo* clearance of 20.4 ml/min is reported (Klotz et al., 1976; Igari et al., 1983). To model the extra-hepatic metabolism of diazepam in rats, hepatic clearance was set equal to three-quarters of total clearance (CL_{iv} = 0.75 × CL_{tot} = 15.3 ml/min; CL_r = 0.25 × CL_{tot} = 5.1 ml/min).

The intravenous injection of 1.2 mg/kg used by Igari et al. (1983) in their experiments on Wistar rats was simulated.

Specific parameters for diazepam in humans
Diazepam is not part of the default substances provided by Simcyp's human model. We used the physico-chemical characteristics specified by Simcyp's rat model (molecular weight: 284.7 g/mol; log octanol to water partition coefficient: 2.9; monoprotic base with a *pKa* of 3.4). Blood binding parameters where obtained from the literature (fraction unbound in plasma, fu_p: 0.022; blood to plasma ratio: 0.58) (Klotz et al., 1975; Klotz, 1985; Jones and Larsson, 2004). Tissue-to-plasma concentration ratios (K_p) were automatically estimated by the corresponding ADME model. For *in vivo* clearance, CL_{iv}, the published value of 1.596 L/h was used (Klotz et al., 1975).

To simulate Klotz et al. (1975) data, a two-minute intravenous infusion of 0.1 mg/kg diazepam was simulated for a representative healthy volunteer, followed up for 72 h.

Specific parameters for ciclosporin A in rats
Ciclosporin A was not one of the drugs predefined for in Simcyp's rat model. We used the physico-chemical properties provided by the Simcyp's human model (molecular weight: 1202 g/mol; log octanol to water partition coefficient: 4.3; neutral compound). Blood binding parameters where obtained from the literature (fraction unbound in plasma, fu_p: 0.062; fraction unbound in blood, fu_b: 0.05 for intravenous doses up to 6 mg/kg and 0.094 for doses above) (Kawai et al., 1998; Tanaka et al., 2000). The tissue-to-plasma concentration ratios

Tab. 1: Tissue-to-blood partition coefficients used for ciclosporin A in rats.

Tissue	Dose (mg/kg)		
	1.2	6	30
Adipose	8.36	7.77	4.94
Bone	2.81	3.76	1.08
Brain	0.01[a]	0.30	0.53
Gut	6.39	6.49	2.55
Heart	7.29	5.74	2.52
Kidneys	9.54	12.90	4.42
Liver	14.63	13.64	5.67
Lungs	8.85	6.32	5.79
Muscle	1.22	1.46	0.71
Skin	5.17	3.04	1.12
Spleen	9.15	8.64	3.61

[a]The brain concentration was below the limit of detection in Tanaka et al., (2000).

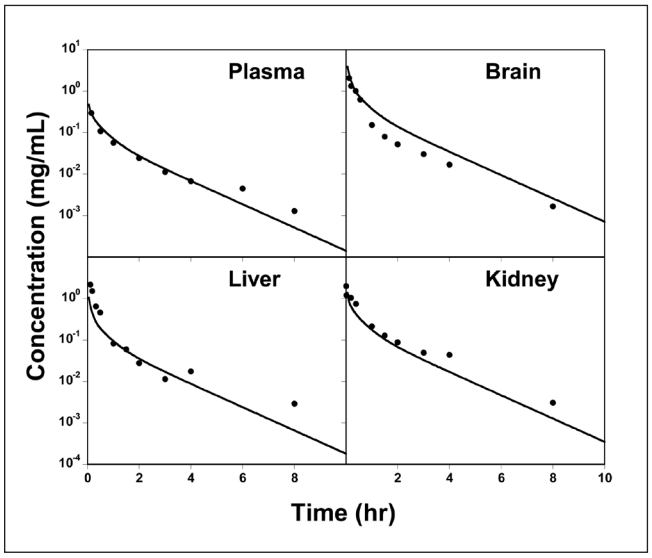


Fig. 3: Model predictions and published data on diazepam pharmacokinetics in the rat.
The rat PBPK model was run to simulate the intravenous administration of 1.25 mg/kg diazepam to Wistar rats (Igari et al., 1983). The solid lines represent the predicted time courses. The dots represent the corresponding data (average concentrations measured in 3 to 5 rats).

(or partition coefficients, K_p) (Tab. 1) were derived by multiplying published tissue-to-blood ratios (Tanaka et al., 2000) by fu_p / fu_b.

In vivo plasma clearance values were derived from Tanaka et al. (2000) by multiplying the blood clearance by fu_p / fu_b: (CL_{iv}: 0.82 ml/min for an intravenous dose of 1.2 mg/kg; 0.94 ml/min for 6 mg/kg; 0.35 ml/min for 30 mg/kg). Three experiments on Sprague-Dawley rats with 2-min intravenous infusions of 1.2 mg/kg, 6 mg/kg and 30 mg/kg ciclosporin A, respectively, were simulated (Tanaka et al., 2000).

Specific parameters for ciclosporin A in humans

Simcyp provides default parameter values for ciclosporin A pharmacokinetics in humans. We used the predefined "healthy volunteers" population. To match the characteristics of the population studied by Gupta et al. (1990), the reference body mass of the subjects was set to an average of 64 kg, with age ranging from 24 to 34 years, with 50% males. The duration of the simulated study was 24 h after an oral administration of 10 mg/kg of ciclosporin A. A first-order oral absorption model was used, with rate constant 0.91 h^{-1} (Kawai, et al., 1998).

For extrapolation to a larger population the age range was set from 20 to 70 years. To study variability, Monte Carlo simulations (Bois et al., 1990, 1991; Bois and Paxman, 1992; Rostami-Hodjegan and Tucker, 2007) of 100 subjects were performed.

3 Results

Diazepam in rats and humans

A set of simulations was performed to compare the model predictions to actual data published on rat and human diazepam pharmacokinetics. Only minimal adjustments to the parameter values, as described in the Methods section, were made. Figure 3 shows the results obtained for the rat. Only plasma, brain, liver and kidney data (Igari, et al., 1983) and model simulations are shown, because these are the compartments of primary interest in PREDICT-IV. Similar results are obtained in the other tissues.

Figure 4 presents similar results in human plasma. The experimental data of Klotz et al. (1975) for two healthy volunteers are slightly overestimated, but we have not attempted here to account for the heterogeneity of the human population.

Ciclosporin A in rats and humans

Figure 5 shows the data obtained by Tanaka et al. (2000) in Sprague-Dawley rats after intravenous administrations of 1.2, 6 or 30 mg/kg of ciclosporin. The model-predicted time courses are shown in the same Figure. Here again, no specific adjustment was made to "fit" the data beyond the obvious adaptations mentioned in the Methods section.

For humans, Figure 6 shows the data of Gupta et al. (1990), obtained after oral administration of 10 mg/kg of ciclosporin A to a healthy volunteer, in low fat or high fat diet conditions.

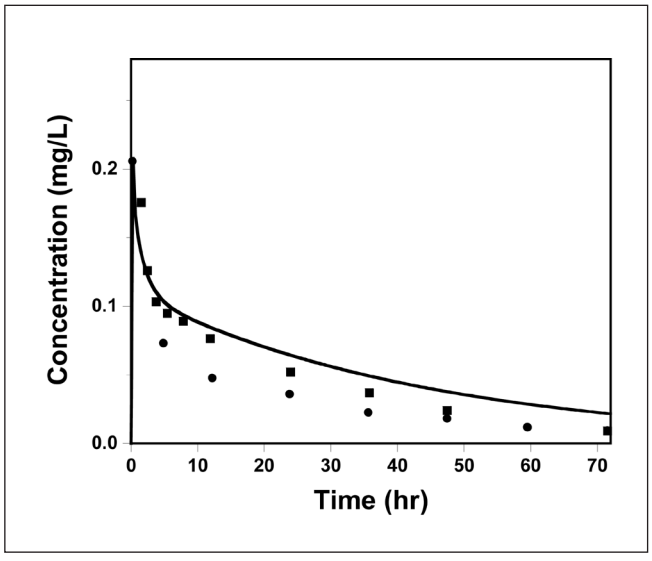


Fig. 4: Model predictions and published data on diazepam time course in human plasma.
The human PBPK model (solid line) was run to simulate the intravenous administration of 0.1 mg/kg diazepam to human healthy volunteers. The data correspond to those obtained in two subjects (circles and squares, respectively) by (Klotz et al., 1975).

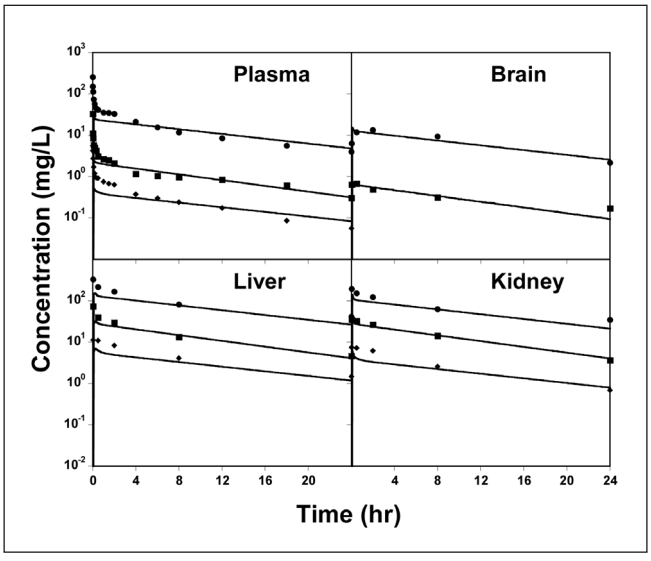


Fig. 5: Model predictions and published data on ciclosporin A pharmacokinetics in the rat.
The rat PBPK model was run to simulate the intravenous administration of 1.2, 6 or 30 mg/kg ciclosporin to Sprague-Dawley rats (Tanaka et al., 2000). The solid lines represent the predicted time courses. The dots (circles: 30 mg/kg dose; squares: 6 mg/kg dose; diamonds: 1.2 mg/kg dose) represent the corresponding data (average concentrations measured in 3 rats).

Monte Carlo simulation of 100 similar subjects (diet not considered) was performed, and the average, 5th and 95th percentiles of the simulated values are plotted.

Figure 7 illustrates the very purpose of all this modelling exercise: predicting the concentration time-course of a substance (here ciclosporin A) in the brain, liver and kidney for a varied human population. We simulated 100 healthy subjects, aged from 20 to 70 years, taking a daily dose of 10 mg/kg of ciclosporin A for 60 days. Note that during the first week of treatment, liver exposure is the highest of the three, but brain exposure after 60 days is higher than liver or kidney exposure.

4 Discussion

Predicting organ toxicity beyond acute effects is a significant scientific challenge. The team of PREDICT-IV is taking an inclusive approach to the problem, coupling *in vitro* testing, *in silico* ADME, PBPK and dose-response models in an integrated predictive tool. An important aspect of the work will be the validation, or benchmarking, of the model predictions with a set of about 30 reference drugs, known for their neurotoxicity, hepatotoxicity or kidney toxicity. To that effect, Simcyp Ltd has already extended its human pharmacokinetic model and developed a rat model. The results presented here are indeed limited to only two compounds, diazepam and ciclosporin, but give an idea of the quality of the predictions obtained with the current model. We will test the approach with a whole set of reference drugs, and we even expect improvements when additional, more precise, ADME properties prediction models are included.

We have performed Monte Carlo simulation here to describe inter-individual variability only to a limited extent, and only for ciclosporin A. This type of simulation will be performed for more diverse populations, potentially more susceptible to organ toxicity. Accessing this type of results is difficult for *in vitro* testing alone, hence the synergy with mathematical modelling, which is one of the goals of PREDICT-IV.

The results presented here go only as far as the prediction of organ exposures after medium-term repeated administration of a drug. We could have extended our simulations to longer times or different doses, but stable values were obtained for ciclosporin A after 60 days of dosing, and we would have gained no more insight with additional follow-up. Obviously this may not be true for other compounds and will have to be assessed on a case by case basis. In any case, long-term target organ exposure is the relevant measure of dose for the prediction of organ toxicity after sustained exposure. We have still to couple our PBPK model to dose-response models for a complete PK/PD approach. The dose-response component will also add a time dimension of its own, but the work on correlated multi-dimensional dose-response models in PREDICT-IV has just begun.

We are focusing on local organ toxicity, for which the target tissue concentration can be considered as the relevant measure of dose. Systemic toxicity would be much more of a challenge,

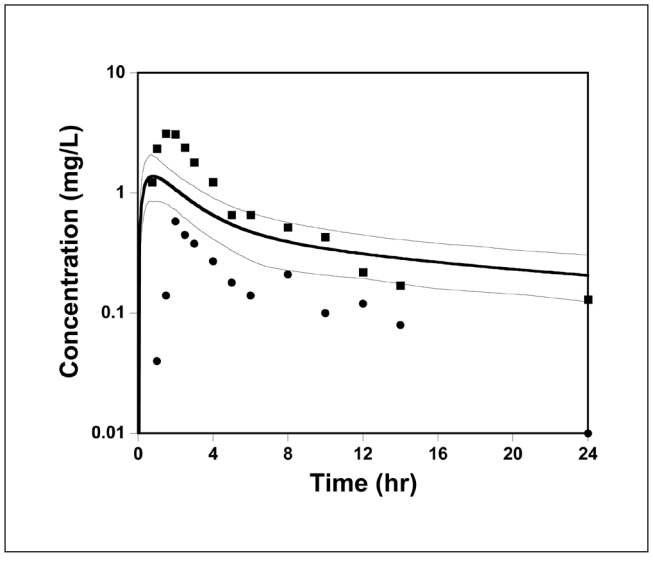


Fig. 6: Modelled and observed ciclosporin A pharmacokinetics in human blood.
The data of Gupta et al. (1990), obtained after oral administration of 10 mg/kg of ciclosporin A to a healthy volunteer, are represented by either circles (low fat diet conditions) or squares (high fat diet). The thick solid line represents the corresponding average model predictions for 100 similar subjects (diet not considered). The thin lines indicate the 5th and 95th percentiles of the 100 simulated values.

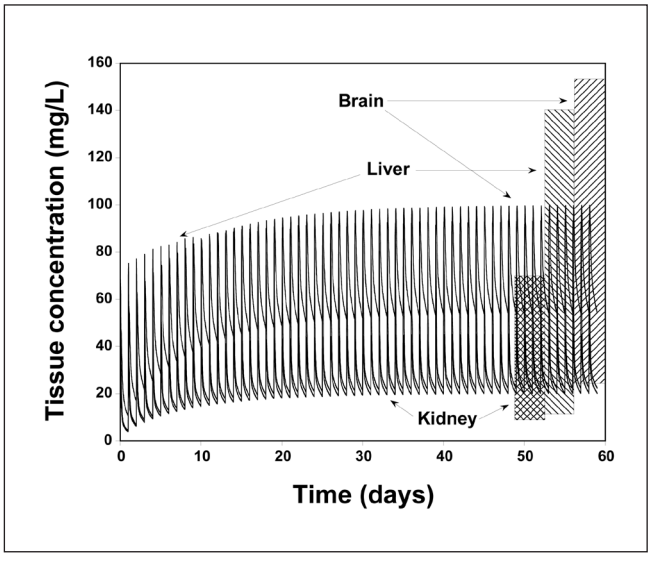


Fig. 7: Predictions of long-term ciclosporin A pharmacokinetics in humans.
Time course of ciclosporin A concentrations predicted in the brain, liver and kidney of healthy subjects, aged from 20 to 70 years, and taking a daily dose of 10 mg/kg of ciclosporin A. The lines represent the average concentrations at any time for 100 simulated subjects. The bars at the right indicate the range (from 5th to 95th percentiles) of concentrations found in the organs after 60 days of treatment. These are relevant values for the prediction of organ toxicity after sustained exposure.

combining effects of the dysfunction of several organs. However, we are only a step away from being able to describe the toxic effects of co-administered substances, since our models are able to account for at least pharmacokinetic and metabolic interactions (Rostami-Hodjegan and Tucker, 2004; Bois, submitted). We are therefore confident in our ability to make significant contributions to the 3Rs agenda and to predictive clinical and environmental toxicology in the near future.

References

Barton, H. A., Chiu, W. A., Setzer, W. et. al. (2007). Characterizing uncertainty and variability in physiologically-based pharmacokinetic (PBPK) models: state of the science and needs for research and implementation. *Toxicol. Sci. 99*, 395-402.

Bois, F. Y. (submitted). Physiologically-based modelling and prediction of drug interactions. *Basic Clin. Pharmacol. Toxicol.*

Bois, F. Y. and Paxman, D. (1992). An analysis of exposure rate effects for benzene using a physiologically based pharmacokinetic model. *Regul. Toxicol. Pharmacol. 15*, 122-136.

Bois, F. Y., Woodruff, T. J. and Spear, R. C. (1991). Comparison of three physiologically-based pharmacokinetic models of benzene disposition. *Toxicol. Appl. Pharmacol. 110*, 79-88.

Bois, F. Y., Zeise, L. and Tozer, T. N. (1990). Precision and sensitivity analysis of pharmacokinetic models for cancer risk assessment: tetrachloroethylene in mice, rats and humans. *Toxicol. Appl. Pharmacol. 102*, 300-315.

Bouvier d'Yvoire, M., Prieto, P., Blaauboer, B. J. et al. (2007). Physiologically-based kinetic modelling (PBK modelling): meeting the 3Rs agenda - The report and recommendations of ECVAM Workshop 63a. *ATLA 35*, 661-671.

Gupta, S., Manfro, R., Tomlanovich, S. et al. (1990). Effect of food on the pharmacokinetics of cyclosporine in healthy subjects following oral and intravenous administration. *J. Clin. Pharmacol. 30*, 643-653.

Howgate, E. M., Rowland Yeo, K., Proctor, N. J. et al. (2006). Prediction of in vivo drug clearance from in vitro data. I. Impact of inter-individual variability. *Xenobiotica 36*, 473-497.

Igari, Y., Sugiyama, Y., Sawada, Y. et al. (1983). Prediction of diazepam disposition in the rat and man by a physiologically based pharmacokinetic model. *J. Pharmacok. Biopharm. 11*, 577-593.

Jamei, M., Dickinson, G. L. and Rostami-Hodjegan, A. (2009a). A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of "bottom-up" vs "top-down" recognition of covariates. *Drug Metab. Pharmacok. 24*, 53-75.

Jamei, M., Marciniak, S., Feng, K. R. et al. (2009b). The Simcyp population-based ADME simulator. *Expert Opin. Drug Metab. Toxicol. 5*, 211-223.

Jamei, M., Turner, D., Yang, J. S. et al. (2009c). Population-based mechanistic prediction of oral drug absorption. *AAPS J. 11*, 225-237.

Jones, A. W. and Larsson, H. (2004). Distribution of diazepam and nordiazepam between plasma and whole blood and the influence of hematocrit. *Ther. Drug Monit. 26*, 380-385.

Kawai, R., Mathew, D., Tanaka, C. and Rowland, M. (1998). Physiologically based pharmacokinetics of cyclosporine A: extension to tissue distribution kinetics in rats and scale-up to human. *J. Pharmacol. Exp. Ther. 287*, 457-468.

Klotz, U. (1985). Estimation of the blood-plasma concentration ratio of diazepam in the rat. *Journal of Pharmacokinetics and Pharmacodynamics 13*, 347-348.

Klotz, U., Antonin, K. H. and Bieck, P. R. (1976). Pharmacokinetics and plasma protein binding of diazepam in man, dog, rabbit, guinea pig and rat. *J. Pharmacol. Exp. Ther. 199*, 67-73.

Klotz, U., Avant, G. R., Hoyumpa, A. et al. (1975). The effects of age and liver disease on the disposition and elimination of diazepam in adult man. *J. Clin. Invest. 55*, 347-359.

Loizou, G., Spendiff, M., Barton, H. A. et al. (2008). Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps. *Regul. Toxicol. Pharmacol. 50*, 400-411.

Parrott, N., Jones, H., Paquereau, N. and Lavé, T. (2005). Application of full physiological models for pharmaceutical drug candidate selection and extrapolation of pharmacokinetics to man. *Basic Clin. Pharmacol. Toxicol. 96*, 193-196.

Rostami-Hodjegan, A. and Tucker, G. (2004). "In silico" simulations to assess the "in vivo" consequences of "in vitro" metabolic drug-drug interactions. *Drug Discov. Today: Technol. 1*, 441-448.

Rostami-Hodjegan, A. and Tucker, G. T. (2007). Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nature Reviews Drug Discovery 6*, 140-148.

Tanaka, C., Kawai, R. and Rowland, M. (2000). Dose-dependent pharmacokinetics of cyclosporin A in rats: events in tissues. *Drug Metab. Dispos. 28*, 582-589.

Woodruff, T. and Bois, F.Y. (1993). Optimization issues in physiological toxicokinetic modeling - a case study with benzene. *Toxicol. Lett. 69*, 181-196.

Acknowledgements

This work was supported by the European Commission, 7th FP project PREDICT-IV (grant agreement #202222) and the French Ministry for the Environment (PRG189_08_DRC03).

Correspondence to

Frédéric Yves Bois
INERIS
Parc ALATA, BP2
60550 Verneuil en Halatte
France
e-mail: frederic.bois@ineris.fr

CAESAR's Approach for Alternative *In Silico* Methods for REACH

Anna Lombardo, Rodolfo Gonella Diaza, Alessandra Roncaglioni and Emilio Benfenati

Istituto di Ricerche Farmacologiche "Mario Negri", Milan, Italy

Summary

CAESAR is an EC funded project devoted to develop and make freely available quantitative structure-activity relationship (QSAR) models for REACH. Five endpoints have been addressed within CAESAR: bioconcentration factor in fish (BCF), skin sensitisation, mutagenicity, carcinogenicity and developmental toxicity (teratogenicity).

These models have been optimised to be reproducible, transparent, freely available, validated and tailored for REACH legislation, and they are now accessible for predicting toxicity and properties of industrial chemicals from their chemical structure through the project website (http://www.caesar-project.eu).

The models are based on experimental data collected from literature and checked, verifying chemical structures, in order to have a robust basis for the models. The models have been validated with external test sets and a series of statistical methods. These models have been implemented to use simple chemical information as input. An indication of the suitability of the model for a specific compound is given based on a similarity analysis.

Keywords: REACH, QSAR, predictive toxicology, in silico models

1 Introduction

REACH (Registration, Evaluation, Authorisation and Restriction of Chemical substances) is the new European legislation on chemicals and their safe use. The aim of REACH is to protect human health and the environment. To do this, it requires industry (manufacturers and importers) to provide detailed physicochemical, toxicological and ecotoxicological information (in function of the tonnage) on the chemicals they produce/import in order to ensure a safe use of these chemicals. The current data is insufficient, and REACH foresees the fulfilling of this data gap. Obviously, the production of the required data with classical toxicological tests implies the use of an increased number of animals for regulatory testing. One of the innovations introduced by REACH legislation is the emphasis on avoiding unnecessary testing and waste of animal lives, so data sharing, extrapolation from similar compounds and alternative testing methods – like *in vitro* or *in silico* ones – are encouraged to limit the amount of newly required tests (Regulation (EC) No 1907/2006).

The so-called "non-testing methods" include Quantitative Structure Activity Relationship (QSAR). QSAR takes advantage of already existing toxicological data to analyse *in silico* if a relationship can be established between the intrinsic chemical properties of the compounds already investigated with animal tests and their activity observed *in vivo*, for instance a specific toxicological behaviour. QSAR methods are already widely applied, for instance in the USA or Canada, for regulatory purposes (Jaworska et al., 2003).

Furthermore, many free and commercial tools are available. For instance, the OECD QSAR Toolbox (http://www.oecd.org/document/54/0,3343,en_2649_34379_42923638_1_1_1_1,00.html) provides data, similarity analysis, read-across and QSAR models and the Joint Research Centre in Ispra (Pavan and Worth, 2008) made available a series of tools to facilitate and promote the computational toxicology.

In the last years, thanks to new techniques, advanced algorithms and to new molecular descriptors, many new *in silico* models have been developed, which can in principle be used

to satisfy the requirements of REACH. However, models for regulatory purposes require special attention and validation. Although the REACH legislation promotes the use of alternative tests, there are still problems regarding the acceptability of the *in silico* models.

One of these is the data quality; good experimental data are necessary to build a good model, but databases with reliable data exist only for some endpoints. A second problem is the applicability domain of the models: each model had to establish for which chemicals it can give good predictions.

The CAESAR project aims to cover the lack of data using *in silico* models for five REACH-relevant endpoints chosen among those with the highest animal consumption. In this project *in silico* models were developed to predict the biological activity of chemicals for five REACH-relevant endpoints chosen among those with the highest animal consumption.

Tab. 1: List of endpoints ordered according to the total number of animals estimated to be used to satisfy REACH requirements as indicated by van der Jagt et al. (2004).

Endpoints
1. Two-generation reprotoxicity
2. Developmental toxicity studies
3. Further mutagenicity studies
4. Skin sensitisation
5. Long-term fish toxicity
6. Developmental toxicity screening
7. Accumulation
8. Short-term repeated dose
9. Carcinogenicity
10. Sub-chronic toxicity
11. Long-term repeated toxicity
12. Short-term fish toxicity
13. Acute inhalation toxicity
14. Acute dermal toxicity
15. Acute oral toxicity
16. *In vivo* eye irritation
17. *In vivo* skin irritation
18. Long-term bird toxicity

They were chosen using two criteria:
– To address those endpoints estimated to require the highest number of animals for testing overall
– To check the availability of experimental data for chemical compounds to be used as examples for building the predictive models

For this purpose the list of endpoints globally requiring the largest number of *in vivo* experiments within REACH was extrapolated from the publication compiled by van der Jagt et al. (2004) (Tab. 1). Some of the endpoints requiring the largest number of animals to satisfy REACH requirements could not be addressed with QSAR due to the paucity of experimental data. For this reason, on the basis of the criteria indicated above, i.e. the endpoints listed in Table 1 and data availability, the five endpoints chosen for the development of the quantitative structure-activity relationship (QSAR) models were:
– Bioconcentration factor,
– Skin sensitisation,
– Mutagenicity,
– Carcinogenicity,
– Developmental toxicity (teratogenicity).

Using non-testing methods it is possible to reduce the number of animal tests and the costs of the tests. Moreover, with QSAR it is possible to proceed in batch, testing many compounds at the same time, providing a rapid way of screening a large number of compounds to prioritise test need.

2 Materials and methods

Dataset preparation

Once the five endpoints were selected, although this choice was also driven by a preliminary analysis of available toxicological data, a dataset was prepared for each of the investigated activities. Starting from the data reported in the original references (Tab. 2), a time-consuming process of data checking was established to ensure data quality and to develop the QSAR models on a sound basis.

The quality of the data, which is important for all QSAR models, is even more fundamental in the case of models for regulatory purposes. For this reason, toxicity data from the original sources were checked and compared with other sources wherever possible.

Tab. 2: Characteristics of the five datasets used to develop QSAR models, their splitting into training and test sets and reference to the original experimental values.

Datasets	Total number of compounds for each dataset	Number of compounds in the training set	Number of compounds in the test set	Original source of experimental data
BCF	473	387	95	Dimitrov et al., 2005
Skin sensitisation	209	167	42	Gerberick et al., 2005
Mutagenicity	4204	3367	837	Kazius et al., 2005
Carcinogenicity	805	644	161	Gold et al., 2008
Developmental toxicity	292	234	58	Arena et al., 2004

Furthermore, besides the quality of the experimental data of the compounds, all chemical structures of these substances were processed in order to increase the quality of the data used as input. In this way we verified that quite a large number of compounds present in these data collections were not suitable for QSAR modelling, because the information on the chemical structure was incomplete or misleading. Within CAESAR much time was spent on checking all the chemical structures of the thousands of compounds listed in Table 2.

At least two independent persons verified the structures individually with the help of web browsers specific for chemical structure searching and verifying the matching with the IUPAC names. Some compounds were omitted if insufficient data were available to establish an accurate structure, or if they presented some characteristics not suitable for QSAR (e.g. mixtures, complexes).

Chemical descriptors and fragments were used to encode the chemical structure of each compound. Only 2-dimensional descriptors were used, because they are simpler and because no improvement in model results was found when using 3D-descriptors alone or combined with 2D-descriptors. 2D-descriptors are numerical values insensitive to a change in the molecular conformation, which occurs when the spatial arrangement of atoms in the molecule is modified through the free rotation of the atoms about single chemical bonds.

There are two main advantages of using 2D-descriptors in this kind of application (QSAR models for regulatory purposes):

1. They do not require optimisation of the conformation, which requires users with some computational chemistry skills and is most typically done manually or using stochastic approaches. Thus 2D-descriptors produce more reproducible results than 3D-descriptors. Reproducibility is a highly desirable feature in case of models for regulatory purposes.
2. They are calculated more rapidly.

Several main classes of 2D-descriptors were used for this project:

– Constitutional descriptors can be defined as descriptors reflecting the molecular composition of a compound without connectivity and geometry information.
– Topological descriptors are descriptors reflecting the molecular connectivity without geometry information.
– Fragments represent the counting of specific functional groups or atom types in a molecule.
– Physico-chemical descriptors refer to physico-chemical molecular properties, such as boiling point, melting point, water solubility, lipophilicity, free energy of solvation, etc. Among them, the octanol-water partition coefficient, LogP, when given as logarithm, has been the attribute used by most investigators to correlate structure and toxic effects. The octanol-water partition coefficient is a measure of the hydrophobicity of a substance and of its ability to pass cell membranes. It is important to mention that LogP cannot describe chemical features such as the compound reactivity, meaning the possibility for a substance to dissociate or react, giving a number of by-products.

These classes, together with the software package used for computing them in CAESAR, are summarised in Table 3.

Finally, each dataset was divided into a training set and a test set. The training set consisted of the chemicals provided as learning examples to the algorithms to search for the best relationship between descriptors and toxicity, while the test set represented a group of chemicals used to assess the predictivity of the QSAR models.

20% of the compounds were assigned to the test set according to the chemical composition of each dataset. Table 2 shows the number of compounds used for the five QSAR models and how they were split into training and test sets.

Modelling approaches

Within CAESAR we used methods to produce continuous values (regression methods) or categories (classifiers), depending on the endpoints, the kind of information available on the toxicological endpoint and REACH specifications. A wide range of algorithms was used, such as multiple linear regression (MLR), support vector machines (SVM), adaptive fuzzy partition (AFP), artificial neural networks (ANN), etc.. Genetic algorithm (GA) was in some cases used to select the most useful descriptors.

In some cases hybrid models were obtained, combining more than one mode or different modelling philosophies. This improved the overall performances.

Implementation of models

The models (which means the algorithms which calculate the property based on chemical descriptors) developed within CAESAR are freely available and were implemented using Java. In most cases we had to re-write the models in order to avoid the use of commercial software for the calculation of chemical descriptors. We made a special agreement with the developers of Dragon, and we also used descriptors developed by the US EPA within a collaboration agreement.

Tab. 3: Classes of 2D-descriptors used for modelling purposes and software package used for computing them.

2D-descriptor classes	Software tools
Constitutional	Codessa[a] Dragon[b] MDL QSAR[c]
Topological	Codessa[a] Dragon[b] MDL QSAR[c]
Fragments	Dragon[b]
Physico-chemical	ACD[d] Pallas[e] Dragon[b]

[a] Codessa, v.2.21, Gainesville, FL: University of Florida, 1995.

[b] Dragon, v.5.4, Talete Srl, Milan, Italy, 2005.

[c] MDL QSAR v.2.2, MDL Information Systems Inc., San Leandro, CA, 2003.

[d] ACD/Lab v.10.0, LogD Suite. Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2006.

[e] Pallas v.3.0, PrologD module. CompuDrug Chemistry Ltd., Budapest, Hungary, 1995.

In this way, it has been possible to develop a special web site, which can be used to calculate the property of interest starting from the simple chemical structure. The file formats supported are SMILES string and sdf files explicitly including all hydrogen atoms.

The application developed is a Java based platform that includes one model for each endpoint. It has been developed like a container in which all the integrated models are handled as separate blocks, like plug-ins, that can be updated separately from the rest of the software. This kind of structure allows integrating more models into the application in the future.

The application has been developed as a Client-Server application. This means that the user interacts directly with a graphic interface (the client-side application), where it is possible to select the molecules for which toxicity is to be predicted, visualise the results and export results/reports. The client interacts via internet connection with the server on which the server-side application is installed. This application integrates the predictive models as modules. The server-side application takes the molecules as inputs from the client, submits them to the correct module (where all calculations are made) and then sends the results to the client application.

Tool for similarity and applicability domain

The safe use of QSAR models for regulatory purposes should take advantage of the availability of tools to assess if a QSAR model can be used for the particular chemical of interest. Indeed, a QSAR model for regulatory purposes has to be a good model from a technical point of view (general condition for the quality of the model) and for the specific application for REACH (and thus has to be optimised for that purpose). But this is not enough, because this refers to a general assessment of the model. A hypothetical user might be interested in a single compound or in a series of chemicals. Thus, the general assessment should be focussed on the specific interests of the user.

For this purpose, we developed an additional tool capable of showing the chemicals present in the training set of the model that are most similar to the compound of interest, along with their predicted and experimental values, and the value of a similarity index. In this way, the user can see and evaluate in a very transparent way whether the queried compound is reasonably similar to any used to optimise the model, and whether predictions for similar compounds were good or not. The similarity index is calculated automatically by Dragon software on the basis of the chemical structure described by a pool of chemical descriptors and fragments.

3 Results

The applet

Many models have been developed based on commercial software for computing the chemical descriptors. So far, one model for each endpoint was implemented: one model providing continuous values for BCF and four models classifying the compound as toxic or non-toxic for the remaining endpoints. The classification models are skewed in the direction of being conservative (avoiding as much as possible the situation of assuming the compound to be non-toxic while it is toxic). Anyway, for the models developed for the most complex endpoints (particularly carcinogenicity and developmental toxicity) the role of QSAR models is as support for a general assessment, which is based on a series of data and information also in view

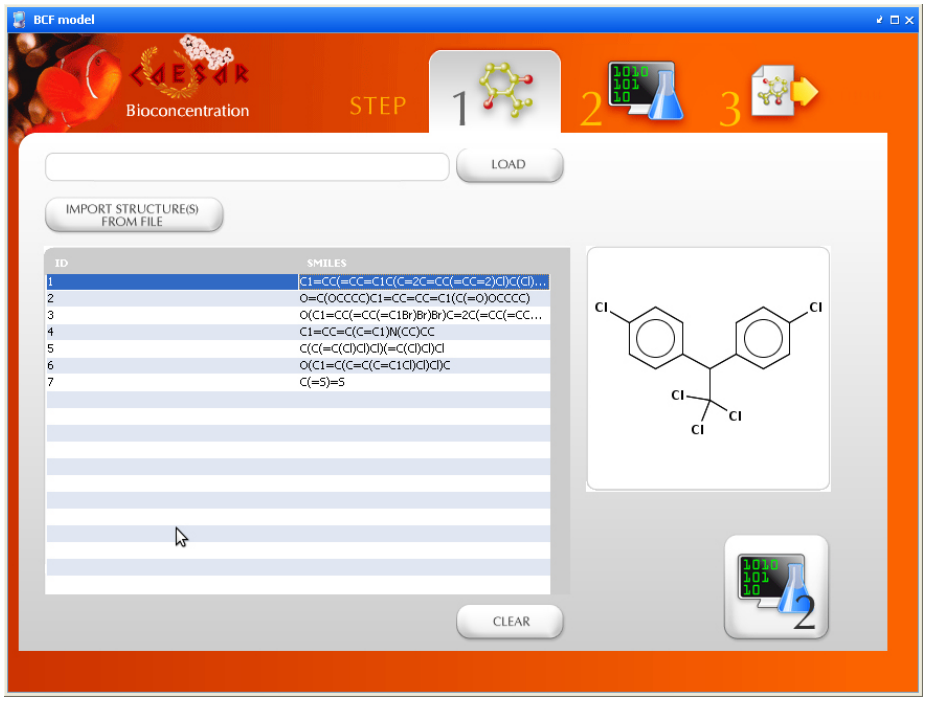


Fig. 1: A screenshot of the panel for using the CAESAR model.
Example of the first step (importing molecules) applied to the BCF model.

of the quite limited number of the compounds available to build up the model.

In the future, other complementary models might be implemented to offer a more balanced view or to enlarge the chemical space covered by these models.

In Figure 1 is shown the graphical interface for the use of the CAESAR models using the BCF as example. Compounds might be inserted manually or copied from other sources and then uploaded. Alternatively they can be imported from other files in one of the two supported file formats (.smi or .sdf files). The list of compounds and their SMILES codes (independent of the uploaded file format) are displayed in the table with the chemical formula of the selected compound (highlighted with a coloured background) on the right hand side.

Moving to the second step (Fig. 2), results are displayed both in terms of the intrinsic units used in the model (logarithmic ones) and the predicted final BCF value. Some remarks might also be displayed highlighting those predictions that might be unreliable because they are out of the applicability domain of the model.

For each compound, in addition to the prediction of the endpoint properties, the six most similar molecules found in the whole dataset used for building and testing the model can be displayed. From the analysis of these chemicals it is possible to verify how well the model was behaving in that specific chemical space and to speculate about the consistency of the prediction.

Finally, results can be exported in a tabulated format (very useful if many compounds have been calculated) or in a more detailed .pdf format that maintains all the details about each single prediction (remarks and similar compounds). An example of the .pdf report is shown in Figure 3.

The quality protocol

The main aims of quality protocol adopted were:

- Conformity: The models conform to requirements.
- Reliability: The models produce a reliable prediction, according to specified criteria.
- Reproducibility: The models have to give the same results when tested by a second user.
- Transparency: The deliverables and all model components must be well documented; all steps, processes, algorithms and a theoretical description of the system are accessible to the interested potential user.
- Testability: Rigorous tests for validation have been performed.
- Efficiency: The models must use resources in the best and most effective manner.
- Portability: The software can be easily ported on different environments and computers.

All these conditions were verified. As already described, all structures were checked individually, and data were evaluated in order to use values of high quality only.

As described, these models have been implemented using Java. As a result, all CAESAR models have been rebuilt in this language, and this required a complete check of all the equations and results. Thus, this activity also addressed the reproducibility of the results, repeating the model-building in another situation.

The quality control of the final model involved also the common information technology testing, with debugging activities, and checks done with up to 900 compounds tested simultaneously.

The models were tested on different browsers and operative systems.

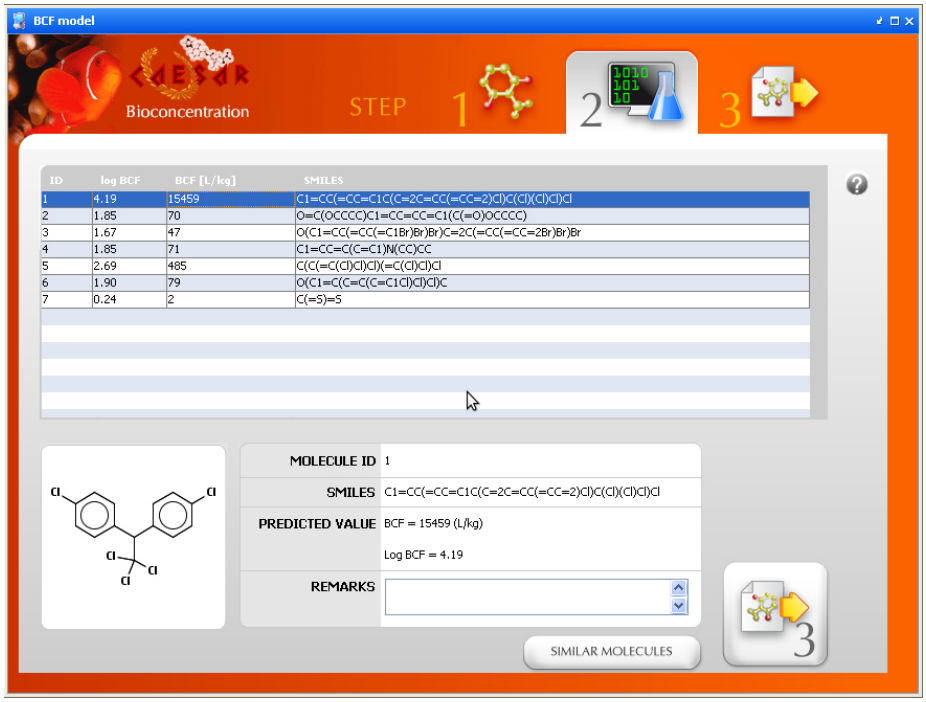
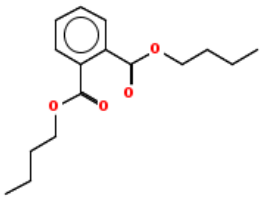


Fig. 2: A screenshot of the panel for using the CAESAR model.
Example of the second step (display of the predicted toxicities) applied to the BCF model.

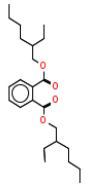
CAESAR QSAR model for bioconcentration factor (BCF) in fish

Prediction for the compound no. 2: O=C(OCCCC)C1=CC=CC=C1(C(=O)OCCCC)



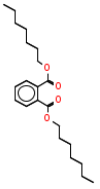
BCF value: 70 (L/Kg) whole body weight
Log BCF value: 1.85
Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



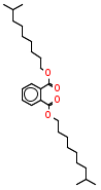
Dataset id: 300
SMILES: O=C(c1c(cccc1)C(=O)OCC(CCCC)CC)OCC(CCCC)CC
Similarity: 0.88

Experimental Log BCF: 1.19
Predicted Log BCF: 1.51



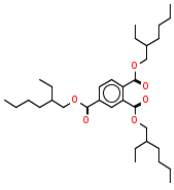
Dataset id: 301
SMILES: O=C(c1c(cccc1)C(=O)OCCCCCCC)OCCCCCCC
Similarity: 0.874

Experimental Log BCF: 1.06
Predicted Log BCF: 1.43



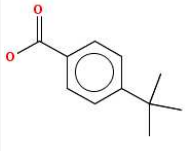
Dataset id: 302
SMILES: O=C(c1c(cccc1)C(=O)OCCCCCCCC(C)C)OCCCCCCCC(C)C
Similarity: 0.798

Experimental Log BCF: 1.16
Predicted Log BCF: 0.77



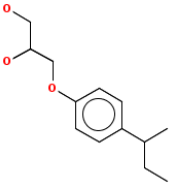
Dataset id: 303
SMILES: O=C(c1c(ccc(c1)C(=O)OCC(CCCC)CC)C(=O)OCC(CCCC)CC)OCC(CCCC)CC
Similarity: 0.719

Experimental Log BCF: 0.27
Predicted Log BCF: 0.74



Dataset id: 111
SMILES: CC(c1ccc(cc1)C(=O)O)(C)C
Similarity: 0.698

Experimental Log BCF: 0.21
Predicted Log BCF: 1.54



Dataset id: 93
SMILES: CCC(c1ccc(cc1)OCC(CO)O)C
Similarity: 0.693

Experimental Log BCF: 0.63
Predicted Log BCF: 0.71

Fig. 3: Example of the format containing the report of the results for one molecule.

4 Discussion

CAESAR developed new QSAR models specific for REACH for the following endpoints: bioconcentration in fish, skin sensitisation, mutagenicity, carcinogenicity, developmental toxicity.

These models have been planned to allow easy access and use: industry and regulators can freely use the models via the Internet. Three simple steps allow the user to introduce the chemical(s), get the prediction, and save or print it. This simplicity does not mean that the models are weak. Instead it shows that the effective use of the models has been one of our main concerns.

The CAESAR models are sound and scientifically based. Global data on about 7,000 compounds have been used to build up the models. The chemical structures were double-checked individually, even though their sources were of high quality. We took into consideration the regulatory purpose of the models, not only in the quality check of the input data but also in assessing the intended use of the models, for instance verifying that the number of false negatives was low, and a similar approach was used for the continuous values.

Many tens of different models were developed, and we chose the best of them. Validation of the models was done according to sound statistical criteria, evaluating internal and external validation (external test set).

The models provide guidance on their applicability. For this, different approaches have been used. We defined a series of chemical groups/classes for which the model does not offer optimal performance. Furthermore, the models automatically highlight if any reason for concern for the ideal use of the model appears, e.g. boundaries of the descriptor space. Finally, the model shows the predicted value/category, the experimental one and the error of the six most similar compounds in relation to the target compound for prediction. Thus, the user has a clear, immediate appreciation whether similar compounds were included when developing the models and how accurate the predictions were for them. A similarity score between compounds is also provided to guide the user. The models can run hundreds of compounds in batch within a few seconds, starting from very simple and easily available chemical formats such as SMILES and .sdf.

Thousands of predictive models have been published in recent years, but typically they are not suitable for regulatory purposes, because they have not taken into account one or more factors that are essential for validation or quality assurance or for the specific application, which is a very delicate matter. The typical approach is to develop a model and then propose its use. However, it is clear that this is a generic approach, not addressing the specific application it should be dedicated to.

The particular application of REACH needs to consider the different stakeholders. This means that the issues to be evaluated are not only theoretical but also practical. Furthermore, other aspects, such as cultural barriers and ease of use should be investigated.

CAESAR addressed all these critical factors. We dedicated special attention to the users and to the access to the models. For this reason we preferred to develop models that are simple to use and which give reproducible results.

We underline that other approaches have preferred to develop flexible models, which can give different results for the same compound depending on the user. This approach, in our opinion, has some drawbacks in the specific case of models for regulatory purposes, because different results can be obtained, for instance, by the industrial or regulatory user, or in different countries. This complicates the assessment of the results and requires a successive, separate, evaluation of the results.

We preferred to follow an approach like that used by the US EPA. The same model is available to all actors: industry and regulators. Thus, the same rules apply and the model is transparent and reproducible. This provides the maximum possibility of exploitation, because the model is directly suitable for the regulatory application. The models can be used to process a large number of chemicals in batch. This fact also increases the possible use and thus the impact of the models.

Industry can use the models to provide data for REACH. The same model is available to regulators, who can verify the correctness of the values. The models have been optimised to reduce false negatives, and this also increases the acceptability for the regulatory bodies. Furthermore, models can have an impact on industry, because they can use the model to plan new compounds, and assess in advance if a certain structure may present critical aspects. Thus, the impact for industry is high.

The models are freely available, and this makes their use highly competitive. Indeed, some of the addressed endpoints are expensive when done experimentally. It has been criticised that REACH will impose a high cost on industry, but the use of CAESAR's model will solve this issue. Furthermore, as explained, the use of the models can anticipate future potential problems related to certain chemical structures. As a result, industry will have a powerful tool to produce safer chemicals, which is the ultimate target of REACH.

Regulators will also benefit from CAESAR's models. Indeed, the models are easy to use, can be adopted for checks of all chemical structures (with restrictions as defined), and, in addition, they can be used to screen large numbers of compounds.

Recently, it was found that the number of chemicals potentially requiring evaluation, on the basis of the pre-registration of the compounds, is much higher than was expected. Originally, the number of dossiers which were to be evaluated by regulators was quite low, a percentage of the dossiers submitted by industry. The higher number of pre-registered chemicals may further reduce the actual percentage of checked dossiers. From this perspective, the availability of tools, such as those developed by CAESAR, is highly beneficial for regulators, as they allow easy and rapid processing of a high number of compounds, even all of those pre-registered.

Thus, the potential impact of CAESAR's models, if adopted by regulators for screening purposes, is that regulators could easily process all compounds to identify critical issues with

some of them. This can improve the evaluation of the chemicals for the European market, and so minimise the possibility of having toxic compounds on the market.

The models developed within the CAESAR project can be used to fulfil the requests of REACH, because they include all those high quality factors that are needed to make the use of (Q)SARs acceptable for regulatory purposes, and all components of the models were carefully assessed and described. Moreover, the CAESAR models cover the most important endpoints for REACH for which no suitable models exist.

The models use an innovative approach: the automatic calculation of a wide range of physicochemical properties for each chemical from a 2D structure. The procedure is reproducible and fast and avoids any error associated with manual operations.

Besides the specific models, the impact of CAESAR lies also in the development of a platform for the QSAR models. The use of this platform is very simple, because calculation is done in the background. Results are clearly presented, indicating similar compounds and assessing the applicability domain. This platform is now available to integrate further models in the future.

The CAESAR project has had the positive impact of merging different scientific communities to work with regulators and stakeholders. The gathering of scientists and other stakeholders under CAESAR and the increased use of a common forum has further established networks and defined common targets between communities with diverse interests.

References

Arena, V. C., Sussman, N. B., Mazumdar, S. et al. (2004). The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: comparative analysis of logistic regression and decision tree models. *SAR QSAR Environ. Res. 15*, 1-18.

Dimitrov, S., Dimitrova, N., Parkerton, T. et al. (2005). Baseline model for identifying the bioaccumulation potential of chemicals. *SAR QSAR Environ. Res. 16*, 531-554.

Gerberick, G. F., Ryan, C. A., Kern, P. S. et al. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis 16,* 157-202.

Gold, L. S., Slone, T. H., Williams, C. R. et al. (2008). DSSTox carcinogenic potency database summary tables – all species: SDF file and documentation. *Version*: CPDBAS_v3b_1481_10Apr2006, www.epa.gov/ncct/dsstox/sdf_cpdbas.html.

Jaworska, J. S., Comber, M., Auer, C. and Van Leeuwen, C. J. (2003). Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environ. Health Perspect. 111*, 1358-1360.

Kazius, J., McGuire, R. and Bursi, R. (2005). Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem. 48*, 312-320.

Pavan, M. and Worth, A. (2008). Publicly-accessible QSAR software tools developed by the Joint Research Centre. *SAR QSAR Environ. Res. 19*, 785-799.

Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L 396, 30.12.2006, 1-849.

van der Jagt, K., Munn, S., Tørsløv, J. and de Bruijn, J. (2004). Alternative approaches can reduce the use of test animals under REACH. Addendum to the report: Assessment of additional testing needs under REACH Effects of (Q)SARS, risk based testing and voluntary industry initiatives, *JRC Report*, EUR 21405 EN.

Acknowledgements

We are indebted to the entire research team involved in the CAESAR project and comprising tens of scientists in nine laboratories across Europe for their valuable contribution. We thank US EPA for the collaboration with Dr A. Richard (Research Triangle Park) for the data checking and with Dr M. Todd and Dr D. Young (Cincinnati) for the software to calculate chemical descriptors. We gratefully acknowledge the financial contribution of the European Commission's CAESAR (Contract SSPI-022674) project.

Correspondence to

Dr. Emilio Benfenati
Istituto di Ricerche Farmacologiche "Mario Negri"
via G. La Masa 19
20156 Milan, Italy
e-mail: emilio.benfenati@marionegri.it

The *In Silico* Model for Mutagenicity

Giuseppina Gini[1], Thomas Ferrari[1] and Alessandra Roncaglioni[2]

[1]DEI, Politecnico di Milano, Italy; [2]Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy

Summary

Mutagenic toxicity is the capacity of a substance to cause genetic mutations. This property is of high public concern, because it has a close relationship with carcinogenicity and reproductive toxicity.
In experiments, mutagenicity is assessed by the Ames test on Salmonella. The estimated inter-laboratory reproducibility rate of Salmonella test data is only 85%. This shows the intrinsic limitation of the in vitro test and opens the road to other assessments like an in silico model.
So far a widely used method is to check for the presence of structural alerts (SAs). However the presence of SAs alone is not a definitive method to prove the mutagenicity of the compound; the substituents can change the classification.
So statistically based methods are developed with the final target of obtaining a cascade of systems with tailored properties.
The integrated system has been developed on a set of a few thousand molecules.

Keywords: in silico methods, QSAR, intelligent systems

1 Introduction and questions

Mutagenic toxicity is the capacity of a substance to cause genetic mutations. This property is of high public concern, because it has a close relationship with carcinogenicity and reproductive toxicity (Benigni et al., 2008): most mutagenic substances are suspected carcinogenic substances in case a genotoxic mechanism is considered.

In experiments, mutagenic toxicity can be assessed by various test systems. One crucial point was the creation of cheap and short-term alternatives to the rodent bioassay, the main tool of the research on chemical carcinogens. With this intent Bruce Ames created a series of genetically engineered *Salmonella typhimurium* bacterial strains, each strain being sensitive to a specific class of chemical carcinogens (Ames, 1984). The Ames test is an *in vitro* model of chemical mutagenicity and carcinogenicity and consists of a range of bacterial strains that together are sensitive to a large array of DNA-damaging agents (Ashby, 1985).

An interesting point is the reliability of such experimental tests: as discussed in other papers (Piegorsch and Zeiger, 1991) the estimated inter-laboratory reproducibility rate of Salmonella test data is 85%. This observation will be taken into account in our model that will make use of the available data without using new *in vitro* testing.

Today regulators request the availability of mutagenic potency to correctly label and restrict mutagens/carcinogens and the exposure to them. Another important use of mutagenicity testing is in drug discovery, where mutagens should be stopped in the development of drugs. In environmental protection regulators need to understand the mutagenic potential of chemicals in order to control or limit their use.

Our aim is to develop a QSAR (Quantitative Structure Activity Relationship) model based on available data and to develop it for regulatory purposes; i.e. to reduce the number of false negatives. There is an argument that, if the main aim of QSAR modelling is simply prediction, the attention should be focused on model quality and not on its interpretation. Another argument is that it is dangerous to attempt to interpret models, since correlation does not imply causality, as discussed in (Livingstone, 2000).

On this basis, we can differentiate predictive QSARs, where the focus is best prediction quality, from descriptive QSARs, where the focus is descriptor interpretability. Our first aim is predictive QSAR; however our model is also quite interpretable.

The first step in making a QSAR model is the calculation of molecular descriptors. We limited our models to descriptors computed using the MDL software, including general descriptors and fingerprints. Fingerprints are used to encode structural characteristics of a chemical compound into a fixed bit vector (Durant et al., 2002).

QSARs have already been developed for mutagenicity. The availability of large data sets of non congeneric compounds, the most notable provided and analysed by Kazius et al. (2005), makes it possible to construct more robust models. A few other papers have been already published on this core data set (Liao et al., 2007; Zheng et al., 2006).

2 Materials and Methods

Two methods for predictions are used.
The first one consists in detecting in the molecule the particular structural fragments already known to be responsible for the toxic property under investigation. In the mutagenicity/carcinogenicity domain, the key contribution in the definition of such toxicophores comes from Ashby (Ashby, 1985) and is grounded on the electrophilicity theory of chemical carcinogenesis developed by Miller and Miller (1981). Every subsequent effort starts from knowledge collected by Ashby to derive more specific rules. It is important to mention that so far the mutagenicity structural alerts (SAs) are sound hypotheses that derive from chemical properties and have a sort of mechanistic interpretation; however their presence alone is not a definitive method to prove the mutagenicity of the compound towards Salmonella; the substituents present in some cases are able to change the classification. SAs are in practice rules that state the condition of mutagenicity given the presence and the absence of peculiar chemical substructures.

The second method we propose uses statistics, in particular an effective method to build non-linear models, as developed under the name Support Vector Machines (Vapnik, 1995).

The integration of both methods will give the final result, making the QSAR both predictive and interpretable.

2.1 Data

Following quality checks (IRFMN and CSL) the Kazius database was pruned and modified to 4225 compounds: 2358 classified as mutagens and 1867 classified as non-mutagens by the Ames test. For developing and evaluating the model we split them into a training set (80%) and a test set (20%). For each compound molecular descriptors were calculated with MDL-QSAR software, including both substructures and global descriptors.

A subset of 27 descriptors has been automatically selected with the BestFirst search method, using as subset evaluator the 5-fold cross-validation score on the training set. In short, BestFirst algorithm searches the space of attribute subsets by greedy hill climbing (considering all possible single attribute additions and/or deletions at a given point), with a backtracking facility to explore also non-improving nodes. The same subset of 27 descriptors has been obtained either searching forward, starting from the empty set, or with a bi-directional search starting from the 10 top rated attributes by a single attribute evaluator (Relief), both with 3 steps of backtracking.

The resulting dataset has been normalised by dividing each descriptor column by its maximum absolute value. Table 1 shows the selected substructural descriptors, while Table 2 shows the global descriptors. It is interesting to explain their meaning.

– *Gmin* = the minimum atom_level E-state value in a molecule. The E-State descriptor Gmin is a measure of the most electrophilic atom in the molecule and the polarity of the molecule. Mechanistically, an electrophilic centre is important for covalent bond formation with nucleophilic DNA, and so it is not surprising that Gmin is found to be important in modelling.

Tab. 1: The 23 local descriptors.
Modelling method for a statistical QSAR: Support Vector Machines

Symbol	Definition
$SsCH_3$_acnt	Count of all ($-CH_3$) groups in molecule
$SdCH_2$_acnt	Count of all ($=CH_2$) groups in molecule
$SssCH_2$_acnt	Count of all ($-CH_2-$) groups in molecule
SdsCH_acnt	Count of all (= CH –) groups in molecule
SaaCH_acnt	Count of all (CH) groups in molecule
SsssCH_acnt	Count of all (> CH –) groups in molecule
SdssC_acnt	Count of all (= C <) groups in molecule
SaasC_acnt	Count of all (= CH =) groups in molecule
SaaaC_acnt	Count of all (= CH =) groups in molecule
SssssC_acnt	Count of all (> C <) groups in molecule
$SsNH_2$_acnt	Count of all ($-NH_2$) groups in molecule
StN_acnt	Count of all (≡ N) groups in molecule
SdsN_acnt	Count of all (= N –) groups in molecule
SaaN_acnt	Count of all (= N =) groups in molecule
SsssN_acnt	Count of all (> N –) groups in molecule
SdaaN_acnt	Count of all (= N =) groups in molecule
SsOH_acnt	Count of all (– OH) groups in molecule
SdO_acnt	Count of all (= O) groups in molecule
SssO_acnt	Count of all (– O –) groups in molecule
SaaO_acnt	Count of all (≡ O ≡) groups in molecule
SHsOH_Acnt	Count of all [– OH] groups in molecule
SHother_Acnt	Count of all [other] groups in molecule
SHCHnX_Acnt	Count of all Halogen on C with 1 or 2 H atoms

– *idwbar* = Bonchev-Trinajstic mean information content based on the distribution of distances in the graph
– *LogP* = partition coefficient between octanol and water
– *nrings* = Number of rings in a molecular graph: cyclomatic number (i.e. the smallest number of bonds which must be removed such that no ring remains)

Atom types are classifications based on element and bonding environment. Atom type assignments are used in functional group identification, hydrogen addition, and hydrogen bond identification, and to determine VDW radii.

Except for the first capital "*S*", each lower case letter represents a bond:

– each "s" within an atom type designation represents a single bond to that atom
– each "d" within an atom type designation represents a double bond to that atom
– each "t" within an atom type designation represents a triple bond to that atom
– each "a" within an atom type designation represents an aromatic bond to that atom

We can observe that a few of them match known SAs.

The SdsN descriptor (for the nitrogen atom type \-N\=) is associated with the azo group, a structural alert. Molecules with larger SdsN descriptor values tend to have larger calculated output values.

SsssN is the atom count of all tertiary nitrogens in molecules. Tertiary nitrogen group alerts occur when the nitrogen is attached to either an aromatic or partially unsaturated ring. SaasC counts aromatic carbons with an attached substituent atom. It is not an alert per se; however, it reflects the nature of structural alerts attached to the ring system.

2.2. The statistical model

We created the statistical model using Support Vector Machines (SVM). SVM can use linear models to implement non-linear class boundaries. The input space is mapped into a higher (maybe infinite) dimensional space by a function ϕ, and a linear model constructed in the new space can represent a non-linear decision boundary in the original space. In the transformed space, the algorithm calculates the maximum margin hyperplane, i.e. the linear model that gives the greatest separation between the classes. The instances that are closest to it are called *support vectors*.

In our model we chose the *Radial Basis Function* as the kernel. A complete environment to develop SVM models is the open source LibSVM library[1], containing C++ and Java implementation of SVM algorithms with high-level interfaces (Python, Weka and more).

The optimal parameterisation of the model can be fully automated by one of the scripts included in LibSVM. With this tool it is possible to perform an almost exhaustive grid-search in the 2-dimensional parameter space of the objective function, using as evaluation criterion a cross-validation on the training set: the best assignment found was $(C, \gamma) = (8, 8)$.

With this parameterisation a model was trained on the training set, and its prediction ability was evaluated on the test set, normalised with the same scale factors used for the training set. Moreover, its robustness was assessed by a stratified 10-fold cross-validation. Table 3 reports the accuracy of the obtained model.

The accuracy of the model is very high. However, for the scope of CAESAR, we may try to reduce the FN rate. To this end we can apply another check and see if some widely known SAs can be used to detect other mutagens.

2.3. A model using SAs

As stated above, the available knowledge on mutagenicity is expressed in terms of SAs. A recent compilation of those alerts has been chosen as the knowledge base of our approach. To this end we considered the set of 30 SAs for mutagenicity derived by Benigni and Bossa from several literature sources. This rulebase is implemented as a module of Toxtree. (Developed by Ideaconsult Ltd. under the terms of a JRC contract. Software available at http://ecb.jrc.ec.europa.eu/qsar/: a java open source wrapper for structure-based predictions inclusive of a few other plugins on some toxicological endpoints.) In this realisation, the SAs are coded into SMARTS (SMiles ARbitrary Target Specification) strings, and the compounds in SMILES strings. Therefore the SAs detection is accomplished basically as a SMARTS[2] matching task. SMARTS strings are a text representation of substructures. To be matched, both the SMILES and the SMARTS strings are translated into graphs and the two graphs are compared.

The Benigni/Bossa rulebase was evaluated on the same set of 4225 chemical structures of CAESAR. The prediction ability on the entire data set, compared with the respective Ames Test results, is summarised in Table 4.

With respect to the SVM model, the SA model shows an increase of 48% of FP and an increase of 5% of FN. However, even if good sensitivity is exhibited by a low False Negative (FN) rate, the toxicity is often overestimated (low specificity), compromising the overall performance. This highlights the apparent drawback of the SAs: their compilation is engineered to individuate candidate mutagens by detecting the presence

Tab. 2: The 4 global descriptors

MDL code	Definition
MDL187	Smallest atom E-State value in molecule
MDL198	Bonchev-Trinajstic mean information content
MDL226	Calculated value of LogP
MDL230	Number of rings (cyclomatic number) in a molecular graph

Tab. 3: Accuracy of the SVM model

TRAINING SET *3380 chemicals*	**Predicted** ***mutagen***	**Predicted** ***non-mutagen***
mutagen	***1766***	***122***
non-mutagen	*137*	***1355***
Correct classification rate: ***92.3%***		
TRAINING SET *845 chemicals*	**Predicted** ***mutagen***	**Predicted** ***non-mutagen***
mutagen	***407***	***63***
non-mutagen	*79*	***296***
Correct classification rate: ***83.2%***		

[1] available at http://www.csie.ntu.edu.tw/\textasciitilde{}cjlin/libsv

[2] Daylight Theory Manual available at http://www.daylight.com/

of (presumed) toxicophores; just the remainder is labelled as non-mutagen.

3 Final results

So far we have seen how it is possible to achieve good prediction ability through a statistical approach, and also quite good results in predicting Salmonella mutagenicity through Toxtree software.

How can be these models be made more suitable for regulatory purposes? An answer is to address with special care the reduction of FNs, the hazardous compounds predicted as safe. There are various tricks to implement such enhancement in learning algorithms simply by throwing the model off centre, but all attempts in this direction will unavoidably raise several new FPs for each FN removed, since a just trained model is already in its best equilibrium. From here arose the idea of a trained classifier supervised by an expert layer: the aim is to refine the good statistical separation between classes supplied by SVM, not by introducing a perturbation in the optimality of the model, but by applying a complementary knowledge-based filter in order to allow an accurate identification of misclassified mutagens (FNs), even if isolated. In other words, instead of making the model more (or too much) sensible to mutagens, our intent is to equip it with an additional device to be applied to non-mutagenic predictions, skilled in what it has had difficulty to learn.

In practice, although every SA should be trusted if evaluated on a random set of molecules, here we are considering only that portion of compounds already presumed non-mutagenic by SVM, i.e. cleaned for the most part of mutagens. This means that while the FP rate spawned by each rule is unchanged (since all the non-mutagens should still be present), the rate of caught true positives (TPs) will decrease, because just a few mutagens are left. Hence a selection is needed to extract just a subset of the rulebase skilled in finding the mutagens potentially subject to misclassification by the SVM model. Having such a large data set, the above selection can be carried out in a straightforward way. The predictions obtained by cross-validating the model on the training set (3380 compounds) shall be representative of its general prediction ability, so a filter fixing the inaccuracies of such a meta-model will probably provide even for defects of the original one.

An integrated model was arranged cascading the two techniques: a trained SVM classifier with an additional expert facility for FNs removal based on SAs. The SVM classifier is the one described previously, while the rulebase for the expert filter was extracted from the Benigni/Bossa SAs set after an analysis of their individual effect, evaluated on those structures of the training set labelled non-mutagenic by 10-fold cross-validating the model.

This spotlights two different subsets of SAs (see Appendix). The former (10 SAs) is the set of "good" rules: each of them showed a balance of more FNs caught than FPs spawned once evaluated on the cross-validated predictions. Their supposed capacity to refine the SVM model prediction ability is confirmed by the proof on the test set. The latter (5 SAs) is the set of "suspicious" rules, i.e. those ones with a still remarkable FNs removal power but a higher misclassification rate. As can be seen in Table 5, the FNs removal carried out by the first set of rules improved both sensitivity and accuracy similarly, either in the calibration on the training set or the validation on the test set. 13% of FNs are cleaned from the "safe" prediction on the test set and this benefits even the correctly classified prediction rate with a slight increment.

By applying the second set of rules (Tab. 6), the performances in classification accuracy are not noticeably downgraded if compared with those of the basic SVM model, but about a third of FNs (32%) are removed from the overall predictions, boosting the sensitivity over 90%.

The statistics on the final model are described in Table 7.

A global overview of the performances of the combined model is illustrated in Figure 1, where an interpretation of the set of "suspicious" rules is given: it can extract the more suspect

Tab. 4: Results of the SA model on the test set

TEST SET *845 chemicals*	Predicted *mutagen*	Predicted *non-mutagen*
mutagen	***404***	***66***
non-mutagen	*117*	***258***
Correct classification rate: ***78.3%***		

5: Confusion matrix on the test set after using the first set of 10 rules

TEST SET *845 chemicals*	Predicted *mutagen*	Predicted *non-mutagen*
mutagen	***427***	***43***
non-mutagen	*109*	***267***
Correct classification rate: ***82.1%***		

Tab. 6: Confusion matrix after applying both sets of rules

TEST SET *845 chemicals*	Predicted *mutagen*	Predicted *non-mutagen*
mutagen	***415***	***55***
non-mutagen	*86*	***289***
Correct classification rate: ***83.3%***		

Tab. 7: Statistics about the cascade model

CAESAR test set	**Suspicious taken as non-mutagenic**	**Suspicious taken as mutagenic**
accuracy:	***83.3%***	***82.1%***
sensitivity:	***88.3%***	***90.9%***
specificity:	***77.1%***	***71.2%***

compounds from "safe" prediction with good accuracy, if related to the very low number of real mutagens still present. The so obtained global model for mutagenicity prediction has been released through the portal of the CAESAR project.

Our final cascade model is described in Figure 2. Its logic flow is very clear, and its statistical analysis defined.

4 Discussion

Our hypothesis that a QSAR approach was a good method to build models of non-congeneric compounds has been proven. The two-step method proposed in our CAESAR model for mutagenicity demonstrated that the QSAR method is more apt to screen the data set than the SAs approach.

In our implementation, besides improving accuracy, we are also biased toward reducing the number of false negatives, as required by regulators. As we have seen, the first screening is based on statistical correlation between small fragments and the mutagenicity property, and we only check the presence of SAs upon a negative outcome. Since regulators use this method manually, we are so able to provide them with a similar check.

However, there are important differences between our method and the traditional SAs. As we already pointed out, the known SAs are biased toward pollutants and carcinogenic molecules; in our approach we can use only the first screening to deal with drugs and other families of compounds. SAs are a fixed list of substructures, while the MDL keys used in the QSAR phase are in a number automatically selected to give better performance to the correlation with the endpoint. So the keys are automatically derived from the reduction algorithm, not extrapolated from human experts. They can cover or not the known SAs. In this sense we are really performing data mining and deriving a set of keys that can become, in principle, new SAs in case the chemical classes considered are new and an interpretation about reactivity is available.

In terms of accuracy our model, which uses powerful algorithms, can reach accuracy very near to the rate of the reproducibility of the experimental data in different laboratories.

In terms of interpretability of our model, the first step can be understood in terms of the few global descriptors used and the MDL keys. However we should remember that the interpretability of non-linear models does not depend on simple relations between input and output, and the mix of the descriptors cannot be translated into rules. The second step is obviously defined in terms of rules, stating that the presence of any of the SAs and the absence of external conditions would put the compound in the mutagenicity class.

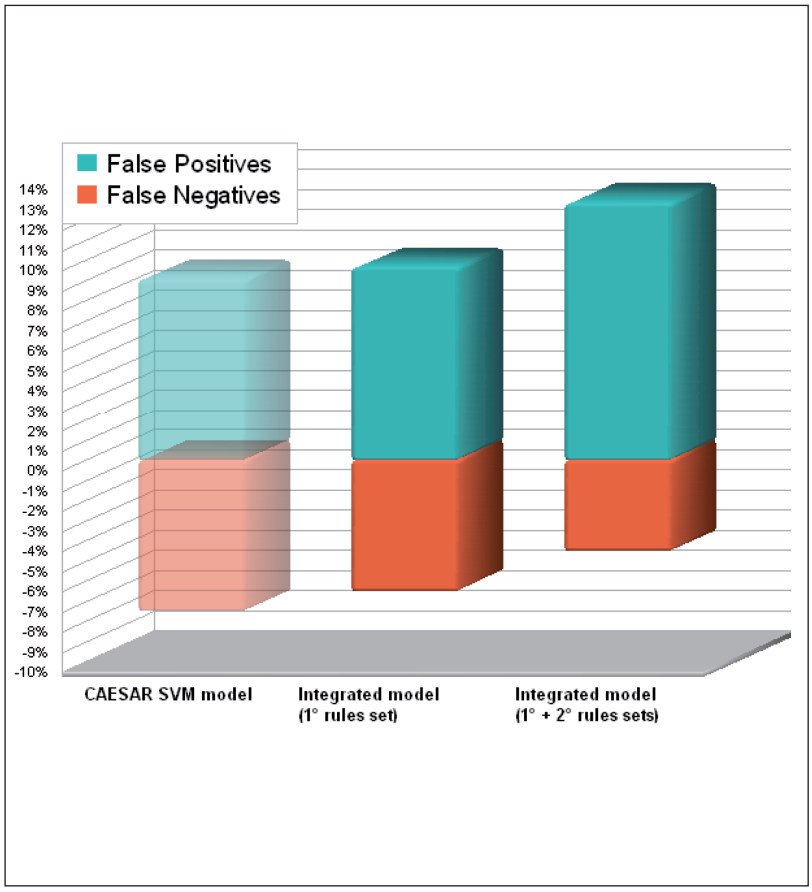


Fig. 1: Reduction of false negatives

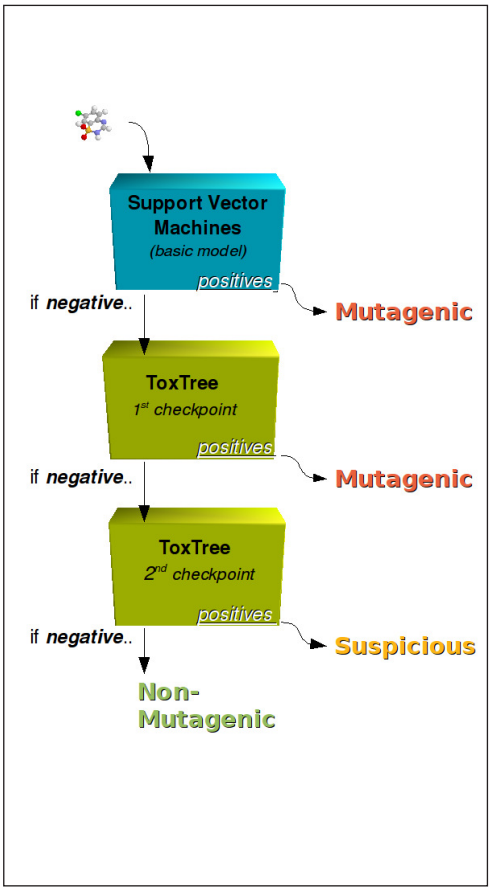


Fig. 2: The cascade model

References

Ames, B. N. (1984). The detection of environmental mutagens and potential. *Cancer 53*, 2030-2040.

Ashby, J. (1985). Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environ. Mutagen 7*, 919-921.

Benigni, R., Bossa, C., Jeliazkova, N. G. et al. (2008). The Benigni/Bossa rulebase for mutagenicity and carcinogenicity – a module of toxtree. Technical Report EUR 23241 EN, European Commission – Joint Research Centre.

Chang, C.-C. and Lin, C.-J. (2001). LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. A practical guide to support vector classification. Download from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

Kazius, J., Mcguire, R. and Bursi, R. (2005). Derivation and validation of toxicophores for mutagenicity prediction. J. Med. Chem. 48(1), 312-320.

Durant, J. L., Leland, B. A., Henry, D. R. and Nourse, J. G. (2002). Reoptimization of mdl keys for use in drug discovery. J. Chem. Inf. Comput. Sci. 42, 1273-1280.

Liao, Q., Yao, J. and Yuan, S. (2007). Prediction of mutagenic toxicity by combination of recursive partitioning and support vector machines. *Molecular Diversity 11*, 59-72.

Livingstone, D. J. (2000). The characterization of chemical structures using molecular properties: a survey. *J. Chem. Info Comput. Sci. 40(2),* 195-209.

Miller, J. A. and Miller, E. C. (1981). Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer 47*, 2327-2345.

Piegorsch, W. W. and Zeiger, E. (1991). Measuring intra-assay agreement for the ames salmonella assay. In L. Hotorn (ed.), *Statistical Methods in Toxicology, Lecture Notes in Medical Informatics*. (35-41). Springer-Verlag.

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag.

Zheng, M., Liu, Z., Xue, C. et al. (2006). Mutagenic probability estimation of chemical compounds by a novel molecular electrophilicity vector and support vector machine. *Bioinformatics 22(17),* 2099-2106.

Acknowledgements

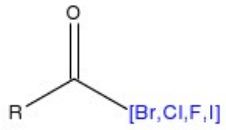
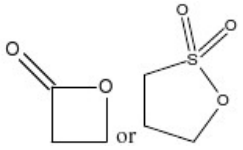
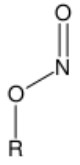
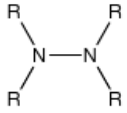
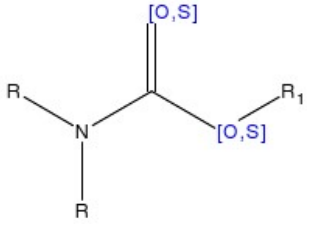
Partial support for his work has been provided through the CAESAR project of the European Union.

Correspondence to

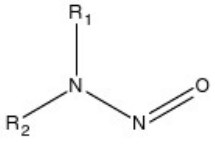
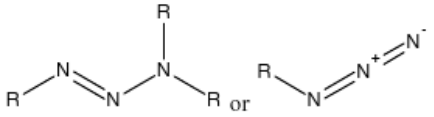
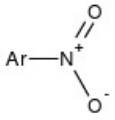
Giuseppina Gini
DEI, Politecnico di Milano
piazza L. da Vinci 32
20133 Milano, Italy
e-mail: gini@elet.polimi.it

The *In Silico* Model for Mutagenicity
Appendix A

1st Ruleset:

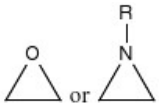
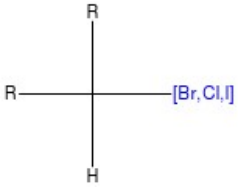
Structural Alert	SMARTS or details
SA_1: Acyl halides  R = any atom/group, except OH, SH	[!$([OH1,SH1])]C(=O)[Br,Cl,F,I]
SA_6: Propiolactones or propiosultones 	[O,S]=C1[O,S]CC1 OR O=S1(=O)(CCCO1)
SA_9: Alkyl nitrite  R= any alkyl group	O=[NX2]OC
SA_13: Hydrazine 	[N+0]!@;-[N+0](=[!O;!N]) OR [N+0]([#1,*])!@;-[N+0]([#1,*])
SA_16: alkyl carbamate and thiocarbamate  R = Aliphatic carbon or hydrogen R1 = Aliphatic carbon	[NX3]([CX4,#1])([CX4,#1])C(=[O,S])[O,S][CX4]
SA_18: Polycyclic Aromatic Hydrocarbons	Three or more fused rings, not heteroaromatic

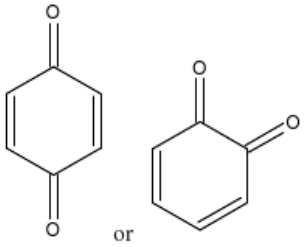
The *In Silico* Model for Mutagenicity
Appendix A

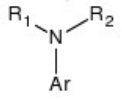
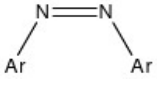
Structural Alert	SMARTS or details
SA_19: Heterocyclic Polycyclic Aromatic Hydrocarbons	Three or more fused rings, heteroaromatic
SA_21: alkyl and aryl N-nitroso groups  R1= Aliphatic or aromatic carbon, R2= Any atom/group	[C,c]N[NX2;v3]=O
SA_22: azide and triazene groups  R= Any atom/group	[N]=[N]-[N] OR [N]=[N]=[N]
SA_27: Nitro-aromatic  Ar = Any aromatic/heteroaromatic ring	• Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded. • Chemicals with a sulfonic acid group (-SO3H) on the same ring of the nitro group are excluded .

The *In Silico* Model for Mutagenicity
Appendix B

2nd Ruleset:

Structural Alert	SMARTS or details
SA_7:Epoxides and aziridines  R = any atom/group	`C1[O,N]C1`
SA_8: Aliphatic halogens  R = any atom/group	

SA_12: Quinones
 
 	
R = any atom/group
	`O=[#6]1[#6]=,:[#6][#6](=O)[#6]=,:[#6]1`
	OR
	`O=[#6]1[#6]=,:[#6][#6]=,:[#6][#6]1(=O)`

SA_28bis: Aromatic mono- and dialkylamine  Ar = Any aromatic/heteroaromatic ring R1 = Hydrogen, methyl, ethyl R2 = Methyl, ethyl	• Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded. • Chemicals with a sulfonic acid group (- SO3H) on the same ring of the amino group are excluded .
SA_29: Aromatic diazo  Ar = Any aromatic/heteroaromatic ring	• Chemicals with a sulfonic acid group (-SO3H) on both rings linked to the diazo group are excluded.

Formation of Mechanistic Categories and Local Models to Facilitate the Prediction of Toxicity

Mark T. D. Cronin, Steven J. Enoch, Mark Hewitt and Judith C. Madden

School of Pharmacy and Chemistry, Liverpool John Moores University, Liverpool, UK

Summary

There is a range of in silico techniques that can be applied to predict the toxicity of chemicals. This paper discusses the use of methods to create "local" models, particularly based around category formation and read-across, to predict toxicity. Specifically, this is illustrated with regard to categories for predicting skin sensitisation and teratogenicity. These were formed using mechanistic and structural similarity techniques to group chemicals. Local QSAR models based on grouping chemicals have the advantage that they are transparent, simple and mechanistically derived. In addition, there are a number of freely available software tools to assist in their derivation. The disadvantages include that they are labour-intensive to develop and restricted to local areas of chemistry.

Keywords: in silico, QSAR, toxicity, chemical category, read-across

1 Introduction

The toxicity and fate of a chemical can be predicted by a number of *in silico* methods. These include the use of models, such as quantitative structure-activity relationships (QSARs) built on a large number of compounds to smaller, more discreet models developed on a rational basis, such as a grouping of similar chemicals (van Leeuwen et al., 2009). There are no strict definitions or cut-offs, and this range of models can be thought of as a spectrum from global to local. In this sense (and in the context of this paper) global models can be thought of as being developed from data for large numbers of compounds (i.e. hundreds rather than tens), crossing broad structural classes and often mechanisms and modes of action. Local models are more likely to be built on smaller numbers of compounds, often with some element of structural and/or mechanistic similarity to them.

There are different reasons, advantages and disadvantages for developing and using global and local models. Global QSARs are, by their nature, generalist in that they cover broad chemical space. They may include a number of descriptors, some of them with no direct physico-chemical significance. In addition, they may be formed with non-linear techniques such as neural networks. Local models are developed to restrict the domain of the model through careful selection of compounds (see below). They will include fewer data and may simply involve read-across or a simple linear technique such as regression analysis.

The distinction between local and global models should not be thought of as a recommendation that one or the other be used. There may be strong arguments in favour of using either or both types of model, depending on the chemical in question, endpoint and context. Global models have the advantage that they are applicable for large numbers of compounds across mechanisms of action and structure. There are many global models for toxicity that can be accessed "off the shelf" either commercially (e.g. expert systems such as TOPKAT and M-CASE) or that are freely available (e.g. the CAESAR models available from www.caesar-project.eu). Local models have the advantage of often being more accurate as they are restricted in domain. They may be more transparent and simpler, providing the user with greater confidence in their application. They do, however, normally require manual construction. A discussion of global versus local QSAR models is given in more detail by Enoch et al. (2008a).

Much has been written about the use of global QSARs to predict toxicity (Bassan and Worth, 2008), and much guidance is available from the European Chemicals Agency. The aim of this study, therefore, was to assess methods to form local QSAR models by the grouping of compounds into categories and to provide illustrations of their strengths and weaknesses.

2 Methods to form local models through the development of chemical categories

A rational method to group compounds together can result in a "category" being formed. If a category is formed and can be populated with data, then read-across may be attempted (in a quantitative or semi-qualitative) sense. Quantitative models may be built within a category using either quantitative read-across (Enoch et al., 2008b) or through the development of local QSARs. These various types of read-across and QSARs built on chemical groupings or category can be thought of as a significant source of local models. They have become more important

as predictive toxicology takes on the challenges of issues such as REACH, especially where data may be sparse, and for the more complex toxicological endpoints such as chronic human health effects (Sakuratani et al., 2008; Johannsen et al., 2008). In addition, as the freely available tools described in this paper (i.e. mechanistic SMARTS strings (Enoch et al., 2008c)); OECD QSAR Application Toolbox; ToxMatch etc.) become more frequently applied, there will be greater emphasis on understanding and predicting these complex effects.

The theory of grouping compounds together is simple, namely that similar compounds will have similar properties and activities (Enoch, 2009). The group of similar compounds is termed a "category" and may also be referred to as a group of analogues. Once formed, if a category can be populated with activity values, e.g. toxicity data, knowledge of activity within a category provides a method for interpolating effects – which is often termed "read-across". The utility and increased acceptance of these methods to group chemicals together is becoming more widespread (Schaafsma et al., 2009; van Leeuwen et al., 2009).

There are a number of methods to group "similar" compounds together to form categories (Enoch, 2009), and guidance is provided into these areas by the OECD and European Chemicals Agency. The main areas on which to form categories are:

- On the basis of structural analogues and/or congeneric series: It is assumed that compounds sharing the same functional group(s) and varying only in chemical sub-groups, such as alkyl chain length, will have similar mechanisms of action and hence read-across can be performed. There are an increasing number of examples of this approach including Fabjan et al., (2006); Sanderson et al., (2009); Veenstra et al., (2009); Walker and Printup (2008).
- Compounds may be grouped together on the basis of being "structurally similar", i.e. chemical similarity, as defined by algorithms to identify them (Pavan and Worth, 2008).
- Mechanisms of action may be used to group compounds together. Grouping can be performed on the basis of structural alerts relating to a mechanism of action. For instance, the electrophilic chemistry underpinning respiratory sensitisation has been defined and allows for the basis of category formation (Enoch et al., 2009b).
- Compounds can be grouped together with similar receptor mediated modes of toxic action. This is the most complex grouping methodology, as it may require capturing 3-D structural information of molecules. An interesting recent example of how this may be achieved is given by Aladjov et al. (2009).

Details of two of these approaches to form categories, as performed in the CAESAR European Union FP6 Project, are described below.

3 Formation of categories on the basis of mechanisms of action for skin sensitisation

Skin sensitisation (or allergic contact dermatitis) is the clinical disease caused by the exposure of the skin to substances that are able to promote an immunological response (such substances are also known as contact allergens). Skin sensitisation is a complex immunotoxicological response, which is often simplified into a number of fundamental steps. The initial exposure to the substance, i.e. induction of contact allergy or sensitisation, is required to induce the immunological procedure. Subsequent challenge of the sensitised individual may result in the elicitation of a response (allergic contact dermatitis or positive test reaction). The dose sufficient for induction is generally larger than the dose sufficient for elicitation (Basketter, 2008).

With regard to making predictions of whether a compound is a sensitiser from a mechanistic standpoint, one must attempt to rationalise the processes that underpin the sensitisation process. Jowsey et al. (2006) rationalised the skin sensitisation process in terms of a number of important processes, namely bioavailability (i.e. skin permeation), protein reactivity, dendritic cell maturation and T-cell proliferation. In terms of making predictions of skin sensitisation, one must hypothesise what is the rate limiting process, without which skin sensitisation will not occur. Of the processes skin permeation, the ability to bind to a relevant protein, the ability to generate a danger signal and the ability of its antigen to be recognised, it can be considered that binding to the relevant protein is a key process (Roberts and Aptula, 2008). Whilst this hypothesis is amenable to description by chemistry and hence convenient for computational approaches, it is as yet unproven, hence the development of a suite of complementary *in vitro* approaches (Natsch and Emter, 2008; Natsch et al., 2009).

The binding of a skin sensitiser to the relevant immunoprotein forms a covalent bond between the molecule and protein. The formation of these bonds is usually through nucleophile-electrophile interactions. Cysteine (-SH) and lysine (-NH2) groups on proteins act as nucleophilic centres, amenable to electrophilic attack by the skin sensitisers (Roberts et al., 2008). The chemistry associated with skin sensitisation can therefore be rationalised, i.e. skin sensitisation is known to be related to a number of organic chemistry mechanisms of action (Schultz et al., 2006). Aptula et al., (2005) describe the possibility of six mechanisms of action (S_N1, S_N2, S_NAr, Michael addition, Schiff base formation, acylation) being associated with sensitisation. This information can be rationalised further to provide an organic chemistry mechanistic basis separating different immunological effects (e.g. skin and respiratory information, Enoch et al., (2009b)).

Since the chemistry underlying toxicological responses such as skin sensitisation can be rationalised, the types of molecules and structural features associated with the chemistry can be defined. As an example, the structural features associated with the Michael acceptor domain have been defined (Schultz et al., 2007, 2009). This information can be supplemented by tests involving chemical reactivity (*in chemico*) measurements to assist in the definition of the exact domains (Natsch et al., 2009). If such testing is performed within an "intelligent testing strategy" then the types of structures and effect of substituents, patterns of substitution and steric hindrance on reactivity can be assessed. All such chemical information can be captured computationally through the use of very simple and freely available techniques for describing chemical information. For example, Enoch et al.

(2008c) have described the chemical fragments associated with protein binding using Smiles ARbitary Target Specification (SMARTS) patterns formed into strings.

We therefore have the mechanistic basis and computational techniques to describe and define compounds that may be associated with protein binding. These have been developed into tools to assist the user to form categories (or groupings) of molecules. The SMARTS strings from Enoch et al (2008c) are available on request from the author. ToxTree (Pavan and Worth, 2008) contains rules to identify compounds as Michael-type acceptors (as defined by Schultz et al., 2007). In addition, and more significantly, as a usable tool since it is linked to databases, the OECD (Q)SAR Application Toolbox contains a profiler for protein binding. All these and other technologies, which are freely available, as well as others, provide the means for the user firstly to profile a chemical to assess whether or not it belongs to one of these mechanisms and secondly to group chemicals together, so that activity may be rationally interpolated within the group – the so-called process of read-across (Koleva et al., 2008).

Once a category has been formed on a mechanistic basis, it can be populated. From this, local models or QSARs may be produced. This can either be in the form of a QSAR model or can apply quantitative read-across. For example, Patlewicz et al. (2003, 2004) brought together a group of compounds that are likely to act as Schiff's bases (e.g., aliphatic and aryl aldehydes). In this case, a two-parameter QSAR was developed incorporating hydrophobicity and electrophilicity descriptors (log P and Taft σ* substituent constant respectively) to predict potency in the local lymph node assay.

In addition, Enoch et al. (2008b) have demonstrated the applicability of quantitative read-across to predict the potency of skin sensitisers. This is a process whereby once a category has been formed, the activity can be related to an appropriate descriptor, in the case of Enoch et al. (2008b) a descriptor of electrophilicity. If compounds are ordered according to that descriptor, then an interpolation can be made by considering the compound with the immediate higher and lower descriptor value. This very simple approach to forming local models was shown to be very powerful.

4 Formation of categories on the basis of chemical similarity for teratogenicity

The use of a mechanistic profiler or rules assumes the user has some knowledge of chemical structure. A number of other methods can be applied to form chemical groupings or categories. The use of structural similarity may provide insights into groupings or categories without recourse to mechanisms of action. At first sight this may appear to be at odds with the concept of mechanistic transparency to form groupings. However, the assumption is that compounds with a "similar" structure will have similar mechanistic properties, even if those properties are not known (Fabjan et al., 2006; van Leeuwen et al., 2009).

A number of algorithms are available to determine the relative similarity of one chemical to another. These "similarity indices" can be calculated by a number of methods and techniques. Generally they provide a number (usually on a scale of 0 to 1 – where 1 indicates identical molecules) relating relative similarity. Thus, for a query molecule, a category can be formed around it by selecting the most similar molecules from a database. For an excellent review of this area, the reader is referred to Nikolova and Jaworska (2003).

An example of category formation using structural similarity is provided by Enoch et al. (2009a). They analysed results from the FDA/TERIS teratogenicity database (Arena et al., 2004) for 290 chemicals (mainly pharmaceuticals). The dataset had been split into a "training set" (from which categories were sought) and a test set for which read-across predictions were made. Teratogenic classifications, made according to FDA guidelines (Briggs et al., 2002), were available for the "training set".

The study indicated that structural similarity can be used to develop categories on which to base read-across predictions. These categories were transparent and usually associated with a mechanistic basis. However, it should be noted that categories could not be developed for all molecules. Whilst this may at first sight seem like a limitation of this method, it should actually be thought of as a strength, as predictions cannot be made (erroneously) for compounds which are not representative of the data set.

5 Strengths and limitations of chemical categories and local QSAR and/or read-across approaches

There is no doubt that chemical categories will be formed with increasing regularity to make assessments of toxicity. From that chemical category either qualitative or quantitative read-across may be applied. In certain circumstances, it may also be possible to develop local QSARs. There are a number of advantages to the use of local models and QSARs.

- As noted in this paper, there is an increasing availability of tools to develop local QSARs and form categories. Many of these tools are free, e.g. OECD (Q)SAR Application Toolbox; Toxmatch; Toxtree; Analog Identification Method (from the US EPA) etc. The freely available tools can be supported and supplemented by commercial products.
- There is increasing acceptance across industry and regulatory agencies that *in silico* methods play an important role in providing toxicological information.
- Category approaches and local QSARs are by their nature transparent, i.e. the "algorithm" or grouping strategy is clear and obvious. Local QSARs are usually developed from a small number of chemicals with a small number of descriptors (three or fewer) using regression analysis.
- Such methods are mechanistically interpretable, i.e. the mechanism and/or mode of action (if known) can be attributed to the model or grouping and will increase confidence in the prediction.
- These grouping methods are easy to develop and describe, although they are not automated.
- Many of these factors make categories and local QSARs

easy to characterise and evaluate under the OECD Principles for the Validation of (Q)SARs.

There are also a number of disadvantages:

- Local QSARs and categories will need to be created on a case-by-case basis. Whilst tools and software are available to develop them, they will require expert input for their development.
- They are limited by the availability of toxicity data to populate the category or chemical grouping.
- Categories may be limited by the tools available to develop them. The profilers within the OECD (Q)SAR Application Toolbox are, in many cases, at an initial stage of development.
- For many endpoints and chemicals, the mechanisms of action may not be known, thus restricting confidence in the category formed.
- There is an assumption that a positive prediction from an *in silico* approach will carry more weight and be more "acceptable" than a negative prediction. It will take a long time for the scientific community to have confidence that a chemical is not associated with a hazard.
- There is a lack of guidance and case studies to assist the (novice) user (although the educational material associated with the OECD (Q)SAR Application Toolbox is to be commended).
- At the time of preparation of this paper, it is not yet known if and how the predictions from categories and/or groupings and local QSAR will be accepted by regulatory agencies.

6 Conclusions

There are a number of approaches to form categories of compounds to allow for the creation of local models and/or QSARs for the prediction of toxicity. This study illustrates the use of mechanistic information based on protein reactivity and chemical similarity indices to develop usable categories for read-across. It is shown that local models are transparent and may provide more accurate results than global models. In addition they have the advantage of being rationalised on a mechanistic basis. There are disadvantages to their use, not least that they are labour-intensive to create and are restricted to specific areas of chemistry.

References

Aladjov, H., Todorov, M., Schmieder, P. et al. (2009). Strategic selection of chemicals for testing. Part I. Functionalities and performance of basic selection methods. *SAR QSAR Environ. Res. 20*, 159-183.

Aptula, A. O., Patlewicz, G. and Roberts, D. W. (2005). Skin sensitization: Reaction mechanistic applicability domains for structure-activity relationships. *Chem. Res. Toxicol. 18*, 1420-1426.

Arena, V. C., Sussman, N. B., Mazumdar, S. et al. (2004). The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: Comparative analysis of logistic regression and decision tree models. *SAR QSAR Environ. Res. 15*, 1-18.

Basketter, D. A. (2008). Skin sensitisation: strategies for the assessment and management of risk. *Brit. J. Dermat. 159*, 267-273.

Bassan, A. and Worth, A. P. (2008). The integrated use of models for the properties and effects of chemicals by means of a structured workflow. *QSAR Comb. Sci. 27*, 6-20.

Briggs, G. C., Freeman, R. K. and Yaffe, S. J. (2002). *Drugs in Pregnancy and Lactation*. 6th ed., Philadelphia: Lippincott Williams and Wilkins.

Enoch S. J. (2010). Chemical category formation and read-across for the prediction of toxicity. In T. Puzyn, J. Leszczynski and M. T. D. Cronin (eds.), *Recent Advances in QSAR Studies: Methods and Applications* (pp 209-219). Dorcht Netherlands: Springer.

Enoch, S. J., Cronin, M. T. D., Madden, J. C. and Hewitt, M. (2009a). Formation of structural categories to allow for read-across for teratogenicity. *QSAR Comb. Sci. 28*, 696-708.

Enoch, S. J., Roberts, D. W. and Cronin, M. T. D. (2009b). Electrophilic reaction chemistry of low molecular weight respiratory sensitizers. *Chem. Res.Toxicol. 22*, 1447-1453.

Enoch, S. J., Cronin, M. T. D., Schultz, T. W. and Madden, J. C. (2008a). An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere 71*, 1225-1232.

Enoch, S. J., Cronin, M. T. D., Schultz, T. W. and Madden, J. C. (2008b). Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem. Res. Toxicol. 21*, 513-520.

Enoch, S. J., Madden, J. C. and Cronin, M. T. D. (2008c). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ. Res. 19*, 555 578

Fabjan, E., Hulzebos, E., Mennes, W. and Piersma, A. H. (2006). A category approach for reproductive effects of phthalates. *Crit. Rev. Toxicol. 36*, 695-726.

Johannsen, F. R., Vogt, B., Waite, M. and Deskin, R. (2008). Mutagenicity assessment of acrylate and methacrylate compounds and implications for regulatory toxicology requirements. *Reg. Toxicol. Pharmacol. 50*, 322 335.

Jowsey, I., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). A future approach to measuring relative skin sensitising potency. *J. Appl. Toxicol. 26*, 341-350.

Koleva, Y. K., Madden, J. C. and Cronin M. T. D. (2008). Formation of categories from structure-activity relationships to allow read-across for risk assessment: toxicity of alpha,beta-unsaturated carbonyl compounds. *Chem. Res. Toxicol. 21*, 2300-2312.

Natsch, A., Emter, R. and Ellis, G. (2009). Filling the concept with data: Integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicol. Sci. 107*, 106-121.

Natsch, A. and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to

the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci. 102*, 110-119.

Nikolova, N. and Jaworska, J. (2003). Approaches to measure chemical similarity – a review. *QSAR Comb. Sci. 22*, 1006-1026.

Patlewicz, G. Y., Basketter, D. A., Pease, C. K. S. et al. (2004). Further evaluation of quantitative structure-activity relationship models for the prediction of the skin sensitization potency of selected fragrance allergens. *Contact Dermatitis 50*, 91-97.

Patlewicz, G., Roberts, D. W., Walker, J. D. (2003). QSARs for the skin sensitization potential of aldehydes and related compounds. *QSAR Comb. Sci. 22*, 196-203.

Pavan, M. and Worth, A. P. (2008). Publicly-accessible QSAR software tools developed by the Joint Research Centre. *SAR QSAR Environ. Res. 19*, 785-799.

Roberts, D. W. and Aptula, A. O. (2008). Determinants of skin sensitisation potential. *J. Appl. Toxicol. 28*, 377-387.

Roberts, D. W., Aptula, A. O., Patlewicz, G. and Pease, C. (2008). Chemical reactivity indices and mechanism-based read-across for non-animal based assessment of skin sensitisation potential. *J. Appl. Toxicol. 28*, 443-454.

Sakuratani, Y., Sato, S., Nishikawa, S. et al. (2008). Category analysis of the substituted anilines studied in a 28-day repeat-dose toxicity test conducted on rats: Correlation between toxicity and chemical structure. *SAR QSAR Environ. Res. 19*, 681-696.

Sanderson, H., Belanger, S. E., Fisk, P. R. et al. (2009). An overview of hazard and risk assessment of the OECD high production volume chemical category-Long chain alcohols [C-6-C-22] (LCOH). *Ecotox. Environ. Saf. 72*, 973-979.

Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P. et al. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. *Reg. Toxicol. Pharmacol. 53*, 70-80.

Schultz, T. W., Rogers, K. and Aptula, A. O. (2009). Read-across to rank skin sensitization potential: subcategories for the Michael acceptor domain. *Contact Dermatitis 60*, 21-31.

Schultz, T. W., Yarbrough, J. W., Hunter, R. S. and Aptula, A. O. (2007). Verification of the structural alerts for Michael acceptors. *Chem. Res. Toxicol. 20*, 1359-1363.

Schultz, T. W., Carlson, R. E., Cronin, M. T. D. et al. (2006) A conceptual framework for predicting the toxicity of reactive chemicals: modeling soft electrophilicity. *SAR QSAR Environ. Res. 17*, 413-428.

van Leeuwen, K., Schultz, T. W., Henry, T. et al. (2009). Using chemical categories to fill data gaps in hazard assessment. *SAR QSAR Environ. Res. 20*, 207-220.

Veenstra, G., Webb, C., Sanderson, H. et al. (2009). Human health risk assessment of long chain alcohols. *Ecotox. Environ. Saf. 72*, 1016-1030.

Walker, J. D. and Printup, H. (2008). Structure-activity relationships for aldehyde categories. *QSAR Comb. Sci. 27*, 475-482.

Acknowledgements

The funding of the European Union 6th Framework CAESAR Specific Targeted Project (SSPI-022674-CAESAR) and OSIRIS Integrated Project (GOCE-037017-OSIRIS) is gratefully acknowledged.

Correspondence to

Mark Cronin, Ph.D.
School of Pharmacy and Chemistry
Liverpool John Moores University
Byrom Street
Liverpool
L3 3AF, UK
e-mail: m.t.cronin@ljmu.ac.uk

Lunch Sessions

SL1: Databases: progress report

The AnimALT-ZEBET Database: A Unique Resource for Comprehensive and Value-Added Information on 3R Alternatives

Daniel Butzke, Antje Doerendahl, Susanne Skolik, Andreas Luch and Barbara Grune

Federal Institute for Risk Assessment (BfR), National German Centre for Documentation and Evaluation of Alternatives to Animal Experiments (ZEBET), Berlin, Germany

Summary

Evaluation of value-added databases represents the most reasonable starting point for any structured search for information on suitable alternative methods. This source provides clear, comprehensive and reliable reviews on the most advanced procedures with relevance to the 3R concept.

At the forefront of these essential resources is the AnimAlt-ZEBET database offered by the German Federal Institute for Risk Assessment (BfR) and accessible online free of charge. The documents of this database compiled by scientific experts provide selected and condensed high-quality information in compliance with specific requirements of scientists, competent authorities and others who are obliged to consider the applicability of specific alternative methods. Thus, the focus of AnimAlt-ZEBET is on (1) essential technical key points, (2) application domains, (3) advances/limitations of the most elaborate protocols, (4) prediction models, (5) opinion(s) of expert panels (e.g. ESAC; ICCVAM), (6) status of validation and acceptance and, most notably, (7) contribution of the respective methods to 3R.

Because the documents of the database are written in a structured manner, they can be used as the feedstock for any up-to-date text mining application, like "semantic landscape"-producing tools. The database currently holds some 140 documents focusing on safety testing of chemicals and drugs, but soon will be expanded to basic sciences as well.

Keywords: text mining, information retrieval, structured search, value-added database, ZEBET

1 Introduction

The amount of accessible information grows rapidly in a largely uncontrolled manner. Buried within this ever-growing "thicket" are innovative ideas and seminal research that may become obscured over time. Scientific meta-databases like *PubMed* provide the possibility to search almost 20 million documents simultaneously via simple and usually general keywords. After retrieving the hit list, great efforts are thus necessary to sort out irrelevant literature, to elaborate an effective combination of search terms (search strategy) for subsequent queries and, at the same time take care of not losing track. As a result, the approach of targeted searching for information is becoming commonly unpopular among scientists. Most never attended a special course on retrieval techniques during their education and are unfamiliar with straightforward concepts (and computer aided tools) for deducing powerful keywords.

Regardless of the issues mentioned, scientists in Germany are obliged to undertake a valid indispensability search prior to applying for authorisation of an animal experiment. The aim of an indispensability search is to exclude all possibility of the presence of (i) a suitable alternative method that can be applied instead of the animal experiment, (ii) usable results from comparable previous animal experiments, and (iii) results from other research suited to anticipate the outcome of the planned experiment. This obligatory search has to consider the current state of scientific knowledge exhaustively. Only when the availability of suitable alternatives or usable scientific results has been excluded in a valid search procedure based upon the current state of knowledge, an animal experi-

ment may be approved as indispensable to reach a vindicatory scientific objective.

To assist scientists and authorising bodies, *ZEBET* at *BfR* – i.e. the German Centre for Documentation and Evaluation of Alternatives to Animal Experiments – provides free online access to its value-added database *AnimAlt-ZEBET*. This source currently contains some 140 method summaries that are written in a structured and standardised manner and concentrate on the most essential facts of the portrayed alternative methods. The added value of these documents is constituted through expert selection of incorporated literature committed to most substantial and reliable information only. Furthermore, the summaries aim to enable users to consider the suitability of the method at hand for their own purposes without needing to collect additional information. Therefore, *AnimAlt-ZEBET* serves as a convenient starting point, especially for scientists and authorising bodies, to perform a structured search for information on suitable alternatives to a particular animal experiment.

2 Effective searching with a structured approach

The basic consideration of a structured search approach is an *a priori* identification of those resources or resource-subsections that most likely contain the required information. Speaking metaphorically: When you return after a sunny day to your car and realise that you have lost the keys, then, of course, before returning to the beach and starting to dig, you may first try to consider those sites most promising. Only a person with too much time and a high frustration threshold would search the whole beach, grain for grain. But exactly this is what scientists attempt when they explore accessible meta-databases by means of simple keyword searches. Elaborating a search strategy (combination of search terms) is essentially the same as adjusting the mesh size of a sieve: If it is set too small you will retrieve the keys but also coins, shells, bottle caps, etc. If it is set too large you will select the bigger flotsam but might miss your keys. Even with a highly adjusted sieve, however, searching the entire beach will prove tedious. (It is a perfect way to find all lost car keys present at that site, though.)

With regard to the employment of search terms, the above certainly seems true for a structured search as well. Due to the restricted amount of information, however, the list of relevant documents retrieved is essentially less voluminous, more reliable and may be evaluated much faster. Thus the fine adjustment of the "mesh size" (search strategy) is easily achievable and, in a topic-restricted environment, a much easier search strategy can be composed due to the redundancy of certain keywords (i.e. topic-defining terms like "3R").

The structured search for scientific literature on a certain candidate alternative method should be inititated from sources that are likely to deliver the most relevant results while requiring the lowest efforts, that is value-added databases such as *Anim-Alt-ZEBET* and *ECVAM DB-ALM*. If no relevant literature can be detected here, the search may then proceed in adequate topic-restrictable databases like *PubMed* or *AGRICOLA*, where the pool of explorable literature can be preselected via subject headings like "animal use alternatives" (currently roughly 2000 documents in PubMed and 1100 in AGRICOLA, respectively). The searcher can find further assistance at the convenient *ALTBIB*-Portal to *PubMed*.

3 AnimAlt-ZEBET at a glance

The fundamental concept behind the documents in the database is to provide selected high-quality information in compliance

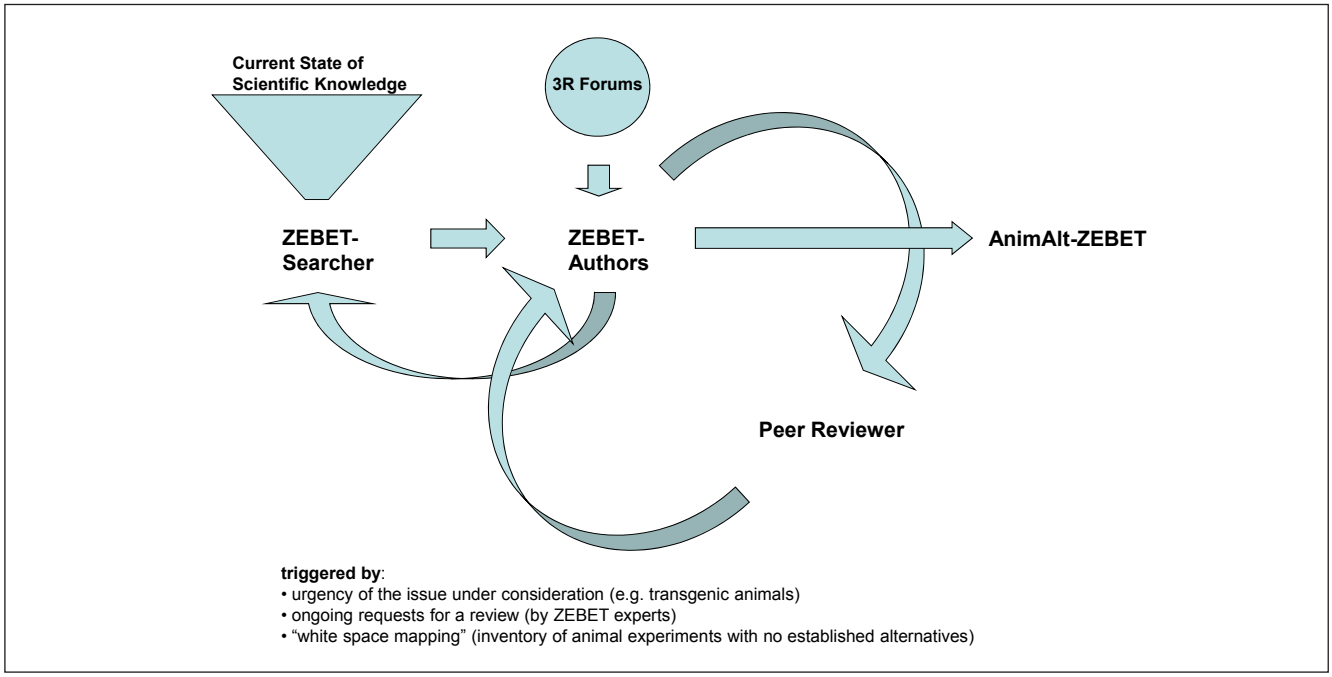


Fig. 1: The workflow of method-summary production for AnimAlt-ZEBET. The process is triggered by the indicated stimuli.

with specific requirements of scientists, competent authorities (authorising bodies) and others who are obliged to consider the applicability/suitability of a certain candidate alternative method. The information provided is therefore focused on, (1) the essential technical key points and required equipment, (2) the application domain, (3) advances/limitations of the most elaborate protocol and its variations, (4) the prediction model, (5) the opinion(s) of expert panels (e.g. *ESAC*; *ICCVAM*), (6) the status of validation and acceptance and, most notably, (7) the contribution of the respective method to the *3R* concept.

The workflow of method-summary production integrates several quality ensuring feedback loops (Fig. 1). Firstly, from the pool of potentially relevant documents collected at *ZEBET*, only the most adequate literature is selected for integration based upon defined criteria. Secondly, the draft summaries are subjected to a stringent internal review process performed by highly recognised experts in the field.

The decision on which method-summaries should be provided next is triggered by (i) the urgency of the addressed issue (e.g. dramatic increase in the number of transgenic animals), (ii) whether there are ongoing requests for a review on animal use applications by national authorising bodies, and (iii) the general desire for "white space mapping" within an inventory of required animal experiments that visualises areas of as yet absent established alternatives. To keep up-to-date with the latest activities and developments in the field of *3R*, the authors also consider relevant forums like *AltTox.org* and *3R* specific portals like *AltWeb*. Nevertheless, care is taken to extract only information relevant for the particular audience.

***3R* Relevance:** *the essential facts*

Background

- **Scope:** *in what context is the experiment required*
- **Animal experiment:** *description of the substituted animal experiment*
- **Alternative approaches:** *comparison of existing alternatives*

Method

- **Principle:** *how does it work*
- **Procedure:** *the essentials of the most elaborate (recommended) protocol and information on the required equipment*
- **Prediction Model:** *method output*

Expert Panel Opinions

wording of ESAC statements, ICCVAM evaluation reports, etc. in extracts

Note

application domain, limitations, performance standards, variations of the basic protocol, etc.

Fig. 2: The conceptual frame of the method-abstract.

Structure and phrasing of the method-summaries are standardised to enable immediate orientation and easy comprehension and to provide the possible feedstock for up-to-date text mining applications. To get a general idea of the method at hand, the reader may simply consult the meta-data fields "Title", "Evaluation", "Status" and "Regulation", in combination with the abstract section "*3R* relevance". Here, users will find highlighted facts to consider the specific objective, the state of development, the acceptance in the scientific/regulatory community, the application domain, and the contribution to the *3R* concept.

Starting with method-summaries that were generated in 2009, there will be a revised structure of the main text (Fig. 2). As explained above, this main text aims to cover all aspects that are necessary to consider the suitability/applicability of a particular method in a given context. Thus, there is a focus on practical considerations, e.g. endpoints, operating schedule, required equipment, limitations, etc. The section "Expert Panel Opinions" provides a detailed picture of the acceptance status of the particular approach and possible objections that might have been raised by official bodies.

Currently, the *AnimAlt-ZEBET* database holds documents that focus on safety testing of chemicals and drugs, but soon it will be expanded to also cover all relevant areas in basic sciences.

4 New online presence in october 2009

AnimAlt-ZEBET is hosted by *DIMDI*, the *German Institute of Medical Documentation and Information*. At this resource one can search the database exclusively or – in a "superbase mode" – blended with other databases like *PubMed*. It is planned to run a stand-alone online presence of *AnimAlt-ZEBET* that will be launched at the BfR website in the future.

References

(Resources as available on 21.08.2009)
AnimAlt-ZEBET, http://www.dimdi.de/static/en/db/dbinfo/zt00.htm
AGRICOLA, http://agricola.nal.usda.gov/
ALTBIB, http://toxnet.nlm.nih.gov/altbib.html
AltTox.org, http://www.alttox.org/
AltWeb, http://altweb.jhsph.edu
ECVAM DB-ALM, http://ecvam-dbalm.jrc.ec.europa.eu/
PubMed, http://www.ncbi.nlm.nih.gov/pubmed/

Correspondence to

Dr. Barbara Grune
Head of Unit ZEBET-Database and Information Services
National German Centre for Documentation and
Evaluation of Alternatives to Animal Experiments
Federal Institute for Risk Assessment
Diedersdorfer Weg 1
12277 Berlin
Germany
e-mail: Barbara.Grune@bfr.bund.de

AltTox.org: Connecting Stakeholders on Issues Concerning Non-Animal Methods of Toxicity Testing

Martin L. Stephens[1] and Sherry L. Ward[2]

[1]Humane Society of the United States, Washington, USA; [2]Biotred Solutions, New Market, USA

Summary

AltTox.org is an interactive website devoted exclusively to non-animal methods of toxicity testing. The emerging paradigm shift in toxicity testing away from adverse outcomes in animals and towards a mechanism-based approach, largely in vitro and in silico, was recently endorsed by the US Environmental Protection Agency and appears to have broad support among many stakeholders. This shift will necessitate communication and coordinated efforts by stakeholders in government, industry, academia, and NGOs. AltTox is designed to provide a common platform for these diverse stakeholders to exchange information and perspectives on relevant science and policy issues. The website consists of two interconnected components: a series of message boards, or forums, and an informational resource center. The forums serve as a communication tool for stakeholders to rapidly exchange information and perspectives. The resource center features concisely summarized and comprehensive information on toxicity testing and alternative test methods. There are numerous ways that stakeholders can communicate with other stakeholders via AltTox and/or contribute content that will then be available to all users. AltTox stakeholders are encouraged to contribute forum postings, meeting reviews, informational resource suggestions (meetings, in vitro testing labs and products, funding sources, etc.), and more. Users can nominate experts who will be invited to submit opinion articles on The Way Forward. A survey is available for providing feedback on the website itself. We welcome comments on stakeholder interest in possible new interactive tools for the website such as a webinar portal for stakeholder use or social media links.

Keywords: in vitro, toxicity testing, forum, 3Rs, alternatives website, AltTox, dissemination

1 Introduction

AltTox.org is a website devoted exclusively to non-animal methods of toxicity testing and is designed to provide a common platform for diverse stakeholders to access and/or exchange information and perspectives on relevant science and policy issues. The target audience includes professionals in industry, government, academia, and nongovernmental organizations.

Recent signs indicate that toxicity testing is in the early stages of a paradigm shift away from its current reliance on *in vivo* methods and toward the use of non-animal methods. Some of these signs include the National Research Council's new vision and strategy for the future of toxicity testing (National Research Council, 2007) and the subsequent endorsement of this approach by the US Environmental Protection Agency (EPA, 2009); the "Tox21" collaboration among various US federal agencies (Kavlock et al., 2009); the European Union's ban on animal testing of cosmetics and their ingredients (European Commission, 2009); and the work of the European Partnership for Alternative Approaches to Animal Testing (EPAA) (EPAA, 2008).

The shift away from animal-based methods in toxicology is being driven by a combination of factors related to science, public health, and animal welfare, as well as by sheer practicalities. It represents the beginning of a new synergism between safety testing and reduced reliance on animal testing.

For these reasons, The Humane Society of the United States (HSUS) and The Procter & Gamble Company (P&G) believed that the time was right for a communication initiative that would accelerate the exchange of information and ideas among stakeholders interested in advancing non-animal methods for toxicity testing. We proposed to accomplish this communication goal through the development of *AltTox.org*, an interactive website launched in December 2007 that encourages stakeholder exchange of technical and policy information on non-animal approaches (*in vitro* methods, *in silico* methods, and integrated testing strategies that incorporate these methods) for toxicity testing.

2 Methods

Early activity in developing the website included naming it (*AltTox* reflects the intersection of alternative methods and toxicology) and developing a logo, a site map (describing the website content and features), page designs, and article content. Interactive forums (message boards) were developed using available (off-the-shelf) software from Simple Machines.

The *AltTox* Management Team consists of Dr. Martin Stephens of the HSUS and Dr. George Daston of P&G. The Webmaster is Ms. Loree Talley, and the Contributing Editor and content writer is Dr. Sherry Ward. European content is contributed by Mr. Troy Seidle. We sought to limit bias and errors by recruiting subject matter experts to oversee the website content (through service on the editorial board) and forum postings (through service as moderators).

2.1 Structure and features of the website

AltTox.org consists of two major components: the *AltTox Website*, a conventional website format that provides various types of informational content on non-animal methods of toxicity testing, and the *AltTox Forums*, which provide an interactive communication platform for stakeholders. While each of these two components can be individually accessed, the Website and the Forums are cross-linked to each other in many places.

AltTox Website (http://www.alttox.org): The *Toxicity Testing Resource Center* (TTRC) is the primary component of the website's informational content and is accessed from the homepage. The TTRC is described in greater detail below. The homepage also features the following sections:

– *About AltTox*: provides a short overview of the website.
– *Get Involved*: describes the many ways that users can access content, submit content, and participate or provide feedback. Organizations are invited to become supporters of *AltTox*.
– *In the Spotlight*: consists of one or more new articles per month that highlight news, events, scientific advancements, and information about the website. The *In the Spotlight* archive now contains 30 articles.
– *Homepage e-mail sign up box*: allows readers to subscribe to the monthly e-newsletter, *AltTox Digest*, which highlights new content and forum postings added during the previous month.
– *Meetings & Events*: provides a comprehensive list of relevant events and grant deadlines; submit information to be listed free-of-charge to *info@alttox.org*.

The TTRC is a concise source of scientific, technical, and policy information on non-animal testing methods - information not easily found elsewhere on the Web. All topics in the TTRC are cross-linked to the relevant forums and to invited expert commentaries. The types of content located in the TTRC include:

– Science-based informational articles on toxicity test methods (*Toxicity Endpoints & Tests*) and relevant technologies (*Emerging Technologies & Approaches*)
– A readily accessible list of validated non-animal toxicity test methods organized according to testing endpoints (*Existing Alternatives*)
– Overviews of policy and procedural issues, such as test method validation and international harmonization
– Overviews of relevant US and EU organizations
– An extensive and unique source of links to relevant organizations, journals, websites, training, *in vitro* testing supplies and services, databases, funding sources, and 3Rs awards (*Informational Resources*)
– Invited commentaries written by scientific and policy experts, providing the latest perspectives on strategies and approaches for moving toward the ultimate goal of a completely non-animal approach to toxicity testing (*The Way Forward*)

AltTox Forums (http://forums.alttox.org): The Forums are interactive communication platforms used to post news, information, and perspectives, as well as to encourage user feedback and commentary. They provide a unique opportunity for stakeholders in the field of toxicity testing to share knowledge and exchange ideas on the development and use of non-animal methods. Interested professionals are invited to register in the forums and either initiate a new topic or reply to an existing message. Registration is not required for reading the forum postings. A brief tutorial is provided to explain the processes of registration and posting messages, although it is not generally needed. Another useful feature is the "alert" function, which allows forum participants to sign up to receive e-mail alerts of new topics or replies to specific topics within a forum.

There are four categories of forums: *Toxicity Endpoints & Tests, Emerging Technologies, Programs & Policies, and Overarching Challenges & Opportunities*. The content scope of each forum is described in greater detail on the website at: http://www.alttox.org/forums. To date, the forums serve primarily as bulletin/message boards for alerting readers to conferences, publications, program documents, and the like, but they are also being used for their primary intended purpose as platforms for science and policy discussions and information exchange. Each forum is moderated by at least one expert in the field.

A new communication tool, the Community of Practice (COP) forums, is a recent addition. The COP forums are intended to operate as centers for discussion and information exchange among working groups, consortia, committees, panels, etc. tasked with working through a specific challenge that will advance non-animal testing methods. Because the new forums are open to the larger *AltTox* readership, they give visibility and transparency to a project, engaging stakeholders that may not otherwise be aware of ongoing work. To date, *AltTox* features two COP forums: DNT (Developmental Neurotoxicity), and *Chemical Reactivity-ChemReactTox*.

2.2 Maintaining the website

Events and other Informational Resources sections are updated continuously as new information is identified or submitted to *AltTox*. New *In the Spotlight* feature stories are posted to the Homepage once or twice monthly. The goal of uploading at least one new *The Way Forward* essay per month has been met thus far. Website content is reviewed by the Editorial Board before being uploaded and at least annually thereafter to identify outdated or missing content. Forum postings are made *ad hoc* by registered forum users. Members of the *AltTox* Team also regularly post messages. One of the roles of the Forum Moderators is to maintain active forums.

2.3 Avenues for stakeholder participation

There are various ways stakeholders can participate actively in the *AltTox* community (http://www.alttox.org/get-involved), including the following:

– Announce – request a free listing for your product, service, meeting, grant program, etc. in the *Informational Resources* section
– Browse – browse the content, including feature articles, fo-

rum postings, and events, and sign up for the e-newsletter to receive monthly updates on new content
– Feedback – fill out a brief survey to assist us in meeting the needs of *AltTox* users
– Nominate – nominate experts as potential commentary authors (*The Way Forward*) or Editorial Board members
– Promote – tell your colleagues and students about *AltTox*
– Sponsor – financially support *AltTox* or one of its activities
– Write – contribute to the website content and forum postings

3 Results

AltTox will reach its second anniversary in December 2009. The website currently receives over 4,000 visits per month, and nearly 1,500 users receive the monthly e-newsletter. Forty-eight *The Way Forward* commentaries have been published to date, the most popular having been accessed more than 1,000 times. As of August 2009, there were 219 registered forum users and 377 forum postings. We consider these numbers to be appropriate for a young website directed toward a niche audience of stakeholders interested in non-animal methods of toxicity testing.

We sought to limit bias and errors by recruiting subject matter experts to oversee the website content (through service on the editorial board) and forum postings (through service as moderators). These experts span the range of topics covered by the website content, and provide geographic coverage of the US, the EU, and Japan (Tab. 1).

The forums provide an opportunity for stakeholders to establish an online community to discuss challenges and opportunities to advance various aspects of non-animal methods for toxicity testing. Some interesting discussions and exchanges of scientific information have taken place in the forums. However, the majority of postings are still of the "message board" variety to alert users to new events, policy articles, and so on. The Community of Practice forums got off to a slow start, but participants continue to claim interest in using this discussion platform. The *AltTox* Management Team welcomes proposals to establish new COP forums.

AltTox just recently began seeking website sponsors, and at this time the site has three sponsors in addition to the two founding organizations. The support from sponsors is intended to enable us to provide additional website features and content for *AltTox* users. It also provides an opportunity for organizations to reach a targeted audience with a link to their products or services and to show their support for reducing animal use in toxicity testing.

4 Discussion

AltTox is the only website dedicated solely to non-animal methods of toxicity testing. We believe that its concise and comprehensive information on such methods is an unparalleled resource on the Internet today.

We are particularly excited by two *AltTox* features, the *AltTox Forums* and *The Way Forward* commentaries. The Forums offer a platform for interactive discussion of scientific and policy issues

Tab. 1: The *AltTox Forum* moderators and editorial board members as of August 2009, their affiliations (for identification purposes only), and their geographic locations.

Forum moderators		
Sandra Coecke	European Centre for the Validation of Alternative Methods	Italy
Claudius Griesinger	European Centre for the Validation of Alternative Methods	Italy
Ian Kimber	University of Manchester	United Kingdom
Hajime Kojima	Japanese Center for the Validation of Alternative Methods	Japan
William Mundy	Environmental Protection Agency	United States
Grace Patlewicz	Dupont	United States
Horst Spielmann	ZEBET – Centre for Documentation and Evaluation of Alternatives to Animal Experiments	Germany
Martin Stephens	The Humane Society of the United States	United States
Editorial board		
Marilyn Aardema	Procter & Gamble	United States
Ed Carney	Dow Chemical	United States
Rodger Curren	Institute for In Vitro Sciences	United States
George Daston	Procter & Gamble	United States
William Dressler	Consultant	United States
Ian Kimber	University of Manchester	United Kingdom
Albert Li	In Vitro ADME	United States
John Lipscomb	Environmental Protection Agency	United States
William Mundy	Environmental Protection Agency	United States
Grace Patlewicz	Dupont	United States
Horst Spielmann	ZEBET – Centre for Documentation and Evaluation of Alternatives to Animal Experiments	Germany
Martin Stephens	The Humane Society of the United States	United States

across geography, stakeholder sectors, and disciplines. This mode of interactive communication seeks to break down artificial barriers among interested parties in government, industry, academia, and animal protection. We believe the Forums provide a platform that will help the toxicity testing community identify challenges and opportunities and will inspire others to take action. This interdisciplinary sharing of knowledge should also encourage innovation, progress, and collaboration. We know of no similar resource serving this purpose.

The Way Forward commentaries from invited experts are intended to stir debate, clarify issues, and suggest approaches for making progress. The diversity of stakeholder interactions driven by the *AltTox Forums* and illustrated by *The Way Forward* commentaries should encourage stakeholders to move beyond their specialties and to sample developments in related fields of non-animal methods and, potentially, to facilitate new collaborative efforts and information sharing.

Developing new policy and research initiatives to support the new vision for toxicity testing in the 21st century will require communication and collaboration among diverse stakeholders. An EPAA workshop concluded that "dissemination of information on the Three Rs has a direct impact on moving the Three Rs methods from R&D to validation, acceptance, and implementation" (Robinson et al., 2008).

The Web 2.0 offers new tools and ways for approaching traditional communication barriers and information overload. The *AltTox Forums*, which may serve as an example, are designed to encourage information exchange, knowledge sharing, and networking. Analogous to other online discussion platforms, especially those directed toward scientists, a large number of stakeholders have been hesitant to participate. One way we have addressed this hesitation was by adding a disclaimer to identify posted comments as independent of the author's affiliation.

The advantage of any online community dedicated to specific goals is that individual members can pool their knowledge and communicate rapidly with one another. *AltTox* is seeking to leverage our participants' knowledge to advance non-animal methods of toxicity testing. That depends critically upon stakeholder participation. In a discussion of online communities for scientists, Sharpton and Jhaveri (2006) noted that "the most fundamental limitation is that the utility of these sites is a direct function of the level of community participation."

We contend that *AltTox*'s value to its stakeholders will increase as more professionals contribute to and participate in the website. The interactions between users in an online community have been described as a way to leverage knowledge and "harness the collective wisdom" of its participants (Sharpton and Jhaveri, 2006). The more participants, the broader the expertise and points of view that are covered, and the greater the information and value created for all. We urge interested parties to tell their colleagues and students about *AltTox*, as appropriate, and encourage them to participate.

AltTox's vision is to create an online community that will, through stakeholder interactions and communications, stimulate progress internationally on the development, validation, acceptance, and implementation of non-animal methods for toxicity testing. The founders and developers of *AltTox* believe that replacement methods will have an important role in the field of toxicity testing in the 21st century and that *AltTox* can facilitate and support progress in this direction.

References

Environmental Protection Agency (EPA) (2009). The U.S. Environmental Protection Agency's strategic plan for evaluating the toxicity of chemicals. Available at: http://www.epa.gov/osa/spc/toxicitytesting/index.htm

European Commission (2009). Cosmetics and animal tests. Available at: http://ec.europa.eu/enterprise/cosmetics/html/cosm_animal_test.htm

European Partnership for Alternative Approaches to Animal Testing (EPAA) (2008). New perspectives on safety. Available at: http://ec.europa.eu/enterprise/epaa/3_activities/3_3_research/new_perspectives.pdf

Kavlock, R. J., Austin, C. P. and Tice, R. R. (2009). Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Analysis 29*, 485-487.

National Research Council (2007). *Toxicity testing in the 21st century: A vision and a strategy*. Washington, DC: National Academies Press.

Robinson, S., Basketter, D., Bussi, R. et al. (2008). Dissemination strategies: how do they influence the uptake of the new Three Rs methods across laboratories and other boundaries? A report of a workshop held by the European Partnership for Alternative Approaches to Animal Testing (EPAA) Working Group 3. *ATLA 36*, 453-458.

Sharpton, T. J. and Jhaveri, A. A. (2006). Leveraging the Knowledge of Our Peers: Online Communities Hold the Promise to Enhance Scientific Research. *PLoS Biol. 4(6),* e199.

Note: Since preparation of the original manuscript, the multiple forums on AltTox.org have been consolidated into one forum.

Acknowledgments

The *AltTox* Management Team wishes to express their thanks and appreciation to all of the Forum Moderators and Editorial Board members who have kindly donated their time in support of this effort, to the experts who prepared The Way Forward essays, and to the stakeholders who have taken the time to post messages in the forums. We would also like to thank the founding organizations, HSUS and P&G, for their support, and our new corporate and organizational sponsors: the Alternatives Research and Development Foundation, the American Chemistry Council, and Lhasa Limited.

Correspondence to

Martin Stephens
The Humane Society of the United States
2100 L Street, NW
Washington, DC 20037
e-mail: mstephens@hsus.org

SL2: Good Cell Culture Practice

The Bologna Statement on Good Cell Culture Practice (GCCP) – 10 Years Later

Gerhard Gstraunthaler

Division of Physiology, Innsbruck Medical University, Austria, and zet – Centre for Alternative and Complementary Methods to Animal Testing, Linz, Austria

Summary

In the 3Rs, replacement of animal testing by in vitro methods is one of the most ambitious goals. In this respect, cell and tissue culture techniques serve as important tools. However, cultured cells in vitro often represent ambiguous models of living organisms. The uncertainty results from the biology of mammalian or human cells in culture and from their variability depending on the in vitro protocols applied and the culture systems used, respectively. Therefore, all efforts should be undertaken to best mimic the in vivo situation for cultured cells and to standardise the high number of variables that are inherent to cell and tissue culture. To this end, in analogy to Good Laboratory Practice (GLP), guidelines for Good Cell Culture Practice (GCCP) were proposed at the 3rd World Congress on Alternatives & Animal Use in the Life Sciences in Bologna, 1999. Since then, GCCP guidelines for cell and tissue culture work were elaborated and were published in two consecutive ECVAM Task Force Reports.

Keywords: cell culture, in vitro systems, quality control, cross-contamination, cell line authentication

1 Background

The cultivation of eukaryotic cells has become a powerful technique in basic cell and molecular biological research, applied biotechnology, and *in vitro* alternatives (Lindl and Gstraunthaler, 2008). Cultured cells form the biological basis of most alternative methods (Hartung, 2007).

The use of *in vitro* systems is likely to increase significantly in the future (Balls et al., 2006), not merely in basic research, but also in more formal applications, including:

- toxicity, safety and quality control testing
- diagnosis, tissue banking
- biomaterial production
- therapeutic applications (e.g. autologous cell transplantation, stem cell therapies)

There will also be significant developments in relation to:

- genetic manipulation
- "omics" technologies (genomics, proteomics, metabolomics)
- biomarkers of susceptibility, exposure and effect
- high throughput screening in pharmacology and toxicology

Despite the wide-spread use and broad applications of cell and tissue cultures a significant number of basic questions and methodological protocols are still unsolved and are handled quite differently in different tissue culture laboratories. Before cell culture could be carried out successfully, two problems had to be overcome:

1. Populations of cells had to be grown from single cells.
2. These populations had to be maintained, alive, for many generations.

In a successful propagation of cells *in vitro*, cells from various tissues should grow and proliferate under appropriate culture conditions, while preserving highly differentiated functions, which closely resemble their ancestor cells *in vivo* (Barnes et al., 1987; McKeehan et al., 1990). Thus, cell proliferation and cell differentiation are two major, albeit opposing, end points in tissue culture. Which of these contrasting goals should be achieved depends on the aim of a selected cell culture study and thus on the culture conditions applied: (i) the supplementation of the culture media with growth factors

or differentiation factors, (ii) the use of specific extracellular matrix components, (iii) the subcultivation intervals and seeding densities, (iv) the feeding cycles, and (v) the choice of stationary cultures versus dynamic media supply in perfusion reactors. In sum, a number of tissue culture parameters have to be defined and coordinated. Therefore, in *Good Laboratory Practice* in the cell culture laboratory, well defined and precisely described culture protocols are mandatory for optimal and reproducible culture conditions and for making possible any interlaboratory comparability of data and scientific results obtained with cultured cells (Gstraunthaler, 2006).

2 Need for GCCP guidelines

The maintenance of high standards is fundamental to all good scientific practice and is essential to securing the reproducibility, credibility, acceptance, and proper application of any results produced. Thus, well-defined and precisely described culture protocols are mandatory to ensure optimal and reproducible culture conditions, and to enable the interlaboratory comparability and exchange of experimental data obtained with *in vitro* culture systems, e.g. in pre-validation and validation studies of *in vitro* alternatives (Balls et al., 1995). To this end, in analogy to Good Laboratory Practice (GLP) (Cooper-Hannan et al., 1999), guidelines for Good Cell Culture Practice (GCCP) were proposed at the 3rd World Congress on Alternatives & Animal Use in the Life Sciences in Bologna, 1999. Since then, GCCP guidelines for cell and tissue culture work were elaborated and were published in two consecutive ECVAM Task Force Reports (Balls et al., 2006; Coecke et al., 2005; Hartung et al., 2002).

The basic principles of good cell culture practice and quality management in cell and tissue culture, that have been identified, are:

- Know what you are working with and what impacts your *in vitro* system.
- Control the quality of what you are working with to achieve validity and maintain reproducibility.
- Record the information required to trace materials and procedures used and to reproduce your experiments.
- Report the information required for your target audience to understand and evaluate your experiments.
- Conduct risk assessments and establish procedures to protect individuals and the environment.
- Educate the people involved to enable high quality and safe research.
- Work in an ethical, responsible and accountable manner, and in compliance with relevant laws and regulations.

The Bologna Statement on Good Cell Culture Practice (GCCP) (Hartung and Gstraunthaler, 2000; Hartung et al., 2000, 2001)

It is the scientist's mission to produce reproducible, reliable and relevant data. In the life sciences, cell culture techniques represent versatile but often ambiguous models of living organisms. The uncertainty resulting from many factors that can cause inherent artefacts, demands the highest level of standardisation, definition and control for the achievement of meaningful results. As far as *in vitro* systems are used to model the *in vivo* situation, all efforts should be undertaken to approximate cell behaviour relevant for the *in vivo* situation. Results should be original, i.e. derived directly from own independent laboratory experiments, or sources/originators of these or analogous data must be cited. Important results should be submitted for publication; obvious conflicting or fraudulent results should be communicated to the same journal.

The Statement

The participants in the 3rd World Congress on Alternatives and Animal Use in the Life Sciences (Bologna, August 29-September 2, 1999) call on the scientific community to develop guidelines defining minimum quality standards in cell and tissue culture, to be called Good Cell Culture Practice (GCCP) in analogy to the OECD Principles of Good Laboratory Practice (GLP) (Cooper-Hannan et al., 1999), which cannot normally be fully implemented in basic research on the grounds of cost and lack of flexibility. Such guidelines should facilitate the interlaboratory comparability of *in vitro* results. One of the intentions of the proponents of this statement is to encourage journals in the life sciences to adopt these guidelines as a condition for the publication of work involving cell and tissue culture.

This Statement should be reviewed and updated at future *World Congresses on Alternatives and Animal Use in the Life Sciences*.

3 Principles

1. Reports on cell culture experiments should include a basic description of the cultured cells (Fedoroff, 1967; Hayflick, 1990; Lindl und Gstraunthaler, 2008; Masters et al., 2000; Schaeffer, 1984, 1989, 1990).

- *Nomenclature* of cell type or cell line in use (code, e.g. ATCC No.).
- *Origin* and *Mode of Culture Initiation* (species, organ, tissue, lineage, mode of transformation, sublines/hybrid cells; in case of humans: donor, disease, biopsy, tumour) (Anderson et al., 1998).
- *Source:* e.g. cell bank (ATCC, ECACC, DMSZ, Riken Gene Bank, etc.), laboratory of origin, original publication/patent (Hay, 1988; Hay et al., 1996).
- *Basic Morphological Description* of cultured cells, including stability of the phenotype.
- *Differentiation State* of cells must be controlled (e.g. by morphology, histochemistry, enzyme/gene expression, growth rates, viability, sensitivity to toxins, scope to stimulate cell functions, surface markers, adherence to matrices). The appropriate measure of differentiation should be independent of the cellular function under study; it should be assessed at least at the beginning and at the end of each series of experiments. Wherever available and feasible, organotypic culture systems should be employed (e.g. sandwich cultures, air/liquid inter-

phase cultures, perfusion cultures, cultures on microporous supports).
- Measures undertaken for cell line *Identification and Authenticity*, e.g. karyotyping, DNA analysis, fingerprinting, tests for cross-contamination (Chatterjee, 2007; Dirks et al., 2005; Markovic and Markovic, 1998; Masters, 2002b; Masters et al., 2001; Nardone, 2007, 2008).
- *Risk Assessment* (risk group and biosafety level, e.g. genetic modification, special care for human and primate cultures, exclusion of infections such as HBV, HCV and HIV, parasites).

2. Culture methods should be precisely defined in standard operating procedures (Freshney, 2005; Lindl und Gstraunthaler, 2008).
- Composition of *Culture Media* for routine cultures (maintenance media) and/or experimental cultures and supplements/additives (e.g. serum, growth factors hormones, antibiotics) should be defined, including name of the supplier; medium volumes used and feeding cycles should be defined; changes of batches of material should be controlled with regard to their influence on the principal endpoints of the study (Butler and Jenkins, 1989; Ham and McKeehan, 1979; Gstraunthaler, 2003; Gstraunthaler et al., 1999).
- *Culture Vessels* (flasks, Petri dishes, bottles, roller cultures, etc.) should be defined, with the name of the manufacturer/supplier.
- *Culture Substratum* (coating material, e.g. collagen, fibronectin or laminin, coating procedure) should be defined, with the name of the manufacturer/supplier.
- *Subcultivation* intervals (cell density, confluent/subconfluent cultures, split ratio, initial passage number, number of passages in culture) should be defined.

3. Animals and materials employed must be of the highest quality.
- The quality of *equipment*, donor *animals/animal facilities*, *cells* and *cell culture* materials should be such as to guarantee reproducible and reliable results. The impact of variation in these materials should be controlled and documented. Equipment and instruments should be maintained and calibrated properly (e.g. control of temperature and CO_2 levels of incubators). All materials employed should be stored under appropriate conditions to protect them from damage, infestations or contaminations.
- Donor animals from an outside source should be kept in *quarantine* for an appropriate period. Similarly, measures should be taken when a cell line is introduced into the laboratory to assure that no infection/contamination of cell lines already present can occur (certificate from supplier, test for most common contaminations, e.g. mycoplasms, growth in antibiotic-free medium for a specified period).
- Tests for frequent *contaminations*, e.g. mycoplasms, must be performed on a regular basis. Results must be discarded in the event of any evidence of contamination of materials.
- *All waste* must be treated properly, i.e. minimising the threat to humans (e.g. toxicity, mutagenicity, teratogenicity), as well as to other cells and animals under study.

4. Data must be analysed and documented properly.
- Data should be subjected to adequate *statistical analysis* without subselection or neglection of data sets. A distinction should be made between variation among replicates within a single experiment and variation among replicate independent experiments.
- A complete documentation of raw data, analysis procedures, lists of materials and equipment employed, as well as all procedures including derivations from the protocol should be maintained and *stored* for at least ten years by the primary investigator.

After publication of the first ECVAM Task Force Report on GCCP (Hartung et al., 2002), a second Task Force was convened at ECVAM in 2003 with a broader range of expertise in cell and tissue culture in order to produce a more-detailed GCCP guidance document which could be of practical use in the *in vitro* laboratory. The report of the second ECVAM Task Force on Good Cell Culture Practice was published in 2005 (Coecke et al., 2005) and was presented at the 5th World Congress on Alternatives & Animal Use in the Life Sciences in Berlin, Germany (Balls et al., 2006; Gstraunthaler, 2006).

The aim of this GCCP guidance is to foster consensus among all concerned with the use of cell and tissue culture systems in order to:
- establish and maintain best cell and tissue culture practice
- promote effective quality control systems
- facilitate education and training
- assist journal editors and editorial boards
- assist research funding bodies and
- facilitate the interpretation and application of conclusions based on *in vitro* work

GCCP complements Good Laboratory Practice, but cannot replace it.

The GCCP guidance is based on six operational principles:

1. Establishment and maintenance of a sufficient understanding of the *in vitro* system and of the relevant factors which could affect it.
2. Assurance of the quality of all materials and methods, maintenance of the integrity, validity and reproducibility of any work conducted.
3. Documentation of the information necessary to track the materials and methods used, to permit the repetition of the work and to enable the target audience to understand and evaluate the work.
4. Establishment and maintenance of adequate measures to protect individuals and the environment from any potential hazards.
5. Compliance with relevant laws and regulations, and with ethical principles.
6. Provision of relevant and adequate education and training for all personnel to promote high quality work and safety.

For the future, the Task Force (Coecke et al., 2005) expressed the hope that the guidance will be used widely to establish and maintain best practice in all aspects of cell and tissue culture work, not only by those who are responsible for such work, but also by those who publish its outcomes or who make important policy decisions based on the information it provides (MacLeod and Drexler, 2001).

4 New threats: Cross-contamination and misidentification of cell lines

Cross-contamination is the erroneous contamination of a given culture with cells of another cell line. Most laboratories culturing cells use multiple cell lines simultaneously. Thus, mix ups of two individual cell lines can occur at any point of day-to-day cell culture due to poor tissue culture practice and/or inadequate quality control (quality management, QM) (Hughes et al., 2007; Lindgren, 2008). In earlier times, due to the widespread use of animal cell lines, *interspecies contaminations*, i. e. contamination of cells from different species, were most common. With the advent of "*omics*" technologies in biomedical research there was a strong tendency towards the use of human cell cultures, either normal or tumour cell lines as well as embryonic and adult stem cells. Thus, *intraspecies contaminations between human cell cultures* have become a new threat (O'Brien, 2001). It takes only a few cells of a rapidly proliferating tumour cell line, such as HeLa (see below), to completely overgrow a culture, so that within a few passages, the slower growing cells of the original culture are wiped out.

Instances of cross-contamination of cell lines, especially human cancer cell lines, are more widespread than generally appreciated (Buehring et al., 2004; Masters, 2002b; Markovic and Markovic, 1998). Since the identification of early cases of cross-contaminated cultures (mainly with HeLa cells) (Masters, 2000, 2002a; Nelson-Rees et al., 1981), a number of cases have continued to be identified. In a study of 252 human cell lines, 18% were found to be cross-contaminants (MacLeod et al., 1999). A recent study of more than 500 human leukaemia-lymphoma cell lines found that 15% were misidentified (Drexler et al., 2002, 2003). Despite periodic reminders from concerned cell culturists (Bubenik, 2000; Chatterjee, 2007; Lacroix, 2008; Nardone, 2007; Stacey et al., 2000) the problem appears to continue and even seems to be neglected by the scientific community (Arlett, 2001; Freshney, 2008; MacLeod and Drexler, 2001; MacLeod et al., 2008; Nardone, 2008).

Misidentification by mislabelling of cultures and cell lines, respectively, can also occur. Inadequate labelling of cryovials and/or poor records of frozen stocks in long-term liquid nitrogen storage are the main pitfalls leading to erroneous tracking of the origin of a frozen cell sample (Freshney, 2005; Lindl und Gstraunthaler, 2008). Thus, complete records of cell cultures and their location in storage freezers on the day of freezing as well as periodic updates of the records and of the inventory are essential to unequivocally track the provenance of a given cell line even after years in liquid nitrogen (Freshney, 2002, 2005).

False Cell Lines: How can they be identified and how can cross-contamination be avoided? (Masters, 2002b; Nardone, 2007)

Today, human cells (cell lines) can be easily identified by means of DNA profiling using highly polymorphic short tandem repeats (STRs) as in forensic applications (Dirks et al., 2005; Masters et al., 2001). STR loci are short, repetitive DNA sequences, 3-7 base pairs in length. The STR profiling technique uses fluorescence-based PCR and multiple dye technology to enable the detection of loci with overlapping size ranges. Using this method, various polymorphic STR loci can be amplified by PCR using commercially available sets of primers. The results are genetic signatures for individual identifications (O'Brien, 2001).

DNA fingerprinting proved to be a powerful tool to identify and to unequivocally distinguish individual cell lines, respectively. STR profiling establishes a DNA fingerprint for every human cell line and can be used as an identification record of a particular cell line (Masters et al., 2001; Yoshino et al., 2006). The method is cheap, commercially available, reproducible and comparable between laboratories.

It can be expected that very soon DNA profiling will be the new standard in the authentication of human cell lines. Cell banks (ATCC, ECACC, DSMZ) offer cell line verification services and set up STR profile databases. It is has already been recommended by several authors that every cell culture laboratory should obtain a DNA profile of every human cell line in use in order to authenticate its identity.

References

Anderson, R., O'Hare, M., Balls, M. et al. (1998). The availability of human tissue for biomedical research. ECVAM Workshop Report 32. *ATLA 26*, 763-777.

Arlett, C. F. (2001). The use of dubious cell lines in research: is trust enough? *Lancet Oncol. 2*, 467.

Balls, M., Coecke, S., Bowe, G. et al. (2006). The Importance of Good Cell Culture Practice (GCCP). Proceedings of the 5th World Congress on Alternatives and Animal Use in the Life Sciences, Berlin, Germany, 2005. *ALTEX 23, Spec. Issue*, 270-273.

Balls, M., Goldberg, A. M., Fentem, J. H. et al. (1995). The three Rs: the way forward. The report and recommendations of ECVAM Workshop 11. *ATLA 23*, 838-866.

Barnes, D., McKeehan, W. L. and Sato, G. H. (1987). Cellular endocrinology: Integrated physiology in vitro. *In Vitro Cell. Dev. Biol. 23*, 659-662.

Bubenik, J. (2000). Cross-contamination of cell lines in culture. *Folia Biol. (Praha) 46*, 163-164.

Buehring, G. C., Eby, E. A. and Eby, M. J. (2004). Cell lines cross-contamination: how aware are mammalian cell cultur-

ists of the problem and how to monitor it? *In Vitro Cell. Dev. Biol. 40*, 211-215.

Butler, M. and Jenkins, H. (1989). Nutritional aspects of the growth of animal cells in culture. J. Biotechnol. 12, 97-110.

Chatterjee, R. (2007). Cases of mistaken identity. *Science 315*, 928-931.

Coecke, S., Balls, M., Bowe, G. et al. (2005). Guidance on Good Cell Culture Practice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice. *ATLA 33*, 261-287.

Cooper-Hannan, R., Harbell, J. W., Coecke, S. et al. (1999). The principles of Good Laboratory Practice: application to in vitro toxicology studies. ECVAM Workshop Report 37. *ATLA 27*, 539-577.

Dirks, W. G., Fähnrich, S., Estella, I. A. J. and Drexler, H. G. (2005). Short tandem repeat DNA typing provides an international reference standard for authentication of human cell lines. *ALTEX 22*, 103-109.

Drexler, H. G., Dirks, W. G., Matsuo, Y. and MacLeod, R. A. F. (2003). False leukemia-lymphoma cell lines: an update on over 500 cell lines. *Leukemia 17*, 416-426.

Drexler, H. G., Uphoff, C. C., Dirks, W. G. and MacLeod, R. A. F. (2002). Mix-ups and mycoplasma: the enemies within. *Leukemia Res. 26*, 329-333.

Fedoroff, S. (1967). Proposed usage of animal tissue culture terms. *Exp. Cell Res. 46*, 642-648.

Freshney, R. I. (2008). Authentication of cell lines: ignore at your peril! *Expert Rev. Anticancer Ther. 8*, 311-314.

Freshney, R. I. (2005). *Culture of Animal Cells. A Manual of Basic Techniques*. 5th Ed., 642pp. Wiley & Sons, Inc.

Freshney, R. I. (2002). Cell line provenance. *Cytotechnology 39*, 55-67.

Gstraunthaler, G. (2006). Standardization in cell and tissue culture – The need for specific glp guidelines in the cell culture laboratory (Good Cell Culture Practice – GCCP). Proceedings of the 5th World Congress on Alternatives and Animal Use in the Life Sciences, Berlin, Germany, 2005. *ALTEX 23, Spec. Issue*, 274-277.

Gstraunthaler, G. (2003). Alternatives to the use of fetal bovine serum: serum-free cell culture. *ALTEX 20*, 275-281.

Gstraunthaler, G., Seppi, T. and Pfaller, W. (1999). Impact of culture conditions, culture media volumes, and glucose content on metabolic properties of renal epithelial cell cultures. Are renal cells in tissue culture hypoxic? *Cell. Physiol. Biochem. 9*, 150-172.

Ham, R. G. and McKeehan, W. L. (1979). Media and growth requirements. *Methods Enzymol. 58*, 44-93.

Hartung, T. (2007). Food for thought ... on cell culture. *ALTEX 24*, 143-147.

Hartung, T., Balls, M., Bardouille, C. et al. (2002). Good Cell Culture Practice. ECVAM Good cell culture practice task force report 1. , 407-414.

Hartung, T., Gstraunthaler, G., Coecke, S. et al. (2001). Good Cell Culture Practice (GCCP) – eine Initiative zur Standardisierung und Qualitätssicherung von in vitro Arbeiten. Die Etablierung einer ECVAM Task Force on GCCP. *ALTEX 18*, 75-78.

Hartung, T. and Gstraunthaler, G. (2000). The standardisation of cell culture proceedures. In M. Balls, A.-M. van Zeller and M. Halder (eds.). *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (1655-1658). Amsterdam: Elsevier.

Hartung, T., Gstraunthaler, G. and Balls, M. (2000). Bologna Statement on Good Cell Culture Practice (GCCP). *ALTEX 17,* 38-39.

Hay, R. J., Reid, Y. A., McClintock, P. R. et al. (1996). Cell line banks and their role in cancer research. *J. Cell. Biochem., Suppl. 24*, 107-130.

Hay, R. J. (1988). The seed stock concept and quality control for cell lines. *Anal. Biochem. 171*, 225-237.

Hayflick, L. (1990). In the interest of clearer communication. *In Vitro Cell. Dev. Biol. 26*, 1-6.

Hughes, P., Marshall, D., Reid, Y. et al. (2007). The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? *BioTechniques 43*, 575-582.

Lacroix, M. (2008). Persistent use of "false" cell lines. *Int. J. Cancer 122*, 1-4.

Lindgren, V. (2008). To err is human; to follow the SOP divine. *Int. J. Cancer 123*, 979-980.

Lindl, T. und Gstraunthaler, G. (2008). *Zell- und Gewebekultur. Von den Grundlagen zur Laborbank*. (406pp, 6. Aufl.). Heidelberg: Spektrum Akademischer Verlag.

MacLeod, R. A. F., Dirks, W. G. and Drexler, H. G. (2008). One falsehood leads easily to another. *Int. J. Cancer 122*, 2165-2168.

MacLeod, R. A. F. and Drexler, H. G. (2001). Cell banks detect false cell lines: journals must act too. *Lancet Oncol. 2*, 467-468.

MacLeod, R. A. F., Dirks, W. G., Matsuo, Y. et al. (1999). Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int. J. Cancer 83*, 555-563.

Markovic, O. and Markovic, N. (1998). Cell cross-contamination in cell cultures: the silent and neglected danger. *In Vitro Cell. Dev. Biol. 34*, 1-8.

Masters, J. R. (2002a). HeLa cells 50 years on: the good, the bad and the ugly. *Nature Rev. Cancer 2*, 315-319.

Masters, J. R. (2002b). False cell lines: the problem and a solution. *Cytotechnology 39*, 69-74.

Masters, J. R., Thomson, J. A., Daly-Burns, B. et al. (2001). Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc. Natl. Acad. Sci. USA 98*, 8012-8017.

Masters, J. R. W. (2000). Human cancer cell lines: fact or fantasy. *Nature Rev. Molec. Cell Biol. 1*, 233-236.

Masters, J. R. W., Twentyman, P., Arlett, C. et al. (2000). UK-CCCR guidelines for the use of cell lines in cancer research. *Br. J. Cancer 82*, 1495-1509.

McKeehan, W. L., Barnes, D., Reid, L. et al. (1990). Frontiers in mammalian cell culture. *In Vitro Cell. Dev. Biol. 26*, 9-23.

Nardone, R. M. (2008). Curbing rampant cross-contamination and misidentification of cell lines. *BioTechniques 45*, 221-227.

Nardone, R. M. (2007). Eradication of cross-contaminated cell lines: a call for action. *Cell Biol. Toxicol. 23*, 367-372.

Nelson-Rees, W. A., Daniels, D. W. and Flandermeyer, R. R. (1981). Cross-contamination of cells in culture. *Science 212*, 446-452.

O'Brien, S. J. (2001). Cell culture forensics. *Proc. Natl. Acad. Sci. USA 98*, 7656-7658.

Schaeffer, W. I. (1990). Terminology associated with cell, tissue and organ culture, molecular biology and molecular genetics. *In Vitro Cell. Dev. Biol. 26*, 97-101.

Schaeffer, W. I. (1989). In the interest of clear communication. *In Vitro Cell. Dev. Biol. 25*, 389-390.

Schaeffer, W. I. (1984). Usage of vertebrate, invertebrate and plant cell, tissue and organ culture terminology. *In Vitro 20*, 19-24.

Stacey, G. N., Masters, J. R. W., Hay, R. J. et al. (2000). Cell contamination leads to inaccurate data: we must take action now. *Nature 403*, 356.

Yoshino, K., Iimura, E., Saijo, K. et al. (2006). Essential role for gene profiling analysis in the authentication of human cell lines. *Human Cell 19*, 43-48.

Correspondence to
Gerhard Gstraunthaler, Ph.D.
Division of Physiology
Innsbruck Medical University
Fritz-Pregl-Strasse 3
A-6020 Innsbruck, Austria
e-mail: gerhard.gstraunthaler@i-med.ac.at

Theme 1:
Areas of animal use

Breakout Sessions

BS11: Basic research

Viable Alternatives: Choosing the Proper Experimental Model for Neuroscience Studies

Randall J. Nelson

The Office of Research Compliance, University of Tennessee Health Science Center, Memphis, Tennessee, USA

Summary

Choosing proper experimental models for neuroscience studies is difficult. Cell cultures, other in vitro preparations or non-animal models may be chosen. However, many neuroscience studies need an intact nervous system to control stages of experimentally induced responses.

The choice of neuroscience model systems involves many things. CIOMS Guidelines and U.S. Government Principles provide important tenets when models are selected for studies. When choosing models, consideration should be given to model appropriateness, relevance, minimization of animal use, data quality, and validity of results. Studies designed with proper experimental controls reduce variance. Good design can modify bad initial hypotheses.

Examples will be given that illustrate refinements to best practices that have lead to reductions in the number of animals used as well as reduction in animal pain and distress while under study. Ultimately these refinements, leading to reductions, may lead to replacements as investigators appreciate that valid results can be obtained with fewer animals and less sentient species.

Coenraad Hendriksen suggests that the 3Rs be implemented with common sense, commitment, and communication. Doing so requires an understanding of experimental basics, blending these into the choice of models, and balancing animal welfare and scientific needs when designing and conducting neuroscience studies.

Keywords: animal models, neuroscience, 3Rs, CIOMS principles, alternative choices

Choosing the proper experimental model for neuroscience studies is difficult at best. Balancing scientific needs with appropriate animal welfare when employing the chosen model must be done with care and planning. What follows is a discussion of unique and common aspects of experimental models in neuroscience: how the 3Rs (Russell and Burch, 1959) together with guidance found in the Council for International Organizations of Medical Sciences (CIOMS) principles and the U.S. Government principles, shape model choices; examples of some model choices that refine and thus reduce animal usage; a brief men-

tion the "3Cs" (Hendriksen, 2000) and how they might apply to models in neuroscience; and finally, "3Bs" that might guide model choices.

– Mogenson (2003) reminds us of the three basic types of animal models. They are:
– Homologous – Symptoms displayed and causes of conditions in the animal model are identical to those in human conditions or diseases.
– Isomorphic – Symptoms displayed by the animal model and humans are similar but those in the animal model need not be provoked by the same cause as in humans.

Partial – Most neuroscience models actually make no attempt to model the entire disease but rather "focus on more limited aspects of the human condition."

As with other dynamic models, it is not always clear whether what is observed is *caused by* the experiment or reflects the animal model's ability to compensate for the experimental intervention. Thus, experimenters may *infer* cause and effect but never can be completely *sure* of their connection.

Choosing models to be used in neuroscience studies involves coming to terms with certain realities. Experimental models used in neuroscience studies often involve studying the working nervous system while the nervous system is working. As with any biological experiment, more data is not always better data. Ideally, good experimental design can overcome bad hypotheses. However, the choice of an inappropriate model can overwhelm even the best experimental design, rendering data collected useless and perhaps resulting in the experiments having to be repeated and additional animals used.

The "3Rs" of Russell and Burch (1959) have served not only as ethical guidance for a half a century thus far, they have resulted in scientists becoming more appreciative of the interdependence of these three aspects of planning, initiating, and executing studies with animals in a responsible manner. Refinement of experimental design, undertaken for scientific reasons (and perhaps efficiency) may result in a reduction of the number of animals used. All too often, however, it is difficult to replace intact animal models with non-animal model systems.

Although non-animal models may be desirable for many reasons, animal models often are the most appropriate. Many agree that animal experimentation has contributed in important ways to generalizable knowledge that ultimately has improved the health and welfare of countless humans and animals. The first of the International Guiding Principles for Biomedical Research Involving Animals (CIOMS, 1985) acknowledges that animal experimentation is at times necessary. It states "The advancement of biological knowledge and the development of improved means for the protection of the health and well-being both of man and of animals require recourse to experimentation on intact live animals of a wide variety of species." These principles may be thought of as giving guidance on the application of the 3Rs, for they were undoubtedly derived with them in mind.

Principles III, V, and VII, when taken together, guide the mindset that should accompany experimental refinement. They embody consideration, compassion, and concern by stating:

– III. Animal experiments should be undertaken only after due consideration of their relevance for human or animal health and the advancement of biological knowledge.
– V. Investigators and other personnel should never fail to treat animals as sentient, and should regard their proper care and use and the avoidance or minimization of discomfort, distress, or pain as ethical imperatives.
– VI. Investigators should assume that procedures that would cause pain in human beings cause pain in other vertebrate species, although more needs to be known about the perception of pain in animals.

Principle IV cautions that being cognizant of model species appropriateness and calculations of the number of animals needed to achieve the right group size should be part of the experimental design and should precede the onset of experimentation. It states:

– IV. The animals selected for an experiment should be of an appropriate species and quality, and the minimum number required to obtain scientifically valid results.

Principle II provides the ethical framework for Russell and Burch's remaining "R," that being replacement. In this principle, non-animal alternatives are mentioned in general terms, but the intent is clear.

– II. Methods such as mathematical models, computer simulation, and *in vitro* biological systems should be used wherever appropriate.

If given the opportunity, some would add a fourth "R" and use parts of five principles to shape thinking regarding an experimenter's "responsibilities." Portions of pertinent principles are:

– V. Investigators and other personnel should never fail to treat animals as sentient…
– VII. [Painful procedures] …should be performed with appropriate sedation, analgesia, or anesthesia …
– IX. … animals that would otherwise suffer severe or chronic pain, distress, discomfort, or disablement that cannot be relieved should be painlessly killed.
– X. The best possible living conditions should be maintained for animals kept for biomedical purposes…
– XI. It is the responsibility of the director of an institute or department using animals to ensure that investigators and personnel have appropriate qualifications.

The common denominator in each of these is "respect." That respect extends to the animals, their sentience and ability to feel pain, their living conditions, and qualifications of the personnel that conduct experimentation involving them.

In considering the 3Rs and how they relate to neuroscience experiments, there may be a "scientific imbalance." Refinements often lead directly to reductions in the number of animals devoted to neuroscience studies. Replacement, however, still most likely takes the form of substituting a less sentient animal model for one traditionally used, rather than replacing animal with non-animal models.

Any time animal models are chosen, those choosing them should proceed with caution. In the book, *The Scalpel and the Butterfly* (Rudacille, 2000), the author provides clarity to some concerns about the use of animal models to investigate human

conditions. She states, "Yet by using animals to construct a biomedically accessible (and definable) human subject, science has focused precisely on those aspects of human physiology and behavior that are most animal-like, ignoring that which does not conform to the model." She goes on to note, "Initially…work focused purely on the physiological, but increasingly…the psychological has been an equally intense focus of study." Despite these concerns, vast strides in the understanding of disease processes and the dynamic workings of the nervous system have been made by the ethical use of animals in neuroscience experimentation.

There have been numerous examples, past and present, where the appropriate choice of an experimental model for neuroscience studies has been crucial. One such example was succinctly presented in "Translational Neuroscience Accomplishments", available from the Society for Neuroscience (2003). It states the following: *"…it was demonstrated that polio was an infectious disease caused by a virus. It could be transmitted to monkeys (but not rabbits, guinea pigs, and mice) by the injection of filtered homogenates of human spinal cord from a patient that died from polio*."

Vaccines developed from animal models virtually eliminated polio from the childhood landscape. Ineffective models may have cost more animal and human lives, as well as time and money. Through this example we begin to see the need for balancing animal welfare and scientific concerns.

Another example involves refining experimental designs by selectively tagging animals that have desired genetic characteristics so that they can be identified with ease. Animals that have been genetically manipulated may have green fluorescent protein (GFP) as their transgene or linked to another protein (see, e.g., Capogna, 2008). When illuminated properly, rodent pups having the GFP construct glow, whereas those without it do not. The identification of those animals possessing desired traits can be accomplished without the use of biopsies and without the need for stress-inducing handling. Moreover, non-invasive selection increases experimental efficiency.

Parkinson's disease is a debilitating neurological disorder involving abnormal dopamine levels in the basal ganglia. The basal ganglia control goal-directed movements via a complex circuit that consists of "go" and "stop" neurons. The stop neurons are normally inhibited. When dopamine levels are low, stop neurons are no longer inhibited and movements grind to a halt. Chemical intervention can relieve symptoms, but this stops being effective after time. Organotypic co-cultures that include parts of the basal ganglia and the neocortex can be kept viable for extended periods of time, form and maintain appropriate functional interconnections, and thus be used to study many parts of the normal physiology of the system that is devastated by Parkinson's disease (see, e.g., Plenz and Kitai, 1998).

Deep brain stimulation in the right part of the brain can relieve the paucity of movement when chemical treatment can't. Finding the right part of the brain to stimulate in humans requires experimentation with animal models that have subthalamic nuclei and an internal segment of the globus pallidus with the right degree of "humanness" to be a predictive model of this disease's symptoms. Recent advances in rodent models allow optical stimulation of target neurons with blue light. This results in neuronal activation and ultimately in improvement of the diseases symptoms (Gradinaru et al., 2009).

Other neurological diseases that have been investigated with carefully chosen animal models include: strabismus (with cats), memory deficits (with rodents and nonhuman primates), retinitis pigmentosa (with genetically defective dogs), stroke (with rodents) and bipolar disease (with guinea pigs). Each of these models was selected because of some unique trait that set the particular model apart as the best partial, isomorphic or ideally, homologous model system.

Sometimes, human subjects become the replacement model used to predict animal behavior. Humans can perform complex behavioral tasks. They give accurate feedback about design flaws and expectations in goal-oriented tasks. Thus they are the ideal species for doing initial, non-invasive experiments that refine experimental designs. Recent work on reward preferences and behavioral strategies has been done using Old World monkeys. This work is patterned after human psychological experiments that were conducted first. One is then left to wonder if this is experimental *refinement* as a result of human *replacement*.

In a similar manner, Dehaene and Changeux (1993) used neural network modeling to predict human brain centers that should control the storage and processing of numbers, coining the term "numerosity detectors" for the elements that would do so. This lead Nieder and Miller (2002, 2004) to discover the numerosity regions in the brains of macaque monkeys. Their work lead Dehaene and co-workers (Piazza et al., 2004) to determine the location of these regions in humans, using non-invasive imaging techniques.

Hendriksen (2000), although not in the context of choosing animal models for neuroscience experiments, suggests that the 3Rs should be implemented with "3Cs": common sense, commitment, and communication. Common sense in choosing animal models certainly involves selecting the right model to answer the scientific questions while not losing sight of animal welfare in the process. This is done through commitment to the 3Rs and to the scientific goals. Communication is key, both within the research group and with others as well so that methods are continually refined, the number of animals involved is reduced, and animals are replaced when possible with other model systems.

Continuing to refine how three letters, which stand for three words, might be used to focus discussions about science and animal welfare in model selection, Nelson's "3Bs" (2008) may be applied here. Originally proposed to guide animal ethics committees in evaluating novel behavioral proposals for animal experimentation, "Basics, Balance and Blend" can easily be applied to animal model selection for neuroscience experiments. Committed and responsible researchers continually have the opportunity to evaluate the basics of the models available and use established ones when possible, thus avoiding the need for model re-validation. They may *reduce* animal numbers as a result. They can determine what the chosen model can "model" and what it cannot, achieving a *balance* between ideals and practical scientific goals. They can then *refine* their experimental goals

accordingly. Prudent scientists should blend the need to use homologous models with doing "initial studies" with partial models, as defined above, and replace whole animal studies with *in vitro* experiments whenever possible.

Choosing the proper experimental model for neuroscience studies is an inexact science. With common sense, commitment, and communication, those making choices can achieve a balance between good science and good animal welfare by blending these basic goals responsibly. In doing so, those choosing can select from among the viable alternatives.

References

Capogna, M. (2008). *Understanding the emotional brain*. eStrategies Projects 5, 66-67. (see http://mrcanu.pharm.ox.ac.uk/publicationsearch.php?au=capogna)

CIOMS Guidelines (1985). International guiding principles for biomedical research involving animals. http://www.cioms.ch/frame_1985_texts_of_guidelines.htm

Dehaene, S. and Changeux, J.-P. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci. 5*, 390-407.

Gradinaru, V., Mogri, M., Thompson, K. R. et al. (2009). Optical deconstruction of parkinsonian neural circuitry. *Science 324*, 354-359.

Hendriksen, C. F. M. (2000). Replacement, reduction and refinement and biologicals: About facts, fiction, and frustration. In M. Balls, A.-M. van Zeller and M. E. Halder (eds). *Progress in the Reduction, Refinement, and Replacement of Animal Experimentation* (51-63). Amsterdam: Elsevier Science.

Piazza, M., Izard, V., Pinel, P. et al. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron 44*, 547-555.

Plenz, D. and Kitai, S. T. (1998). Up and down states in striatal medium spiny neurons simultaneously recorded with spontaneous activity in fast-spiking interneurons studied in cortex-striatum-substantia nigra organotypic cultures. *J. Neurosci. 18*, 266-283.

Mogenson, J. (2003). Animal Models in Neuroscience. In J. Hau and G. L. van Hoosier Jr. (eds), *Handbook of Laboratory Animal Science, Volume II, Animal Models* (95-109). Boca Raton: CRC Press.

Nelson, R. J. (2008). Behavioral Studies and the IACUC: Challenges and opportunities. *ILAR J. 50*, 81-84.

Nieder, A. and Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *PNAS 101*, 7457-7462.

Nieder, A., Freedman, D. J. and Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science 297*, 1708-1711.

Rudacille, D. (2000). *The Scalpel and the butterfly*. New York: Farrar, Strauss and Giroux.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen & Co. LTD.

Society for Neuroscience (2003). Translational Neuroscience Accomplishments. http://www.sfn.org/index.cfm?pagename=gpa_TranslationalAccomplishments.

Correspondence to

Randall J. Nelson
The Office of Research Compliance
The University of Tennessee Health Science Center
910 Madison Avenue, Suite 650
Memphis, Tennessee, 38163 USA
e-mail: rnelson@uthsc.edu

Proteolytic Cleavage of Adherens Junction Components During Fas-Dependent Cell Death in Primary Cultures of Rat Hepatocytes

Mathieu Vinken[1,#], Elke Decrock[2,§], Luc Leybaert[2], Tamara Vanhaecke[1,#] and Vera Rogiers[1]

[1]Department of Toxicology, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Belgium; [2]Department of Basic Medical Sciences-Physiology Group, Faculty of Medicine and Health Sciences, Ghent University, Belgium

[#]M. V. and T. V. are postdoctoral research fellows of the Fund for Scientific Research Flanders (FWO-Vlaanderen), Belgium; [§]E. D. is a doctoral research fellow of the Fund for Scientific Research Flanders (FWO-Vlaanderen), Belgium

Summary

We previously introduced an in vitro model of hepatocellular apoptotic cell death consisting of monolayers of freshly isolated rat hepatocytes that are exposed to a combination of Fas ligand and cycloheximide for a limited time span. The current study was set up to demonstrate the actual applicability of this system for investigating molecular events that typify apoptosis. We specifically focussed on adherens junctions, a group of cell contacts composed of cadherin-catenin complexes that play essential roles in the control of liver homeostasis. It was found that E-cadherin expression gradually declined during the cell death process, whereas both β-catenin and γ-catenin were progressively degraded, yielding a number of proteolytic fragments. These results support the notion that dismantling of adherens junctions during apoptosis depends on proteolytic processing of its components. This outcome, in turn, illustrates the usefulness of the in vitro model as an experimental tool to study apoptosis, whilst reducing animal experimentation.

Keywords: in vitro model, primary hepatocyte, cadherin, catenin, apoptosis

Abbreviations: Ac-DEVD-AFC, acetyl-Asp-Glu-Val-Asp-7-amino-4-trifluoromethylcoumarin; CHX, cycloheximide; FasL, Fas ligand; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; LDH, lactate dehydrogenase; PBS, phosphate-buffered saline solution; $PBSD^{+}$, divalent ion-supplemented PBS; LH, liver homogenate; TBS, Tris-buffered saline solution.

1 Introduction

The liver is the principal site of xenobiotic biotransformation in the body. Inherent to its detoxifying function, this organ, and more specifically the hepatocyte, is also a primary target for chemical-induced cell injury (Jaeschke et al., 2002; Vanhaecke and Rogiers, 2006; Vinken et al., 2006). For many years it was thought that cell death, elicited by toxicants, mainly occurred through necrosis. It has now become clear that an alternative cell death mode, namely apoptosis, predominates in toxicant-induced cell death (Gill and Dive, 2000; Gomez-Lechon et al., 2002; Jaeschke et al., 2002; Raffray and Cohen, 1997). The most prominent signalling pathway that drives apoptotic cell death in the liver is initiated by binding of specific ligands, including the Fas (CD95/Apo-1) ligand (FasL), to their receptors at the cell plasma membrane surface. Hepatocytes abundantly produce the Fas receptor and ligand binding leads to induction of caspase 8. Activated caspase 8 triggers caspase 3, which subsequently cleaves a plethora of cellular components (Mahli et al., 2006; Mahli and Gores, 2008). In fact, several proteins relevant to cytoskeletal architecture are present among the caspase 3 substrates, including keratins (Caulin et al., 1997; Ku et al., 1997), lamins (Orth et al., 1996; Takahashi et al., 1996), gelsolin (Kothakota et al., 1997), vimentin (Caulin et al., 1997) and catenins (Brancolini et al., 1997; Schmeiser et al., 1998). The latter are a family of cytoplasmic proteins that interact with transmembrane cadherins (Wheelock and Johnson, 2003). In the liver, hepatocytes mainly express E-cadherin, β-catenin and γ-catenin (plakoglobin) (Vinken et al., 2006). Cadherin-catenin complexes are the essential building stones of adherens junctions, which connect cells one to another (Wheelock and Johnson, 2003). The caspase 3-mediated breakdown of catenins, and thus the disruption of cell-cell contacts, is believed to underlie the typical morphological changes that are associated with apoptosis, including membrane blebbing and cytoplasmic condensation (Brancolini et al., 1998; Schmeiser and Grand, 1999).

A number of protocols have been described to study hepatocellular apoptosis *in vivo*, such as the direct administration of cell death-evoking toxicants to animals (Furukawa et al., 2000). Such experiments not only raise serious ethical ques-

tions but are also of limited scientific value. Indeed, apoptotic cells are barely detectable *in vivo*, as they are rapidly engulfed by neighbouring phagocytes. During *in vitro* experimentation, where phagocytosis does not take place, the full course of apoptosis can be monitored (Bai et al., 2003; Gill and Dive, 2000; Gomez-Lechon et al., 2002; Raffray and Cohen, 1997). In this respect, we have recently introduced a primary hepatocyte culture system to investigate Fas-mediated apoptotic cell death. In the established model freshly isolated hepatocytes are cultivated in a monolayer configuration and are exposed to a combination of FasL and cycloheximide (CHX) for a limited time span. A battery of well-established cell death markers was used for the biochemical characterisation of this system, and we could demonstrate termination of the apoptotic process and simultaneous transition to a necrotic phenotype in this experimental setting (Vinken et al., 2009). The current study was set up to demonstrate the actual applicability of this *in vitro* model for studying the molecular mechanisms that typify apoptosis. We hereby specifically focussed on adherens junctions, as their orchestrated degradation is a well-known event occurring during apoptotic cell death.

2 Animals, materials and methods

Chemicals and reagents

FasL and CHX came from Alexis (Switzerland) and Sigma (Belgium), respectively. Acetyl-Asp-Glu-Val-Asp-7-amino-4-trifluoromethylcoumarin (Ac-DEVD-AFC) was purchased from Merck (Belgium). Propidium iodide and Hoechst 33342 were from Invitrogen (Belgium), whereas annexin-V-Fluos was obtained from Roche (Germany). All other chemicals were commercially available products of analytical grade and were supplied by Sigma (Belgium), unless specified otherwise.

Hepatocyte cultivation and cell death induction

Procedures for the housing of rats and isolation and cultivation of hepatocytes were approved by the local ethics committee of the Vrije Universiteit Brussel (Belgium). Male outbred Sprague-Dawley rats (200-250g; Charles River Laboratories, Belgium) were kept under controlled environmental conditions (12 h light/dark cycles) with free access to food and water. Hepatocytes were isolated by use of a two-step collagenase method, and cell viability was assessed by trypan blue exclusion (Papeleu et al., 2006). Viable (>85%) hepatocytes were plated at a density of 0.56 x 10^5 cells/cm^2 in William's medium E (Invitrogen, Belgium) containing 7 ng/ml glucagon, 292 mg/ml L-glutamine, antibiotics (7.33 I.E./ml sodium benzyl penicillin, 50 μg/ml kanamycin monosulphate, 10 μg/ml sodium ampicillin, 50 μg/ml streptomycin sulphate) and 10% fetal bovine serum. After 4 h and 24 h, the culture medium was removed and replaced by serum-free medium supplemented with 25 μg/ml hydrocortisone sodium hemisuccinate and 0.5 μg/ml insulin ("hepatocyte culture medium"). Cell death was induced 44 h post-plating by adding hepatocyte culture medium containing 200 ng/ml FasL and 2 μg/ml CHX. Sampling was performed at the start of cell death induction (0 h) and 2 h, 4 h and 6 h thereafter (Vinken et al., 2009).

Determination of the apoptotic and necrotic indices

Hepatocytes, seeded on 35 mm culture dishes, were washed twice with phosphate-buffered saline solution (PBS) containing 1.2 mM $CaCl_2$ and 340 μM $MgCl_2 \cdot 6H_2O$ ($PBSD^+$). Subsequently, cells were stained with 2% annexin-V-Fluos, 3 μg/ml Hoechst 33342 and 1 μg/ml propidium iodide in annexin-V buffer (140 mM NaCl, 5 mM $CaCl_2$, 10 mM Hepes) for 15 min at room temperature. Culture dishes were thereafter rinsed four times with PBSD+ and subjected to fluorescence microscopy (Nikon, Japan). At least five images per culture dish were taken (original magnification x100). The number of cells positive for the concerned marker was counted in each image and expressed relative to the total number of cells present. Analyses were carried out in a blinded fashion.

Measurement of caspase 3-like activity

Hepatocytes, seeded on 60 mm culture plates, were washed twice with cold PBS, resuspended in lysis buffer (10 mM Hepes, 2 mM ethylenediamine tetra-acetic acid, 1 mM phenylmethylsulphonylfluoride, 14.6 μM pepstatin A, 21.6 μM aprotinin, 3.1 μM leupeptin, 2 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate, 5 mM dithiothreitol), and subjected to three freeze-thawing cycles. Cell lysates were centrifuged at 1559 x *g* for 30 min at 4°C, and protein concentrations were determined according to the Bradford procedure (Bradford, 1976) using a Bio-Rad protein assay kit (Bio-Rad, Germany) with bovine serum albumin as a standard. Subsequently, 30 μg protein was added to reaction buffer (10 mM Pipes, 2 mM ethylenediamine tetra-acetic acid, 5 mM dithiothreitol, 1.6 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate). Then, 1.375 mM of Ac-DEVD-AFC substrate was added, and after 1 h incubation at 37°C, fluorescence was measured at an excitation wavelength of 400 nm and an emission wavelength of 505 nm in a Victor plate reader (PerkinElmer, Belgium). The background readings from the buffer were subtracted from the readings from the samples and the enzyme activity was expressed as nmol AFC/min x μg protein.

Measurement of lactate dehydrogenase leakage

Hepatocyte membrane damage was evaluated by determination of the lactate dehydrogenase (LDH) index (Bergmeyer, 1974) using a commercially available kit (Merck, Germany). The LDH index was calculated by the following equation: [100 x LDH activity in supernatant]/[LDH activity in (supernatant+cells)].

Immunoblot analysis

For the preparation of total protein lysates, hepatocytes were harvested from 60 mm culture plates by scraping and washed twice with ice-cold PBS. Cells were homogenised in lysis buffer (50 mM Hepes, 150 mM NaCl, 1 mM ethylenediamine tetra-acetic acid, 2.5 mM ethyleneglycol tetra-acetic acid, 0.1% Tween 20, 10% glycerol) supplemented with 0.1 mM phenylmethylsulphonylfluoride, 5 mM dithiothreitol and 1x protease inhibitor cocktail (Roche, Germany). Following sonication for 10 s, samples were left on ice for 1 h. Cell lysates were spun at 13791 x *g* for 5 min, and protein concentrations were determined in the supernatants according to the Bradford procedure

Tab. 1: Primary antibodies used for immunoblot analysis. (GAPDH, glyceraldehyde-3-phosphate dehydrogenase).

Antigen	Supplier	Cat. No.	Species	Type	Dilution
E-cadherin	BD Biosciences	610181	mouse	monoclonal	1/5000
β-catenin	BD Biosciences	610153	mouse	monoclonal	1/1000
γ-catenin	BD Biosciences	610253	mouse	monoclonal	1/2000
GAPDH	Abcam	ab8245	mouse	monoclonal	1/20000

(Bradford, 1976) using a Bio-Rad protein assay kit (Bio-Rad, Germany) with bovine serum albumin as a standard. Proteins (25 μg) were fractionated on 7.5% sodium dodecyl sulphate polyacrylamide and blotted afterwards onto nitrocellulose membranes (Amersham, United Kingdom). The efficiency of transfer was controlled by reversible Ponceau S staining. Subsequent blocking of the membranes was performed with 5% non-fatty milk in Tris-buffered saline solution (TBS; 20 mM Tris, 135 mM NaCl) containing 0.1% Tween 20. Membranes were incubated for 2 h at room temperature with primary antibody (Tab. 1), followed by incubation for 1 h at room temperature with horseradish peroxidase-conjugated secondary antibody directed against mouse (Amersham, United Kingdom). Excessive antibody was removed by washing the membrane several times in Tween-supplemented TBS. Detection of the proteins was carried out by means of an enhanced chemiluminescence

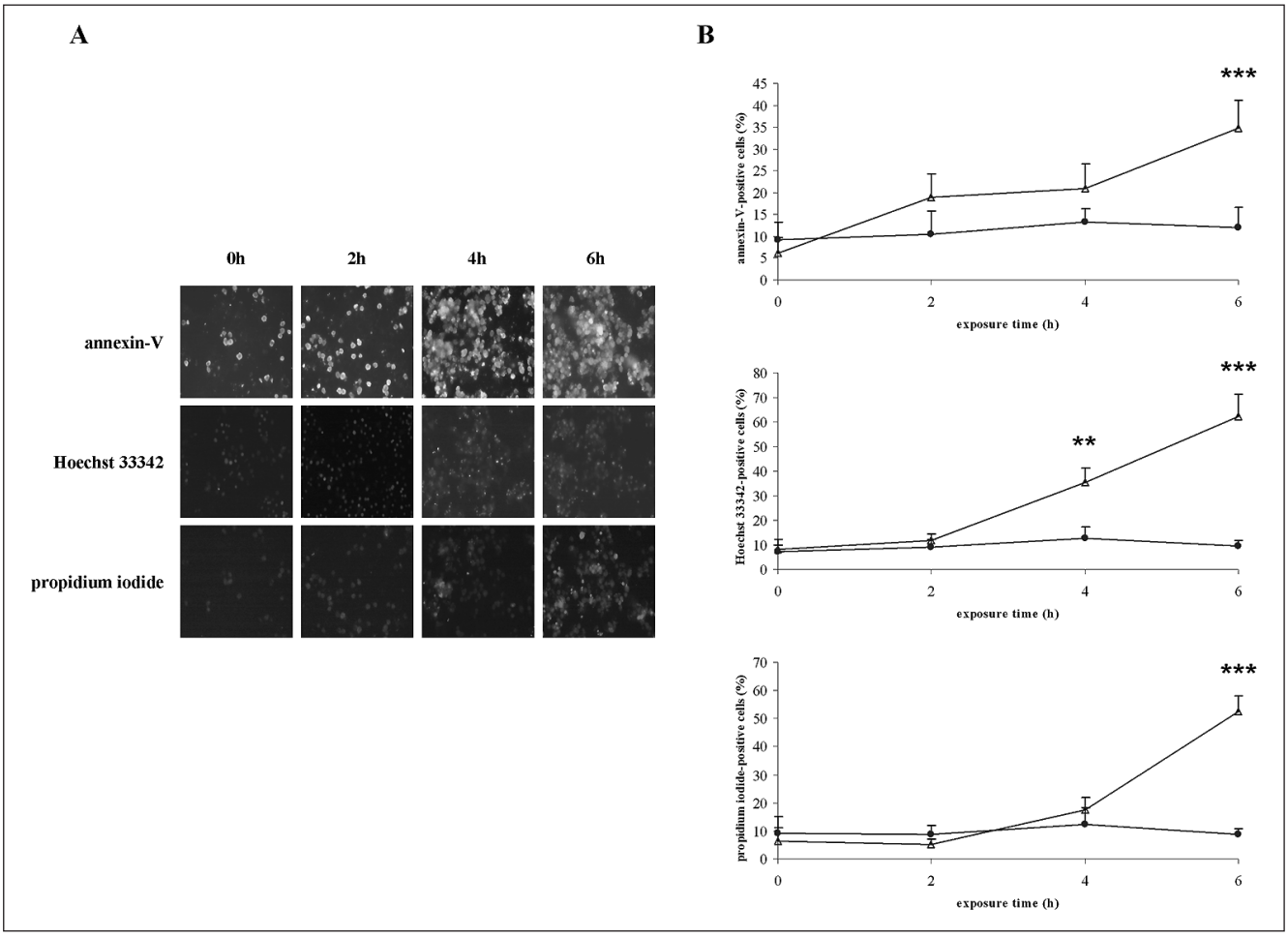


Fig. 1: Characterisation of the cell death response induced by FasL/CHX in primary cultures of rat hepatocytes.
Freshly isolated rat hepatocytes were cultivated in a monolayer configuration and were exposed to 200 ng/ml FasL and 2 μg/ml CHX, starting 44 h post-plating. (A) Samples were taken at the start of the exposure (0 h), and after 2 h, 4 h and 6 h and were stained with annexin-V-Fluos, Hoechst 33342 and propidium iodide, as described in "Animal, materials and methods" (original magnification x100). (B) The number of cells positive for the concerned marker was counted in each image and was expressed relative to the total number of cells present. Data are expressed as means ± standard deviation of three independent experiments. Results were evaluated by one-way ANOVA (repeated measures), followed by *post hoc* Bonferroni tests. (●=control; Δ=FasL/CHX). (**p<0.01; ***p<0.001).

Western blotting system (Amersham, United Kingdom). In order to confirm equal loading of proteins, membranes were further incubated with a monoclonal anti-mouse glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibody (Abcam, United Kingdom).

Statistical analysis

Data are expressed as means ± standard deviation of three independent experiments (n=3). Results were evaluated by one-way ANOVA (repeated measures) followed by *post hoc* Bonferroni tests.

3 Results

Characterisation of the cell death response induced by FasL/CHX in primary hepatocyte cultures

FasL is known to trigger the extrinsic apoptotic pathway upon binding to its receptor at the cell plasma membrane surface, and CHX, a potent inhibitor of protein synthesis, was repeatedly reported to amplify this process (Ni et al., 1994; Rouquet et al., 1996). To chart the time course of the cell death response, we carried out combined *in situ* stainings with annexin-V-Fluos, Hoechst 33342 and propidium iodide at the start (0 h) of exposure of primary hepatocytes to FasL/CHX, and 2 h, 4 h and 6 h thereafter (Fig. 1A,B). Measurements of caspase 3-like activity and extracellular LDH release were performed in parallel (Fig. 2).

An early event during the commitment of cells to apoptosis includes the externalisation of phosphatidyl serine, a negatively charged phospholipid normally restricted to the inner surface of the cell plasma membrane bilayer. This process can be monitored using the phospholipid binding protein annexin-V (Foster, 2000). Indeed, the number of annexin-V-positive cells already increased, though not significantly, after 2 h of cell death induction (Fig. 1B). A key feature of apoptosis includes the cleavage of cellular proteins by caspase 3. The latter can be measured by using the prototypical though not exclusive caspase 3 substrate Ac-DEVD-AFC (Bai et al., 2003). Caspase 3-like activity significantly increased after 2 h and peaked after 4 h of exposure of the primary hepatocytes to FasL/CHX (Fig. 2A). Chromatin fragmentation and condensation, a deleterious result of caspase 3 action occurring in a late phase of the apoptotic process, can be revealed with the dye Hoechst 33342, which intercalates with DNA (Bai et al., 2003; Foster, 2000). Accordingly, Hoechst 33342 fluorescence became more intense towards the final stage of the provoked cell death phenomenon (Fig. 1B). *In vivo*, late apoptotic cells are rapidly engulfed by neighbouring phagocytes. During *in vitro* experimentation, however, in the absence of phagocytes, late apoptotic cells switch to a rather necrotic mode (Bai et al., 2003; Gill and Dive, 2000; Gomez-Lechon et al., 2008). A typical hallmark of necrotic cell death, unlike apoptosis, is the loss of cell plasma membrane integrity. Propidium iodide only permeates cells displaying a disrupted cell plasma membrane and can thus be used to discriminate between necrotic and apoptotic cells (Foster, 2000). Propidium iodide-positive counts indeed culminated 6 h after addition of FasL/CHX to the cell culture medium of primary hepatocytes, indicating a massive apoptotic-to-necrotic transition (Fig. 1B). The acquisition of a necrotic phenotype at this time point was further demonstrated by a drastically increased LDH index, a cytosolic enzyme that leaks to the extracellular environment upon cell plasma membrane damage (Bergmeyer, 1974) (Fig. 2B).

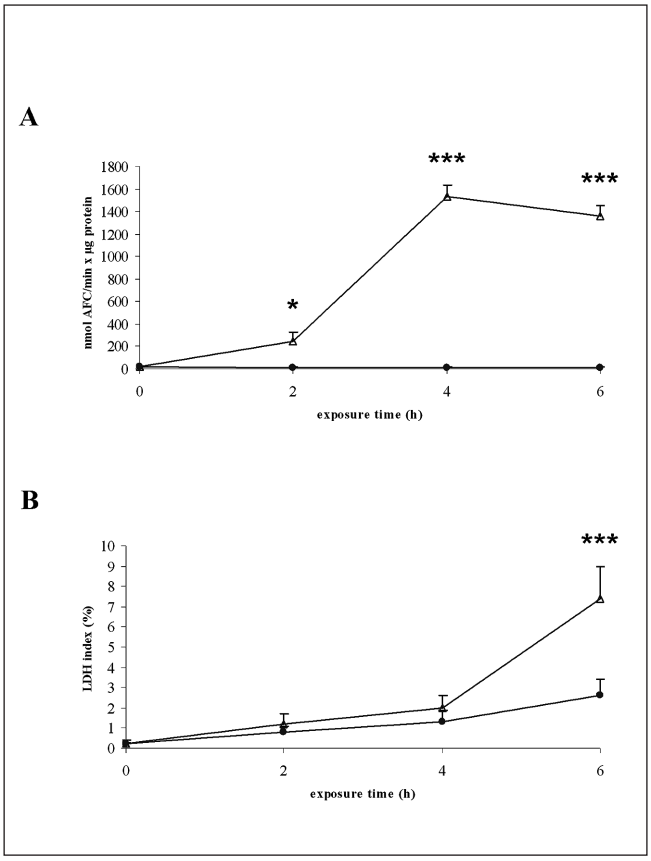


Fig. 2: Modifications in apoptotic and necrotic activities induced by FasL/CHX in primary cultures of rat hepatocytes.
Freshly isolated rat hepatocytes were cultivated in a monolayer configuration and were exposed to 200 ng/ml FasL and 2 µg/ml CHX, starting 44 h post-plating. Samples were taken at the start of the exposure (0 h) and after 2 h, 4 h and 6 h.
(A) Caspase 3-like activity was determined as described in "Animals, materials and methods". Results are expressed as nmol AFC/min x µg protein and represent means ± standard deviation of three independent experiments. Results were evaluated by one-way ANOVA (repeated measures) followed by *post hoc* Bonferroni tests.
(B) The LDH index was determined as described in "Animals, materials and methods" and was calculated by the following equation: [100 x LDH activity in supernatant]/[LDH activity in (supernatant+cells)]. Data are expressed as means ± standard deviation of three independent experiments. Results were evaluated by one-way ANOVA (repeated measures), followed by post hoc Bonferroni tests. (●=control; Δ=FasL/CHX). (*p<0.05; ***p<0.001). (AFC, 7-amino-4-trifluoromethylcoumarin; LDH, lactate dehydrogenase).

Effect of FasL/CHX on adherens junctions proteins in primary hepatocyte cultures

Following the establishment of the *in vitro* model of hepatocellular cell death, we investigated adherens junctions in this experimental setting. More specifically, protein levels of E-cadherin, β-catenin and γ-catenin were studied by immunoblot analysis, thereby using a set of monoclonal antibodies. We included a rat liver homogenate in order to have reference data of adherens junction expression in the rat liver *in vivo* (Fig. 3).

Despite the drastic abolishment of cellular interactions that hepatocytes undergo during their isolation from the liver, cell-cell contacts are restored, at least in part, upon subsequent cultivation of the cells (Vinken et al., 2006). In fact, adherens junction components are known to be abundantly and stably expressed in primary hepatocyte cultures, thus reflecting the need of the cells to re-establish their connectivity (Vinken et al., 2007). In line with this information, E-cadherin, β-catenin and γ-catenin proteins were expressed at similar levels in primary hepatocyte cultures and in the rat liver homogenate (Fig. 3). For E-cadherin, one single signal was observed around 120 kDa in all experimental conditions tested. The E-cadherin steady-state protein levels gradually declined when FasL/CHX was added to the cell culture medium of the primary hepatocytes. The monoclonal antibody used to detect β-catenin also recognised one fragment, *in casu* at 92 kDa, corresponding with the native β-catenin protein. Immunoreactivity for this fragment diminished after 4 h of exposure of the hepatocytes to FasL/CHX. At this time point, a second β-catenin fragment (75 kDa), representing a proteolytic cleavage product, became manifested. Plakoglobin appeared as a dual signal of which the upper band (82 kDa) was clearly predominant in primary hepatocyte cultures. During the triggered cell death process, its importance progressively decreased in favour of the 75 kDa fragment. Moreover, two additional truncated γ-catenin fragments, with approximate molecular weights of 65 kDa and 55 kDa, became detectable after 4 h, but particularly after 6 h of cell death induction (Fig. 3).

3 Discussion

Our research group is addressing the development and optimisation of liver-based *in vitro* systems, particularly primary hepatocyte culture models, for pharmaco-toxicological purposes. We previously described cultivation conditions to specifically study hepatocellular proliferation (Papeleu et al., 2004) and differentiation (Henkens et al., 2007), two critical aspects of the hepatocyte life cycle. Liver homeostasis also relies on the appropriate removal of hepatocytes through apoptosis. In normal liver, the incidence of apoptosis is very low, affecting 0.05% of all liver cells in rats (Qiao and Farrell, 1999). Upon exposure to hepatotoxicants, however, apoptotic activity may drastically increase (Gill and Dive, 2000; Gomez-Lechon et al., 2002; Jaeschke et al., 2002; Raffray and Cohen, 1997). The importance of apoptosis in toxicology has been underestimated for a long time, mainly because of the difficulty of identifying apoptotic cells in the intact organism owing to their rapid phagocytosis. Such constraints can be overcome by the use of *in vitro* experimental settings (Bai et al., 2003; Gill and Dive, 2000; Gomez-Lechon et al., 2002; Raffray and Cohen, 1997). We have recently established and fully characterised an *in vitro* model of Fas-mediated hepatocyte cell death (Vinken et al., 2009). The aim of the present study was to further test the suitability of this system for studying the

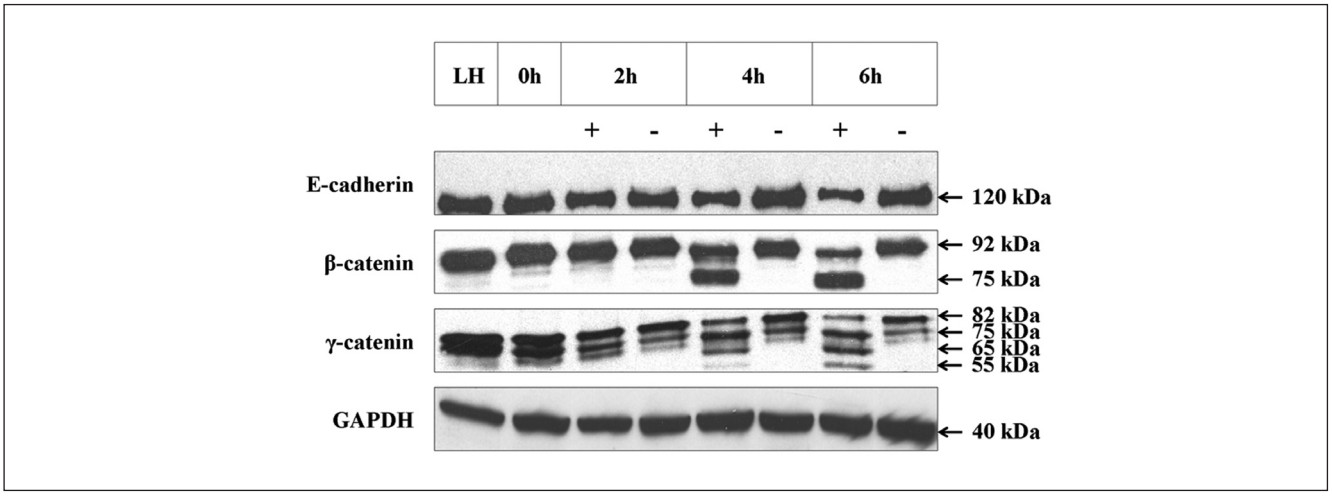


Fig. 3: Alterations in the expression of adherens junction components induced by FasL/CHX in primary cultures of rat hepatocytes.

Freshly isolated rat hepatocytes were cultivated in a monolayer configuration and were exposed to 200 ng/ml FasL and 2 μg/ml CHX, starting 44 h post-plating. This condition is indicated as "+", whereas untreated controls are represented as "-". Samples were taken at the start of the exposure (0 h), and after 2 h, 4 h and 6 h and were subjected to immunoblot analyses as described in "Animals, materials and methods", using primary antibodies (see Tab. 1) raised against E-cadherin, β-catenin and γ-catenin. The different protein fragments and their corresponding molecular weights are indicated by arrows. Blots shown are representative of three independent experiments. (GAPDH, glyceraldehyde-3-phosphate dehydrogenase; LH, liver homogenate).

molecular determinants of apoptosis. We paid specific attention to adherens junctions, a group of specialised cell contacts composed of cadherin-catenin complexes that play essential roles in the control of liver homeostasis (Vinken et al., 2006).

We found that progression through the Fas-triggered cell death process is associated with the active degradation of the native β-catenin and γ-catenin proteins, yielding a number of lower mass proteolytic fragments which started to appear at the time whereby maximal caspase 3-like activity was measured. This finding is consistent with previous reports, showing caspase 3-mediated proteolysis of both β-catenin and plakoglobin during chemical-induced or physical-induced apoptosis in a plethora of cell types (Bannerman et al., 1998; Brancolini et al., 1998; Fuchs et al., 2008; Fukuda, 1999; Herren et al., 1998; Kessler and Muller, 2009; Schmeiser and Grand, 1999; Steinhusen et al., 2000; Su et al., 2005). Although this scenario usually also holds for E-cadherin (Brancolini et al., 1998; Fuchs et al., 2008; Keller and Nigam, 2003; Kessler and Muller, 2009; Schmeiser and Grand, 1999; Steinhusen et al., 2001), we could only detect one signal that gradually decreased upon cell death induction without the appearance of additional bands. This failure to detect a clear-cut proteolytic E-cadherin pattern, unlike in most other studies, is probably due to the use of distinct antibodies that recognise different epitopes that are not present on the truncated E-cadherin fragments.

In conclusion, the results of the present study further support the notion that dismantling of adherens junctions during apoptosis, which gives rise to the typical apoptotic phenotype, depends on the concerted proteolytic processing of its building blocks. This outcome, in turn, further illustrates the usefulness of the *in vitro* model of Fas-mediated cell death as an experimental tool to study apoptosis whilst reducing animal experimentation.

References

Bai, L., Wang, J., Yin, X. M. and Dong, Z. (2003). Analysis of apoptosis: basic principles and procedures. In X. M. Yin and Z. Dong (eds.), *Essentials of apoptosis: a guide for basic and clinical research* (239-251). New Jersey: Humana press.

Bannerman, D. D., Sathyamoorthy, M. and Goldblum, S. E. (1998). Bacterial lipopolysaccharide disrupts endothelial monolayer integrity and survival signaling events through caspase cleavage of adherens junction proteins. *J. Biol. Chem. 273*, 35371-35380.

Bergmeyer, H. U. (1974). Lactate dehydrogenase. In H. U. Bergmeyer (ed.), *Methods of enzymatic analysis* (574-579). New York: Academic Press.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248-254.

Brancolini, C., Lazarevic, D., Rodriguez, J. and Schneider, C. (1997). Dismantling cell-cell contacts during apoptosis is coupled to a caspase-dependent proteolytic cleavage of beta-catenin. *J. Cell. Biol. 139*, 759-771.

Brancolini, C., Sgorbissa, A. and Schneider, C. (1998). Proteolytic processing of the adherens junctions components beta-catenin and gamma-catenin/plakoglobin during apoptosis. *Cell Death Differ. 5*, 1042-1050.

Caulin, C., Salvesen, G. S. and Oshima, R. G. (1997). Caspase cleavage of keratin 18 and reorganization of intermediate filaments during epithelial cell apoptosis. *J. Cell Biol. 138*, 1379-1394.

Foster, J. R. (2000). Detection and biomarkers of apoptosis. In R. Roberts (ed.), *Apoptosis in toxicology* (213-232). London: Taylor & Francis.

Fuchs, M., Hermannstadter, C., Hutzler, P. et al. (2008). Deletion of exon 8 increases cisplatin-induced E-cadherin cleavage. *Exp. Cell Res. 314*, 153-163.

Fukuda, K. (1999). Apoptosis-associated cleavage of beta-catenin in human colon cancer and rat hepatoma cells. *Int. J. Biochem. Cell Biol. 31*, 519-529.

Furukawa, S., Usuda, K., Fujieda, Y. et al. (2000). Apoptosis and cell proliferation in rat hepatocytes induced by barbiturates. *J. Vet. Med. Sci. 62*, 23-28.

Gill, G. H. and Dive, D. (2000). Apoptosis: basic mechanisms and relevance to toxicology. In R. Roberts (ed.), *Apoptosis in toxicology* (1-20). London: Taylor & Francis.

Gomez-Lechon, M. J., O'Connor, E., Castell, J. V. and Jover, R. (2002). Sensitive markers used to identify compounds that trigger apoptosis in cultured hepatocytes. *Toxicol. Sci. 65*, 299-308.

Henkens, T., Papeleu, P., Elaut, G. et al. (2007). Trichostatin A, a critical factor in maintaining the functional differentiation of primary cultured rat hepatocytes. *Toxicol. Appl. Pharmacol. 218*, 64-71.

Herren, B., Levkau, B., Raines, E. W. and Ross, R. (1998). Cleavage of beta-catenin and plakoglobin and shedding of VE-cadherin during endothelial apoptosis: evidence for a role for caspases and metalloproteinases. *Mol. Biol. Cell 9*, 1589-1601.

Jaeschke, H., Gores, G. J., Cederbaum, A. I. et al. (2002). Mechanisms of hepatotoxicity. *Toxicol. Sci. 65*, 166-176.

Keller, S. H. and Nigam, S. K. (2003). Biochemical processing of E-cadherin under cellular stress. *Biochem. Biophys. Res. Commun. 307*, 215-223.

Kessler, T. and Muller, H. A. (2009). Cleavage of Armadillo/beta-catenin by the caspase DrICE in Drosophila apoptotic epithelial cells. *BMC Dev. Biol. 9*, 15.

Kothakota, S., Azuma, T., Reinhard, C. et al. (1997). Caspase-3-generated fragment of gelsolin: effector of morphological change in apoptosis. *Science 278*, 294-298.

Ku, N. O., Liao, J. and Omary, M. B. (1997). Apoptosis generates stable fragments of human type I keratins. *J. Biol. Chem. 272*, 33197-33203.

Malhi, H. and Gores, G. J. (2008). Cellular and molecular mechanisms of liver injury. *Gastroenterology 134*, 1641-1654.

Malhi, H., Gores, G. J. and Lemasters, J. J. (2006). Apoptosis and necrosis in the liver: a tale of two deaths? *Hepatology 43*, S31-S44.

Ni, R., Tomita, Y., Matsuda, K. et al. (1994). Fas-mediated apoptosis in primary cultured mouse hepatocytes. *Exp. Cell Res. 215*, 332-337.

Orth, K., Chinnaiyan, A. M., Garg, M. et al. (1996). The CED-3/ICE-like protease Mch2 is activated during apoptosis and cleaves the death substrate lamin A. *J. Biol. Chem. 271*, 16443-16446.

Papeleu, P., Vanhaecke, T., Henkens, T. et al. (2006). Isolation of rat hepatocytes. *Methods Mol. Biol. 320*, 229-237.

Papeleu, P., Loyer, P., Vanhaecke, T. et al. (2004). Proliferation of epidermal growth factor-stimulated hepatocytes in a hormonally defined serum-free medium. *ATLA 32*, 57-64.

Qiao, L. and Farrell, G. C. (1999). The effects of cell density, attachment substratum and dexamethasone on spontaneous apoptosis of rat hepatocytes in primary culture. *In Vitro Cell. Dev. Biol. Anim. 35*, 417-424.

Raffray, M. and Cohen, G. M. (1997). Apoptosis and necrosis in toxicology: a continuum or distinct modes of cell death? *Pharmacol. Ther. 75*, 153-177.

Rouquet, N., Carlier, K., Briand, P. et al. (1996). Multiple pathways of Fas-induced apoptosis in primary culture of hepatocytes. *Biochem. Biophys. Res. Commun. 229*, 27-35.

Schmeiser, K. and Grand, R. J. (1999). The fate of E- and P-cadherin during the early stages of apoptosis. *Cell Death Differ. 6*, 377-386.

Schmeiser, K., Hammond, E. M., Roberts, S. and Grand, R. J. (1998). Specific cleavage of gamma catenin by caspases during apoptosis. *FEBS Lett. 433*, 51-57.

Steinhusen, U., Badock, V., Bauer, A. et al. (2000). Apoptosis-induced cleavage of beta-catenin by caspase-3 results in proteolytic fragments with reduced transactivation potential. *J. Biol. Chem. 275*, 16345-16353.

Su, J. M., Wang, L. Y., Liang, Y. L. and Zha, X. L. (2005). Role of cell adhesion signal molecules in hepatocellular carcinoma cell apoptosis. *World J. Gastroenterol. 11*, 4667-4673.

Takahashi, A., Alnemri, E. S. and Lazebnik, Y. A. et al. (1996). Cleavage of lamin A by Mch2 alpha but not CPP32: multiple interleukin 1 beta-converting enzyme-related proteases with distinct substrate recognition properties are active in apoptosis. *Proc. Natl. Acad. Sci. USA 93*, 8395-8400.

Vanhaecke, T. and Rogiers, V. (2006). Hepatocyte cultures in drug metabolism and toxicological research and testing. *Methods Mol. Biol. 320*, 209-227.

Vinken, M., Decrock, E., De Vuyst, E. et al. (2009). Biochemical characterisation of an in vitro model of hepatocellular apoptotic cell death. *ATLA 37*, 209-218.

Vinken, M., Henkens, T., Snykers, S. et al. (2007). The novel histone deacetylase inhibitor 4-Me2N-BAVAH differentially affects cell junctions between primary hepatocytes. *Toxicology 236*, 92-102.

Vinken, M., Papeleu, P., Snykers, S. et al. (2006). Involvement of cell junctions in hepatocyte culture functionality. *Crit. Rev. Toxicol. 36*, 299-318.

Wheelock, M. J. and Johnson, K. R. (2003). Cadherins as modulators of cellular phenotype. *Annu. Rev. Cell. Dev. Biol. 19*, 207-235.

Acknowledgements

This work was supported by grants from the Research Council of the Vrije Universiteit Brussel (OZR-VUB), the Fund for Scientific Research Flanders (FWO-Vlaanderen) and the European Union (FP6 projects carcinoGENOMICS and LIINTOP). Special thanks go to the European Partnership on Alternatives Approaches to Animal Testing (EPAA) for awarding this work with the "Best Poster Presentation" on the occasion of the 3rd annual EPAA meeting (Brussels-Belgium, November 3, 2008) and for subsequently funding participation of the first author of this paper to the 7th World Congress on Alternatives and the Use of Animals in Life Sciences (Rome-Italy, August 30-September 3, 2009). The authors are also grateful to Mrs. Evelien De Rop for scientific collaboration and to Mrs. Sofie Wijthouck and Mr. Bart Degreef for their excellent technical assistance.

Correspondence to

Mathieu Vinken, Ph.D., Pharm.D.
Department of Toxicology, Faculty of Medicine and Pharmacy
Vrije Universiteit Brussel
Laarbeeklaan 103
1090 Brussels
Belgium
e-mail: mvinken@vub.ac.be.

Session BS12: Chemicals and pesticides

Initiatives to Decrease Redundancy in Animal Testing of Pesticides

Bennard van Ravenzwaay

Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany

Summary

Two well-documented examples of studies that can be eliminated from the data requirement list without apparent impact on the quality of risk assessment are presented. Database evaluations demonstrated no clear difference in sensitivity between 3-month and 12-month dog studies. From a regulatory perspective, only two compounds were identified in which the NOAEL for the 12-month dog study was more than 2-fold lower than in any other study. Evaluation of the contribution of the mouse cancer study demonstrated that in 5% of all cases the mouse cancer study was used for reference dose setting and in 1.5% it was used for selective cancer classification. It is suggested that regulatory agencies periodically review their study requirements for redundant studies.

Keywords: 12 month dog, carcinogenicity mice, obsolete studies

1 Introduction

In 2003 members of ILSI – HESI started a project called ACSA (agrochemical safety assessment). The purpose of this project was to evaluate how toxicity testing for agrochemicals could be done if the testing requirements were to be designed from scratch. The background of this question was the notion that our society continues to add testing requirements to an already substantial package without really looking at the overall picture, i.e. what value does each study have relative to the risk assessment and risk management process. The ACSA group was made up of distinguished academics, people from regulatory authorities and toxicologists from industry.

The results of this work were published in 2006 in a special issue of *Critical Reviews in Toxicology*. The main conclusions can be summarised as follows:

1) With the current study requirements we are focusing too much on long-term, very low exposure issues (which require substantial resources in toxicity testing) rather than on short-term or intermittent exposure scenarios.
2) Improved and early recognition of relevant toxicological effects will help to address the particular profile that a chemical may have in a more focused way.
3) Several studies are either of limited or no use (for risk assessment) or can be redesigned in such a way that the same amount of information can be obtained with far fewer resources.

An example of a redesigned study is the extended 1-generation study as an alternative to a full 2-generation study. The feasibility and sensitivity of such a study design has been positively evaluated for one compound by BASF and is under evaluation for several other compounds by other companies. With this study animal resources can be significantly reduced. However, an even more effective alternative to redesigning a study is to evaluate whether a study is useful at all, relative to risk assessment requirements. This is the true alternative to alternatives. Before a data requirement can be abandoned, however, careful examination needs to be done to assess the significance of the study for risk assessment purposes.

2 Results

There are at least two well-documented examples of studies that can be eliminated from the data requirement list without apparent impact on the quality of risk assessment. One of these

studies is the 12-month dog study. The question whether dog studies are necessary at all for risk assessment has been evaluated by several authors and their conclusion was that the dog as a non-rodent species is important in risk assessment. For about 30-50% of all investigated agrochemicals (depending on the data base), the dog was the species which provided a lower no observed adverse effect level in 90 day studies relative to the rat. Thus, the 90 day dog study provides essential information and needs to be performed.

In the past three decades it has also been necessary to perform a 12-month toxicity study in dogs for global regulatory acceptance. The large number of studies performed during this period has provided an extensive database that has been used to address the value of studies of different duration. Spielmann and Gerbracht (2001) found no clear difference in sensitivity between 3-month and 12-month studies. The distribution of the ratios between the lowest observed effect levels (LOEL) of the subchronic and chronic studies (insecticides, herbicides, fungicides) did also not show a different distribution pattern.

Doe et al. (2006) evaluated sensitivity based on NOAEL while looking at the impact on the regulatory outcome, if the 12-month dog study had not been performed. In this endeavour they looked at the lowest NOAEL of the standard set of 4 systemic toxicity studies (90-day rat, 2-year rat, 90-day dog, 1-year dog) and compared the result with and without consideration of the 12-month dog study. By the ratio of these values they found that for only two compounds the lowest NOAEL from the other three studies would be more than twice the NOAEL for the 12-month dog study. For one of these there are confounding factors with the 1-year studies. Both groups thus concluded that the 12 month dog study does not provide essential data for risk assessment.

The reason for the lack of increased sensitivity with longer duration of exposure in dogs is probably related to the total life expectancy. In rat studies the extension of exposure from 3-months to 1 year (chronic) or 2 years (cancer) takes the time of exposure relative to life expectancy from 12% to 50% or 100%. In dogs the 3-month study is about 2% of life expectancy, the 12-month study not more than 8%. In the EU the 12-month dog study is not an absolute data requirement anymore, and the US-EPA has also indicated that this study does not necessarily need to be performed. However, the world is much larger than the EU and US and we need global acceptance before a data requirement can be completely eliminated. The 12-month dog study is not necessary for risk assessment purposes. To speed up the process of acceptance of this fact, an interim solution may be suggested which should be easier to accept for those who still have doubts. In those cases in which the NOAEL/LOEL values of a 3-month rat study are lower than those obtained in the 3-month dog study – thus demonstrating the higher sensitivity of the rat in a study of similar duration – it should not be necessary to do any further dog studies.

The second study that does not contribute significantly to risk assessment is the carcinogenicity study in mice. In this study a total of 400 mice are treated for at least 18 months with a compound and are evaluated for the induction of cancer. Billington et al. (2010) evaluated a total of 195 agrochemicals with adequate cancer studies in mice for the contribution of this study type to risk assessment. With respect to the setting of a reference dose for 10 chemicals (i.e. 5% of all cases) the mouse cancer study was used. For all of these cases the NOAEL in the mouse study was close to the NOAEL obtained from other species and in 9 cases this value was between the NOAEL and LOEL of the mouse study.

The relative insensitivity of mice was also recently demonstrated in an evaluation performed by the Fraunhofer Gesellschaft within the context of an ECETOX Task Force in which the NOAELs and LOELs for chemicals were compared. One of the results of this exercise, which involved a very large database, was that on average the NOAEL/LOEL values in rats are about 2-fold lower than in mice.

The contribution of the mouse cancer study to its actual purpose, i.e. the identification of carcinogens, is even more disappointing upon detailed examination (Billington et al., 2010). Approximately 10% of the 200 agrochemicals showed mouse-specific tumours. However, most of the tumours (70%) were liver tumours. These tumours are very often associated with a very high spontaneous background incidence of certain mice strains (e.g. the B6C3F1 strain) and are not considered to be relevant for humans (if the chemical has no genotoxic properties).

Other tumour types not necessarily relevant for human risk assessment are those related to an irritating mechanism. In the end, only in 1.5% of the investigated cases the mouse carcinogenicity study resulted in a classification as "limited evidence of a carcinogenic effect".

These examples show that it is worthwhile to look at the usefulness of existing data requirements. It is not unlikely that we may find other studies that also serve very little purpose. A possible candidate for the list of redundant studies is the acute dermal LD_{50}. This study has received very little attention, and there are no extended review papers available concerning its use in risk assessment. A small analysis performed using the former ECB "classlab" database indicated that a total of 4133 chemicals are classified. A total of 66 substances were classified as "harmful" following dermal administration. Out of these 16 were not classified based on oral or inhalation toxicity, but 3 of these were corrosive, leaving 13. A total of 42 substances were classified as "toxic" following dermal administration. Out of these 17 had a lower oral or inhalation classification, but 7 were corrosive, leaving 10. Three substances were classified as "very toxic" based on the dermal LD_{50} study. Thus, the total number of selective (relevant) classifications based on LD_{50} dermal studies is = 13+10+3 = 26. In other words, less than 1% of the relevant classifications are based on the dermal LD_{50} study.

3 Conclusion

A general continuation of the three study types presented here in a "check the box" fashion is not sensible, because the data obtained with these studies are not used in risk assessment. In special cases, in which, for example, kinetic information would

indicate that the dog is more appropriate than the rat, these studies could still be performed (guidelines are available), but we should all think carefully before we perform studies that serve little purpose. This can be achieved by a regular review of the usefulness of all studies requested in standard packages, be it for agrochemical registration or for studies performed under REACH which follow a similar standard package profile. To achieve this objective, dialogue between authorities (those demanding data) and registrants (those responsible for the development of data) is essential. It is time we start this dialogue.

References

Billington, R., Lewis, R. W., Mehta, J. M. and Dewhurst, I. (2010). The mouse carcinogenicity study is no longer a scientifically justifiable core data requirement for the safety assessment of pesticides. *Crit. Rev. Toxicol. 40(1),* 35-49.

Spielmann, H. and Gerbracht, U. (2001). The use of dogs as second species in regulatory testing of pesticides. Part II: Subacute, subchronic and chronic studies in the dog. *Arch. Toxicol. 75(1),* 1-21.

Doe, J. E., Boobis, A. R., Blaker, A et al. (2006). A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. *Crit. Rev. Toxicol. 36*, 37-68.

Correspondence to

Dr. Bennard van Ravenzwaay
Experimental Toxicology and Ecology
BASF SE
67056 Ludwigshafen, Germany
e-mail: bennard.ravenzwaay@basf.com

BS13: Cosmetics

Skin Sensitisation: The Colipa Strategy for Developing and Evaluating Non-Animal Test Methods for Risk Assessment

Gavin Maxwell[1], Pierre Aeby[2], Takao Ashikaga[3], Sandrine Bessou-Touya[4], Walter Diembeck[5], Frank Gerberick[6], Petra Kern[6], Monique Marrec-Fairley[2], Jean-Marc Ovigne[7], Hitoshi Sakaguchi[8], Klaus Schroeder[9], Magali Tailhardat[10], Silvia Teissier[7] and Petra Winkler[11]

[1]Unilever, Sharnbrook, UK; [2]Colipa, Brussels, Belgium; [3]Shiseido Research Centre, Yokohama-Shi, Japan; [4]Pierre Fabre, Castres, France; [5]Beiersdorf, Hamburg, Germany; [6]The Procter & Gamble Company, Cincinnati, OH, USA and Strombeek-Bever, Belgium; [7]L'Oréal, Aulnay sous Bois, France; [8]Kao Corporation, Tochigi, Japan; [9]Henkel AG & Co, Düsseldorf, Germany; [10]LVMH Recherche, Saint Jean de Braye, France; [11]Johnson & Johnson, Neuss, Germany

Summary

Allergic contact dermatitis is a delayed-type hypersensitivity reaction induced by small reactive chemicals (haptens). Currently, the sensitising potential and potency of new chemicals is usually characterised using data generated via animal studies, such as the local lymph node assay (LLNA). There are, however, increasing public and political concerns regarding the use of animals for the testing of new chemicals. Consequently, the development of in vitro, in chemico or in silico models for predicting the sensitising potential and/or potency of new chemicals is receiving widespread interest.

The Colipa Skin Tolerance task force currently collaborates with and/or funds several academic research groups to expand our understanding of the molecular and cellular events occurring during the acquisition of skin sensitisation. Knowledge gained from this research is being used to support the development and evaluation of novel alternative approaches for the identification and characterisation of skin sensitising chemicals. At present three non-animal test methods (Direct Peptide Reactivity Assay (DPRA), Myeloid U937 Skin Sensitisation Test (MUSST) and human Cell Line Activation Test (hCLAT)) have been evaluated in Colipa interlaboratory ring trials for their potential to predict skin sensitisation potential and were recently submitted to ECVAM for formal pre-validation. Data from all three test methods will now be used to support the study and development of testing strategy approaches for skin sensitiser potency prediction.

This publication represents the current viewpoint of the cosmetics industry on the feasibility of replacing the need for animal test data for informing skin sensitisation risk assessment decisions.

Keywords: skin sensitisation, allergic contact dermatitis, animal alternatives, replacement, risk assessment

1 Introduction

Allergic contact dermatitis (ACD) is a delayed-type hypersensitivity reaction induced by small reactive chemicals (haptens) that can result from the induction of skin sensitisation (Smith and Hotchkiss, 2001). Guinea pig models have historically been used to identify whether a chemical has the potential to induce skin sensitisation in humans. More recently, a refined and reduced method, i.e. the murine local lymph node assay (LLNA), has been employed (Basketter et al., 2002). Sensitiser potency information generated using the LLNA can be used to predict a safe level of human exposure (using a Quantitative Risk Assessment (QRA) approach (Felter et al., 2002, 2003)). There are, however, increasing public and political concerns regarding the use of animal testing for the hazard characterisation of new chemicals, and consequently the development of *in vitro*, *in chemico* or *in silico* models for the replacement of animal testing is a high priority (EC, 2003; US National Research Council, 2007).

The current replacement rationale for skin sensitisation is that data from several non-animal test methods will need to be combined as part of a testing strategy to produce useful skin sensitiser potency information (Jowsey et al., 2006; Maxwell et al., 2008; Natsch et al., 2009). Each non-animal test method aims to model a key pathway involved in the induction of skin sensitisation, for example: skin exposure to the chemical and its subsequent skin bioavailability; initiation of the skin's innate immune response; chemical reactivity with skin protein; activation of epidermal (termed Langerhans cells (LC)) or dermal dendritic cells (DC); and finally, the sensitiser-specific T cell response (Basketter and Maxwell, 2007). Colipa (the European Cosmetics Association) currently collaborates with academic and industrial research groups in each of these key areas to increase and apply our fundamental knowledge of skin sensitisation to non-animal method development (Aeby et al., 2005, 2008; COLIPA, 2007). This publication aims to provide a progress report on all Colipa-funded research, method development and method evaluation activities, as well as providing a cosmetic industry viewpoint on the feasibility of replacement of animal testing for skin sensitisation.

2 Skin exposure and bioavailability

The extent and frequency of skin exposure to a sensitising chemical has long been known to play a key role in determining the potential for and severity of ACD (Basketter et al., 2006; Upadhye and Maibach, 1992). Exposure information for various product types has been generated and is routinely used to inform consumer safety risk assessments (e.g. Loretz et al., 2006, 2005, 2008). However as non-animal test methods are adopted there is an increasing need to understand skin exposure at the cellular and molecular level to enable appropriate interpretation and extrapolation of *in vitro* or *in chemico* data. Consequently, Colipa has invested in exploring the application of new technologies for the prediction of skin bioavailability and for evaluating the role of metabolism in skin sensitisation.

Toxicokinetic model for prediction of epidermal bioavailability
In order to better understand the penetration and skin disposition of new test chemicals, Prof. Jerry Kasting at the University of Cincinnati, USA has developed a toxicokinetic model to allow a more accurate prediction of the epidermal bioavailability of potential skin sensitisers. This computer model of skin absorption is based on skin physico/chemical data and aims to predict key skin parameters such as; C_{max} (the peak mid epidermal concentration of freely diffusing permeant); AUC_{120} (the area under the concentration-time profile of freely diffusing permeant 120 hours after dosing) and percentage of dose absorbed.

Skin metabolism
Understanding the metabolic competency of *in vivo* human skin and its role in the activation and clearance of sensitising chemicals is relevant for both the development and evaluation of *in vitro* test methods. Consequently Colipa is currently working with Prof. Ellen Fritsch at University of Düsseldorf, Germany and Dr Rob Edwards at Imperial College London, UK to evaluate the metabolic competency of human skin relative to *in vitro* human epidermal equivalents and immortalised keratinocyte models. The Imperial College project aims to profile the ecpression of metabolic enzymes in human skin means of immunochemistry and proteomics, while the University of Düsseldorf project focuses on the characterisation of basal and inducible metabolic enzyme activity. Initial results from the Imperial College group have already demonstrated that human skin may express far fewer cytochrome P450 enzymes than previously documented.

3 Modification of skin proteins by chemical sensitisers

Chemical sensitiser-specific T cells do not recognise the chemical itself but instead recognise a conjugate of the chemical (hapten) in the context of a protein fragment (a peptide), which is presented to T cells by antigen-presenting cells, such as LC or DC. In addition, chemical sensitisers can act either as a hapten (i.e. chemical is inherently reactive and will covalently bind to amino acid side chains), pro-hapten (chemical requires metabolic conversion to a protein-reactive species) or as pre-hapten (chemical conversion to a protein-reactive species). Colipa is supporting the development of next generation peptide binding test methods in addition to funding research to explore the correlation between sensitisation potential and chemical reactivity.

Direct Peptide Reactivity Assay (DPRA)
The DPRA protocol was developed by Dr. F. Gerberick at Procter & Gamble, Cincinnati, USA in collaboration with Prof. J. P. Lepoittevin at the University of Strasbourg, France (Gerberick et al., 2007). The test method aims to model protein haptenation in chemico by measuring the depletion of two synthetic peptides (containing either a single cysteine or lysine side-chain as a reaction target) using High Performance Liquid Chromatography (HPLC). The prediction model for the DPRA uses a classification tree approach to assign different reactivity categories to the depletion data from the cysteine and lysine synthetic peptides (Gerberick et al., 2007). Using this approach the DPRA achieved 89% concordance with LLNA potency categories (i.e. weak, moderate, etc.), and following two rounds of ring trials interlaboratory reproducibility was also found to be robust. Consequently, the current protocol has been transferred to the European Centre for the Validation of Alternative Methods (ECVAM) for phase 3 pre-validation (Hartung et al., 2004).

Next generation peptide binding assay
Colipa continues to support the ongoing collaboration between Procter & Gamble and the University of Strasbourg, which is currently focussed on increasing the applicability domain of the current test method to include poorly soluble chemicals and prohaptens that require oxidation to become active.

4 Skin innate immune response

Keratinocytes are the predominant epidermal cell type and are known to play a key role in the activation of LC in response to a

pathogenic stimulus or physical/chemical stress (Cumberbatch et al., 1992; Griffiths et al., 2005). Although keratinocyte activation in isolation is unlikely to be directly predictive of skin sensitisation hazard, Colipa remains interested in the role that epidermal keratinocytes and/or dermal fibroblasts could play in the activation of LC and dermal DC. Consequently a collaboration was recently, initiated with Dr A. Ghaem Maghami at University of Nottingham, UK and Dr. J. Haycock at University of Sheffield, UK, to aims to establish an immunocompetent skin equivalent model using keratinocytes, fibroblasts and DC.

5 Sensitiser-induced dendritic cell activation

The next key step in the induction of the skin sensitisation process is the uptake and processing of the haptenated skin proteins by immature DCs in response to the inflammatory signals released by activated epidermal keratinocytes and dermal fibroblasts (Cumberbatch et al.,1992; Griffiths et al., 2005). During this process DCs mature to an activated state, which can be measured experimentally by assessing changes in cell surface receptor expression (e.g. co-stimulatory molecules such as CD83, CD86 or adhesion molecules such as CD54) and cytokine secretion (e.g. IL-1β, IL-8) (Basketter and Maxwell, 2007). Advances in the generation of immature DCs *in vitro* from human peripheral blood monocytes and the availability of cell lines with DC-like phenotypes have led to the development of several non-animal test methods based upon the measurement of DC activation biomarkers following exposure to the test chemical.

Myeloid U937 Skin Sensitisation Test (MUSST)

Drawbacks of peripheral blood derived DCs are their complex preparation procedures, sourcing issues and their inherent donor-to-donor variability. As a possible alternative, human myeloid leukaemia cell lines represent good candidates as DC surrogates. Protocols based on the U937 cell line were developed by L'Oréal (Ade et al., 2006) and Cosmital SA, Procter & Gamble (Python et al., 2007). The MUSST protocol is based upon the following steps (Fig. 1): U937 cells are exposed for 48 h in 96 well plates to a range of chemical concentrations. CD86 expression (FITC coupled monoclonal antibody) and cell viability (propidium iodide exclusion) are then measured by flow cytometry. A chemical is predicted as a sensitiser if it induces a dose-dependent increase of CD86 expression at non-toxic doses (viability >70%) within 2 independent experiments. Historically, the MUSST's concordance with yes/no predictions of sensitisation potential derived from human clinical data has been determined by L'Oréal to be 85% (based on 99 chemicals), and following two rounds of Colipa ring trials the inter-laboratory reproducibility of the MUSST was found to be robust. Consequently the current MUSST protocol has been transferred to ECVAM for phase 3 pre-validation.

Human Cell Line Activation Test (h-CLAT)

The h-CLAT was developed in a collaboration between Kao (Sakaguchi et al., 2006) and Shiseido (Ashikaga et al., 2006). The outline of the h-CLAT protocol (Fig. 2) is as follows: THP-1 cells (an acute monocytic leukaemia cell line (Tsuchiya et al., 1980)) are pre-cultured for 48 h or 72 h and are treated with the test chemical for 24 h. After removing the test chemical, the expression of two cell surface antigens, CD86 and CD54, is measured by specific antibody staining and subsequent detection by flow cytometry. Two of three independent measurements at any dose should exceed the positive criteria (CD86 >150% or CD54 >200%) in order to be judged as positive. The concordance with yes/no predictions of sensitisation potential derived from

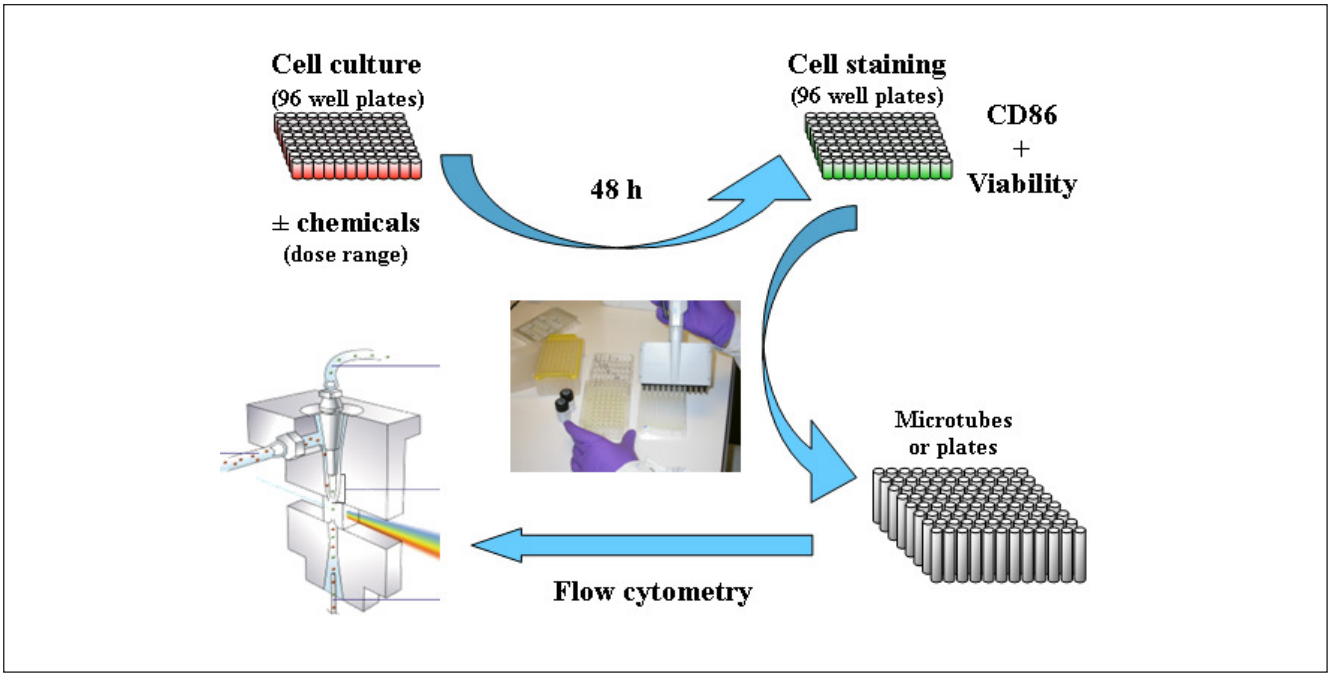


Fig. 1: Overview of MUSST protocol

LLNA data was found to be approximately 85% (Sakaguchi et al., 2006, 2009), and following two parallel sets of ring trials involving Colipa members and Japanese cosmetic companies the h-CLAT inter-laboratory reproducibility was found to be acceptable. Consequently the current h-CLAT protocol has been transferred to ECVAM for phase 3 pre-validation, in parallel with the MUSST and DPRA.

6 New biomarkers for DC activation

Intracellular signalling pathways

The early intracellular signalling pathways involved in sensitiser-induced DC activation and maturation are currently not well understood. To address this knowledge gap, Colipa recently collaborated with Prof. M. Serres at University of Claude Bernard in Lyon, France where DCs derived from human peripheral blood monocytes (MoDC), U937 and h-CLAT were treated with a large panel of sensitisers and non-sensitisers prior to ERK 1/2 and p38 MAP kinase (MAPK) activity being measured by the Face™ method (Trompezinski et al., 2008). The results of this study demonstrated that all sensitisers markedly activated p38 MAPK and inhibited Erk1/2 in MoDC, many of them via a predominant role of oxidative stress as previously reported. Similar data were also obtained in the U937 cell line, however the THP-1 cell line was found to respond differently, with Erk1/2 pathways becoming activated (rather than inhibited) following treatment with sensitiser. In all studies non-sensitisers had no significant impact on either p38 MAPK or Erk 1/2 kinase activity. Discussions are ongoing with respect to how best to apply these pathways as predictive biomarkers in existing or new DC activation test methods.

Gene expression changes

In order to exploit new predictive biomarkers for sensitiser-induced DC activation, Colipa has sponsored work at Procter & Gamble (Cincinnati, USA) to develop a high throughput method for the measurement of DC gene expression changes previously identified by Affymetrix GeneChip® analysis of sensitiser-treated MoDC. This analysis revealed 173 genes that were significantly modulated (Ryan et al., 2004), and 29 of these genes were selected for further evaluation of predictive potential using RT PCR (Gildea et al., 2006). In order to develop a high throughput method the Luminex® xMAP® technology was employed for concurrent measurement of gene expression changes, and MUTZ-3, THP-1 and U937 cell lines were all evaluated as potential surrogates for the MoDC model. All models were assessed for expression changes in all 29 genes in response to 5 sensitisers and 3 non-sensitisers, however the conclusion of the work was that, due to a lack of sensitivity, none of the cell lines was an appropriate replacement for the MoDC model.

In addition, Colipa has recently identified two projects for funding that aim to develop recombinant DC models capable of fluorescing in response to sensitiser treatment; the first project is a collaboration with Prof. A. Takashima in University of Toledo, USA that aims to develop a multi-parameter DC biosensor system for detection of skin sensitisers through transfection with a fluorescent protein or luciferase reporter construct using promoters of literature- or experimentally-identified gene biomarkers. The second project focuses on the development of a DC migration reporter model based upon coupling a luciferase construct to gene promoters known to control DC migration from the skin to the lymph nodes with Dr J. Pease at Imperial College London, UK.

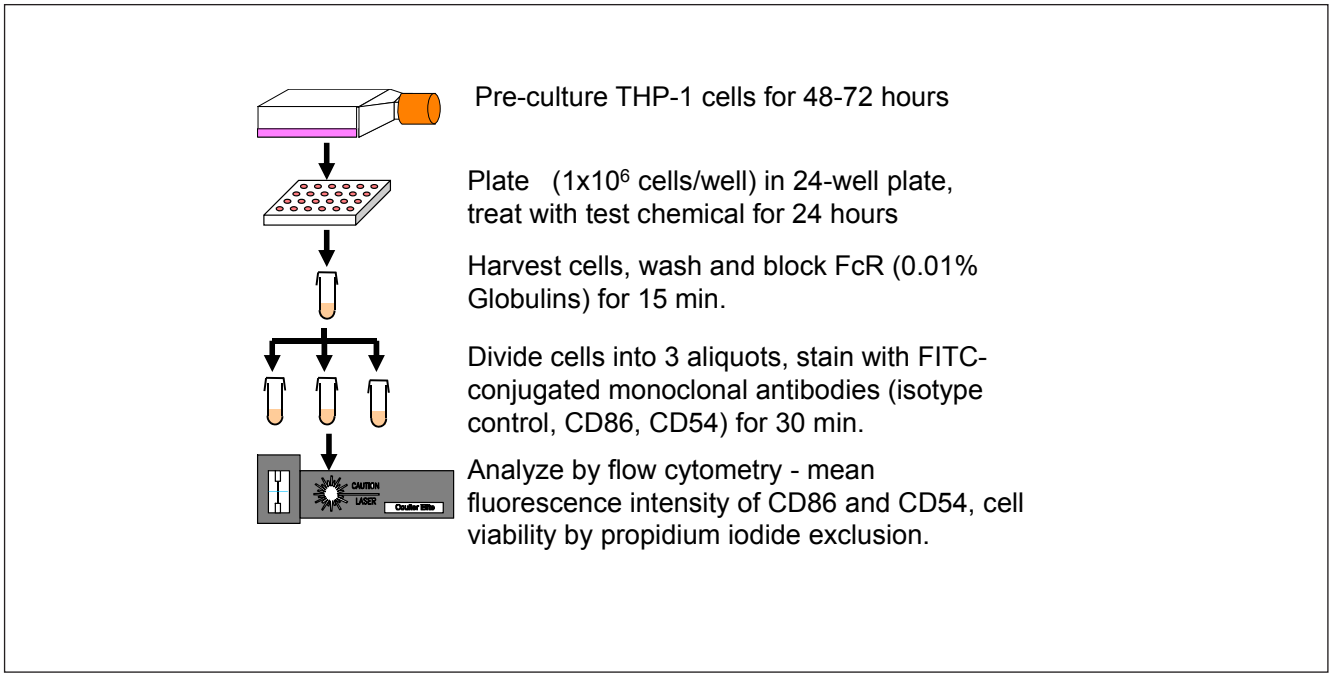


Fig. 2: Overview of h-CLAT protocol

7 Sensitiser-specific T cell responses

Despite sensitiser-specific T cell proliferation representing the hypothesised key pathway in the induction of skin sensitisation (Jowsey et al., 2006; Maxwell and MacKay, 2008), *in vitro* T cell test methods have proven difficult to apply to chemical-induced skin sensitisation due to their low sensitivity for chemical allergens, the difficulty of *ex vivo* DC: T cell co-culture, the complexity of the T cell repertoire and the inherent genetic variability within human populations (Guironnet et al., 2000; Hauser, 1990; Rougier et al., 2000). Consequently a test method capable of generating predictive information on sensitiser-induced T cell responses remains a key gap at present, and Colipa is keen to identify new ways of resolving these issues.

T cell priming assay (TCPA)

For the past two years Colipa has collaborated with Prof. J. F. Nicolas at INSERM in Lyon, France to develop a human DC: T cell co-culture approach based on a protocol that was originally developed using murine tissues (Vocanson et al., 2006). The TCPA protocol is based around co-culturing of sensitiser-haptenated, irradiated MoDC with T cells that have been depleted for the $CD25^+$ regulatory T cell subset. The rationale, which proved successful in the mouse system, is that removal of the regulatory T cell fraction will lower the threshold at which a T cell proliferative response can occur. In so doing, the sensitivity of the test method for weak and moderate sensitisers increases relative to standard DC: T cell co-culture protocols. However initial results, although demonstrating that the basic protocol can be adapted to utilise human tissues, suggest that reproducibility may still be an issue.

8 Discussion

To conclude, it is worth sharing the conclusions of the recent Colipa Skin Tolerance conference and Colipa "Testing Strategies for Skin Sensitisation" workshop (held on 23rd-24th June 09 and 25th March 09, respectively), as these discussions offer perspective on the remaining challenges. For test method development, the continuing lack of a robust, reproducible *in vitro* method for predicting the sensitiser-specific T cell responses is seen as a key gap. For testing strategy development, it is felt that the lack of connectivity between different skin sensitisation biomarkers and the absence of an obvious gold standard dataset for human sensitiser potency will make establishing the clinical relevance of *in vitro* or *in chemico* biomarkers particularly challenging. In addition, the need for flexibility in testing strategy development, to accommodate new test methods and bespoke solutions for particular ingredients or exposure scenarios, represents another criterion to be accommodated. Despite these gaps, the replacement of the need for animal testing for skin sensitisation risk assessment is viewed as ultimately achievable, and the next couple of years should set the timeline for this achievement.

References

Ade, N., Martinozzi-Teissier, S., Pallardy, M. and Rousset, F. (2006). Activation of U937 cells by contact sensitizers: CD86 expression is independent of apoptosis. *J. Immunotoxicol. 3*, 189-197.

Aeby, P., Ashikaga, T., Diembeck, W. et al. (2008). The COLIPA Strategy for the Development of In Vitro Alternatives: Skin Sensitisation. *AATEX 14, Spec. Issue*, 375-379.

Aeby, P., Basketter, D. A., Diembeck, W. et al. (2005). Dendritic Cell Research Projects, Proceedings 5th World Congress 2005, *ALTEX 23, Spec. Issue*, 226-228.

Basketter, D. and Maxwell, G. (2007). In vitro approaches to the identification and characterization of skin sensitizers. *Cutan. Ocul. Toxicol. 26*, 359-373.

Basketter, D. A., Jefferies, D., Safford, B. J. et al. (2006). The impact of exposure variables on the induction of skin sensitization. *Contact Dermatitis 55*, 178-185.

Basketter, D. A., Evans, P., Fielder, R. J. et al. (2002). Local lymph node assay - validation, conduct and use in practice. *Food and Chemical Toxicology 40*, 593-598.

COLIPA (2007). Colipa, Serious About Cosmetics - Serious about Alternative Methods. http://www. colipa. com/site/index. cfm?SID=15588&OBJ=20702&back=1.

Cumberbatch, M., Gould, S. J., Peters, S. W. et al. (1992). Langerhans Cells, Antigen Presentation, and the Diversity of Responses to Chemical Allergens. *Journal of Investigative Dermatology 99*, 107-S108.

European Commission (2003). Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products. *Official Journal of the European Union 2003. L66*; 26-35.

Felter, S. P., Ryan, C. A., Basketter, D. A. et al. (2003). Application of the risk assessment paradigm to the induction of allergic contact dermatitis. *Regulatory Toxicology and Pharmacology 37*, 1-10.

Felter, S. P., Robinson, M. K., Basketter, D. A. and Gerberick, G. F. (2002). A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. *Contact Dermatitis 47*, 257-266.

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M. et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree lmodel approach. *Toxicological Sciences 97*, 417-427.

Gildea, L. A., Ryan, C. A., Foertsch, L. M. et al. (2006). Identification of gene expression changes induced by chemical allergens in dendritic cells: opportunities for skin sensitization testing. *J. Invest Dermatol. 126*, 1813-1822.

Griffiths, C. E., Dearman, R. J., Cumberbatch, M. and Kimber, I. (2005). Cytokines and Langerhans cell mobilisation in mouse and man. *Cytokine. 32*, 67-70.

Guironnet, G., Dalbiez-Gauthier, C., Rousset, F. et al. (2000). In vitro human T cell sensitization to haptens by monocyte-derived dendritic cells. *Toxicology in Vitro 14*, 517-522.

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular ap-

proach to the ECVAM principles on test validity. *Altern. Lab Anim 32*, 467-472.

Hauser, C. (1990). Cultured epidermal Langerhans cells activate effector T cells for contact sensitivity 6. *J Invest Dermatol. 95*, 436-440.

Jowsey, I. R., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). A future approach to measuring relative skin sensitising potency: a proposal. *Journal of Applied Toxicology 26*, 341-350.

Loretz, L. J., Api, A. M., Babcock, L. et al. (2008). Exposure data for cosmetic products: Facial cleanser, hair conditioner, and eye shadow. *Food and Chemical Toxicology*. In press.

Loretz, L., Api, A. M., Barraj, L. et al. (2006). Exposure data for personal care products: hairspray, spray perfume, liquid foundation, shampoo, body wash, and solid antiperspirant. *Food and Chemical Toxicology 44*, 2008-2018.

Loretz, L. J., Api, A. M., Barraj, L. M. et al. (2005). Exposure data for cosmetic products: lipstick, body lotion, and face cream. *Food and Chemical Toxicology 43*, 279-291.

Maxwell, G., Aleksic, M., Aptula, A. et al. (2008). Assuring consumer safety without animal testing: a feasibility case study for skin sensitisation. *Altern. Lab. Anim. 36*, 557-568.

Maxwell, G. and MacKay, C. (2008). Application of a systems biology approach to skin allergy risk assessment. *Altern. Lab. Anim. 36*, 521-556.

Natsch, A., Emter, R. and Ellis, G. (2009). Filling the concept with data: integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicol. Sci. 107*, 106-121.

Python, F., Goebel, C. and Aeby, P. (2007). Assessment of the U937 cell line for the detection of contact allergens. *Toxicology and Applied Pharmacology 220*, 113-124.

Rougier, N., Redziniak, G., Mougin, D. et al. (2000). In vitro evaluation of the sensitization potential of weak contact allergens using langerhans-like dendritic cells and autologous T cells. *Toxicology 145*, 73-82.

Ryan, C. A., Gildea, L. A., Hulette, B. C. et al. (2004). Gene expression changes in peripheral blood-derived dendritic cells following exposure to a contact allergen. *6. Toxicology Letters 150*, 301-316.

Sakaguchi, H., Ashikaga, T., Miyazawa, M. et al. (2009). The relationship between CD86/CD54 expression and THP-1 cell viability in an in vitro skin sensitization test--human cell line activation test (h-CLAT). *Cell Biol. Toxicol. 25*, 109-126.

Sakaguchi, H., Ashikaga, T., Miyazawa, M. et al. (2006). Development of an in vitro skin sensitization test using human cell lines; human Cell Line Activation Test (h-CLAT). II. An inter-laboratory study of the h-CLAT. *Toxicology in Vitro 20*, 774-784.

Sens-it-iv (2006). Novel Testing Strategies for In Vitro Assessment of Allergens - project summary. http://www. sens-it-iv. eu/files/start/Sens-it-iv_Summary. pdf.

Smith, C. K. and Hotchkiss, S. A. (2001). *Allergic Contact Dermatitis*. London, UK: Taylor & Francis.

Trompezinski, S., Migdal, C., Tailhardat, M. et al. (2008). Characterization of early events involved in human dendritic cell maturation induced by sensitizers: cross talk between MAPK signalling pathways. *Toxicol. Appl. Pharmacol. 230*, 397-406.

Tsuchiya, S., Yamabe, M., Yamaguchi, Y. et al. (1980). Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int. J. Cancer 26*, 171-176.

Upadhye, M. R. and Maibach, H. I. (1992). Influence of area of application of allergen on sensitization in contact dermatitis. *Contact Dermatitis 27*, 281-286.

US National Research Council (2007). *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. Committee on Toxicity and Assessment of Environmental Agents* National Research Council, Washington, DC.

Vocanson, M., Hennino, A., Cluzel-Tailhardat, M. et al. (2006). CD8+ T cells are effector cells of contact dermatitis to common skin allergens in mice. *J. Invest Dermatol. 126*, 815-820.

Acknowledgements

In our effort to develop in vitro skin sensitisation approaches, Colipa is closely collaborating with ECVAM (Hartung et al., 2004) and the research consortium entitled "Novel Testing Strategies for *In Vitro* Assessment of Allergens" (Acronym: Sens-it-iv) sponsored by the European Union Framework Program 6 (Sens-it-iv, 2006).

Correspondence to

Gavin Maxwell
Safety and Environmental Assurance Centre
Unilever Colworth
Sharnbrook
Bedford MK44 1LQ, UK
e-mail: Gavin.maxwell@unilever.com

BS19: Education and training

Education and Training in the 3Rs

Jan van der Valk[1], Marilyn Brown[2], Clément Gauthier[3], Brian Howard[4], Tsutomu Miki Kurosawa[5], Margaret Rose[6] and Bert van Zutphen[7]

[1]Netherlands Centre for Alternatives to animal experiments, Utrecht University, The Netherlands; [2]Animal Welfare and Training Charles River, East Thetford, USA; [3]Canadian Council on Animal Care, Ottawa, Canada; [4]Director of Animal Welfare, The University of Sheffield, UK; [5]The Institute of Experimental Animal Sciences, Osaka University Medical School, Yamadaoka, Suita-shi, Osaka, Japan; [6]University of New South Wales, Sydney, Australia; [7]Utrecht University, The Netherlands

Summary

To discuss education and training of persons involved in the design of animal experiments, a workshop was organised during the 7th World Congress on Alternatives & Animal Use. Speakers from Canada, Australia, Japan, United States and Europe gave an overview of education and training requirements and courses in their parts of the world.

During the subsequent discussion with the audience it was concluded that for every study involving animals there should be at least one person who is finally responsible. This person should have been educated in one of the biomedical areas and have completed a laboratory animal science course (LAS). Attitude training was regarded as the most important part of such a course. Furthermore, the content, length, skills training and harmonisation of LAS courses were discussed.

The final conclusion was that education on LAS and the 3Rs not only contributes to reducing animal numbers and animal suffering, but also to better science.

Keywords: education, 3Rs, laboratory animal science, training, guidelines

1 Introduction

When the use of animals in research is inevitable, it is of utmost importance that no more animals are used than strictly needed and that every possible measure to reduce suffering (refinement) is applied. Ideally, all persons involved in the use and care of laboratory animals must be specifically educated and trained. The quality of research and the welfare of the laboratory animal greatly depend on their competence. Implementation of the 3Rs not only depends on knowledge and skills but also on an attitude that is based on respect for animals.

In countries with legislation on the protection of animals used for scientific purposes the law often includes a section on competence of the institutes and persons dealing with experimental animals. But also in countries without such legislative requirements, initiatives have been taken to promote the education of these persons. However, at present there are still major differences between countries (and within countries between institutes) with regard to the programmes that must be taken in order to comply with the requirements. This is mainly due to the fact that only few of the existing regulations and guidelines are specific on length and/or depth of the courses for each of the categories of persons involved in animal experiments and care.

Here we will report on the results of a workshop, held during the 7th World Congress on Alternatives & Animals in the Life Sciences, on the topic of education and training of the category of persons who are responsible for the design of animal experiments (the scientist). The report presents an overview of the presentations of invited speakers, mainly dealing with initiatives that have been taken in several parts of the world (first part) and a summary of the discussion between the panel members and with the audience (second part).

2 Canada (Clément Gauthier)

The Canadian Council on Animal Care (CCAC) is the peer-based organisation overseeing the ethical care and use of animals in research, teaching and testing throughout Canada since 1968. Institutional Animal Care Committees (ACCs) pioneered by the CCAC are the keystone of the Canadian oversight system. ACCs act as local quality control structures responsible for informed decision-making based on science and societal values, while the CCAC provides quality assurance at the national level as quasi-regulatory body.

Adequate training for all personnel is an essential component of any institutional animal care and use program to ensure that animals are used in the most humane and ethical manner. The *CCAC guidelines on: institutional animal user training* was published in 1999 to present theoretical and practical training requirements for animal users including investigators, study directors, post-doctoral fellows, research team members (including veterinarians if involved in research) and graduate students. An accompanying *Recommended Syllabus* indicates the core topics to be covered. Twelve web-based modules on the Core Topics of the *Recommended Syllabus* were posted on the CCAC website (www.ccac.ca) with other resources prior to the mandatory implementation of the *CCAC guidelines on: institutional animal user training* through the CCAC Assessment Program, beginning in 2003. The requirements covered in the *Recommended Syllabus* and the related training modules are similar to those included in FELASA Categories C and B. The theoretical part covered in the twelve CCAC training modules can be taken in about 20 hours. However, the hands-on part is as extensive as the specialised research mandate of the institution and the content of the protocol itself require.

Through its overarching *CCAC policy statement on: ethics of animal investigation* (1989), the CCAC has incorporated adherence to the Three Rs principles of Russell and Burch (1959) as the fundamental basis for the ethical oversight of animal care and use in Canada. This ethical tenet is the third core topic required to be covered in any institutional animal care and use program. In addition to the current CCAC training module on the Three Rs, further training material has been developed under the new Three Rs Program initiated by the CCAC in 2008, namely a microsite on the Three Rs which includes information on Replacement, Reduction, Refinement alternatives and a Three Rs search strategy for investigators.

While the *CCAC guidelines on: institutional animal user training* (1999) covered primarily conventional laboratory animals, additional training material on wildlife and fish has been subsequently posted on the CCAC website, and training material on the use of farm animals is in preparation for posting in 2010.

The CCAC assesses institutional training programs as part of its certification of complete institutional animal care and use programs. However, the training and examination of individuals' competencies is the responsibility of the institution and its ACC as per the *CCAC guidelines on: institutional animal user training* (1999). Accordingly, while there can be agreement on general training principles and course contents through initiatives such as the one undertaken recently by the International Council for Laboratory Animal Science (ICLAS), reciprocal agreements regarding specialised competencies of individuals have to be achieved at the level of institutions.

3 Australia (Margaret Rose)

The *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes* (NHMRC, 2004) governs any use of animals for research, teaching or product testing in the fields of medicine, biology, agriculture, veterinary, environmental or animal sciences. The Code details an ethical framework for deciding if and how animals can be used in these circumstances and provides principles to guide such decisions notably with regard to justification and the critical application of the 3Rs.

A fundamental principle of the Code is that scientists (researchers and teachers) have direct and ultimate responsibility for all aspects of the welfare of those animals they use, emphasising that this responsibility is embedded in the notion of a duty of care. The framework for ethical review and the arrangements for responsible use of animals are intricately linked such that ethical practice is integral to day to day decisions and activities.

An education program that promotes awareness of the issues informs the critical application of principles of the Code and provides the skills and knowledge to support the engagement of scientists that is essential to achieve its goals. Thus, through various educational activities, the Code requires institutions to ensure that scientists are aware of their responsibilities and have the knowledge and skills to undertake specific projects. In the case of students undertaking research training, there is a specific requirement that they receive instruction in their ethical and legal responsibilities as well as in the appropriate methods for animal care and use.

In most institutions it is mandatory that staff attend a course prior to working with animals but, given the diversity of scientific activities, institutions have developed programs customised to their particular needs. Courses range from intensive 2-3 day programs as a general introduction to the ethical and scientific issues to specific programs for wildlife researchers. Course content covers the scope of responsibilities of scientists, the role of the animal ethics committee (AEC) and the relationship between the scientist and the AEC, planning experiments emphasising the 3Rs, monitoring animal wellbeing, the management of pain and distress, animal models, research procedures and record keeping. The relationship between animal welfare and scientific outcomes is a core element of the Code and is highlighted in the course content. Key learning outcomes sought include an awareness of the range of ethical views and an understanding of the responsibilities of a scientist, the process for ethical review, the importance of animal welfare, the scientific basis for the application of the 3Rs and strategies to identify and manage pain and distress. Practical sessions to enable staff to develop confidence in handling animals as well as undertaking specific procedures are offered regularly. These sessions may involve the use of electronic materials that demonstrate appropriate handling

techniques, research methods and species-specific behaviours, including pain-related behaviours, as well as manikins to develop manual skills such as venipuncture or suturing methods.

To date we have not developed a national curriculum for the education of scientists, although this is currently under discussion and would draw on the experiences of extant courses. One consideration is the method of delivery for such courses, noting the need to consider teaching strategies to support adult learning (Dobrovolny et al., 2007). The availability of online course material is seen as being of potential value not only to enhance access to information but also as a way of sharing resources and expertise. But face-to-face discussion also is seen to be important to promote reflective practice and thus underpin the goals of the Code. Further, the publication of evidence-based guidelines, such as those recently published to promote the wellbeing of animals (NHMRC, 2008), provide important resources to support the delivery of courses emphasising the link between animal welfare and scientific outcomes.

In these educational activities, the Replacement of animals is emphasised and informed in a number of ways. Foremost, a heightened awareness of his or her responsibilities on the part of the scientist will support reflective practice and so inform consideration of opportunities to implement the 3Rs (Lloyd, 2009). Another important strategy is to highlight Replacement as the default position when planning a project, critically evaluating the need to use animals to achieve all or part of the scientific aims. Evidence of this approach is seen in the wellbeing guidelines mentioned above, in guidelines concerning specific procedures, such as the production of monoclonal antibodies and in the development of case studies such as those published as part of an on-line course. Further, the use of "local" examples of how animals have been replaced in certain studies provides a tangible context within which to discuss potential opportunities; this being reinforced by the recent establishment of a national prize for alternatives. Knowledge of and access to resources such as websites that provide information about alternatives are also important. Such information is highlighted on the website *Animal Ethics Infolink* (www.animalethics.org.au) that has been developed as an information resource for scientists and AEC members in Australia.

4 Japan (Tsutomu Miki Kurosawa)

All investigators involved in animal experiments, including undergraduate and postgraduate students, are required to have completed education and training as regulated by the notice of Ministry of Education and the Standards Relating to the Care and Management of Laboratory Animals and Relief of Pain in Japan. The training course is established by the IACUC (Institutional Animal Care and Use Committee), and in most institutions a two hour course is presented. In Osaka University Medical School 1,600 participants take part in this course every year. During the course the 3Rs are specifically emphasised, as stated in the Law for the Humane Treatment and Management of Animals (Scientific use of animals). Apart from the formal course, practical training is provided in many institutions. In the practical training of the Graduate School of Medicine at Osaka University, supporting video programmes are used. Anaesthesia is considered most important for refinement and a 30 minute video on this subject is screened. Training includes proper animal handling and minor and major surgery with aseptic techniques. Also, post-surgical care and analgesia are emphasised. So far, the mouse is shown as a model.

Accreditation systems for laboratory animal research institutions are established by three different national organisations, namely Center for Accreditation of Laboratory Animal Care (HS foundation) for the institutions governed by the Ministry of Health, the Japanese Association of Laboratory Animal Facilities of National University Corporations for universities and animal breeding facilities and AAALAC International (Association for Assessment and Accreditation of Laboratory Animal Care). Six institutions are now accredited by AAALAC International.

The personal qualification system for junior and senior laboratory animal technicians and educators is established by the Japanese Society for Laboratory Animal Resources. The Japanese College of Laboratory Animal Medicine, which is a member of the International Association of Colleges of Laboratory Animal Medicine (IACLAM), has certified more than 70 diplomats. IACLAM is planning the international harmonisation of diplomats and their certification system. In practice, these professionals are already mutually recognised. Also in other Asian countries, accreditation by AAALAC International is becoming popular. Through the accreditation process of AAALAC International, the importance of education and training is receiving more attention, and the quality and harmonisation of the contents of education will be more realistic in the future. Also the OIE (World Organisation of Animal Health) Animal Welfare Code will soon be introduced, and this international standard will prove to be effective in improving laboratory animal welfare in Asian countries.

In Japan, international harmonisation of education and training is still rarely discussed.

The Working Group for laboratory animal welfare in OIE is currently discussing the training of veterinarians in laboratory animal medicine as one of the three priority areas. Other topics are laboratory animal transport, regulatory testing and the adoption of alternatives.

5 Europe (Bryan Howard)

European Directive 86/609 (European Council, 1986) recognises that the skills required of a competent researcher extend beyond the core scientific needs of the relevant academic discipline. There is a perceived increasing concern for wider implementation of the 3Rs within the European community, and the current proposal for revision of the Directive (European Parliament, 2009) urges an extension of the definition of competence to include better matching of scientific need with the impact of science on animals. For example, Article 7 of the Directive 86/609 stipulates that "Experiments shall be performed solely by competent authorised persons, or under the direct responsi-

bility of such a person". Competence involves a combination of knowledge, skills, understanding and attitudes. Knowledge is the easiest of these to acquire – attending lectures, reading books, internet sources, etc. The development of understanding usually involves interaction with situations or other people and is best achieved in simulated or real practical situations. Skills and attitudes are something else, and are often acquired by working alongside a suitable role model, although a basic respect for animals and awareness of the theoretical basis of practical skills should be acquired beforehand. Although delivery of rounded competence within structured training programmes may be problematic, key elements can be developed alongside a commitment to lifelong learning.

European Member States have traditionally adopted national approaches to delivering training, but the Federation of Laboratory Animal Science Associations (FELASA) set out to recommend a formal structure for developing such competencies and established a series of working groups, one of which (FELASA, 1995) proposed a uniform educational approach to providing a foundation for the responsible use of animals. FELASA category C training is intended for those designing and directing animal experiments. Prior experience in laboratory animal science can provide a springboard for learning, and hence FELASA's stipulation of a Bachelor or MSc degree in an appropriate biological discipline as a prerequisite. The proposal comprised a course lasting approximately 80 h, involving both theoretical and practical instruction, with provision for alternative modes of study. It included the following topics:

– Biology and husbandry of laboratory animals
– Microbiology and disease
– Health hazards and safe practices in the animal house
– Design and conduct of animal experiments
– Anaesthesia, analgesia and experimental procedures
– Alternatives to animal use
– Ethical aspects and legislation
– Analysis of scientific literature

The Category C syllabus emphasises the importance of the 3Rs in the planning, design and conduct of scientific experiments – courses usually include practical guidance on searching for alternatives, reducing the numbers of animals used by appropriate experimental design and refinement by advising on how to conduct experimental procedures including anaesthesia so as to have minimal impact on the animals' well-being.

FELASA subsequently introduced an accreditation scheme which assures the quality of education and training, promotes responsible and high quality science, facilitates free movement of personnel between countries and assists with further harmonisation (FELASA 2002). This scheme further promotes refinement by insistence on low ratios of students to tutor in practical classes, scrutiny of the content and effectiveness of training and assessment, and reduction by promoting the use of replacement strategies in training and avoiding unnecessary use of animals because of a need to repeat training. Category C training has proven to be very popular, and many establishments in Europe have adopted it as the standard for training laboratory animal scientists.

6 United States (Marilyn Brown)

Laboratory animal use in the United States is primarily governed by two regulations, the Animal Welfare Regulations (AWRs, enforced by the US Department of Agriculture – USDA, CRR, 1985) and the Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS) Policy (PHS, 1996), which is overseen by the Office of Laboratory Animal Welfare (OLAW) of the PHS. In addition, most major users of laboratory animals are involved in the voluntary accreditation program of the Association for the Assessment and Accreditation of Laboratory Animal Care, International (AAALAC, Intl.). Both AAALAC and the PHS require compliance with the Guide for the Care and Use of the Laboratory Animals – the Guide (ILAR, 1996).

The AWRs address the training of individuals involved in animal research. The AWRs require that "Personnel conducting procedures on the species being maintained or studied will be appropriately qualified and trained in those procedures." Section 2.32 of the AWRs specifically addresses personnel qualifications. It is the institution's responsibility to ensure training, and this responsibility is "fulfilled in part through the provision of training and instruction…" The AWRs go further and describe the general areas that must be included: humane methods of animal maintenance and experimentation including: basic needs of each species; proper handling and care; proper pre and post procedural care; aseptic surgical methods; the concept of the 3Rs; proper use of anaesthetics and analgesics; methods to report deficiencies in animal care and use; and how to provide information on alternatives. Training is one of the areas that are evaluated as part of unannounced USDA inspections which occur at least annually.

As previously mentioned, both the PHS and AAALAC expectations are based on the *Guide*, which also has a section on personnel qualifications and training. However, details of researcher training are not provided, except to indicate that they must comply with regulations. Several areas in the Program Description, which is the self assessment document serving as the basis for the AAALAC site visit, ask for details of training at the accredited institution. When conducting site visits, AAALAC site visitors will often review training records and, through observation of activities and questioning research personnel, make an assessment of the adequacy of training at the institution. AAALAC categories site visit findings as either a mandatory item (one that must be corrected to obtain or maintain accreditation) or a suggestion for improvement (something that the Council on Accreditation believes would further improve a satisfactory program). Of approximately 1000 site visits reviewed, 2.8% of sites had mandatory findings related to training. A majority of these (1.9%) were related to occupational health and safety. Suggestions for improvement of training occurred at 4.9% of the institutions with, again, the majority (3.9%) related to occupational health and safety. A review of these data indicates that 97.5% of accredited institutions had NO deficiencies in their training programs which would impact on animal welfare in their training programs. All of the top 100 funded academic institutions and virtually all of the major pharmaceutical and contract research organisations in the U.S. are

AAALAC accredited. These institutions would represent a sizable majority of all vertebrate animals used in research.

Further directions are given in "Education and Training in the Care and Use of Laboratory Animals: A Guide for Developing Institutional Programs" (ILAR, 1991), which is a manual on training that provides additional specific guidance on development of training programs. It is published by the Institute of Laboratory Animal Resources (ILAR) of the National Academy of Science and provides additional specific guidance on development of training programs. The basis for such programs includes a list of subjects that should be included in core material and additional modules that are generally provided on an "as needed" basis. In addition, a full issue of the ILAR Journal was devoted to Training and Adult Learning Strategies for the Care and Use of Laboratory Animals (ILAR, 2007).

In the US, researcher training is not formalised at a national level as it is in some other regions. Using more of a performance based approach, researcher training is individualised at both the institutional level and the individual researcher level, based upon what species and procedures are involved. While this may seem to be a weakness by those who use other systems, the success of training (competency) is regularly assessed externally by regulators and site visitors and internally through biannual thorough review of institutional programs by the Institutional Animal Care and Use Committees. Reports of these reviews must be sent to the Institutional Official, who has legal responsibility to assure compliance and be available for review by external regulators and site visitors. The success of this system can be seen by looking at results of AAALAC accreditation site visit findings.

An additional measure of the status of researcher training was determined by a survey of US Board Certified Laboratory Animal Veterinarians. It was found that 87% of the respondents felt that, at their institution, research training was good to excellent. The most commonly identified strengths of programs included: comprehensive; standardised; tailored to meet needs of scientist; ease of access; training tied to research privileges; and face to face building of trustful relationships. However, areas for improvement were also noted and included such items as: the need for more hands on practical training; the need for more refresher courses and continuing education, and post training assessment and monitoring.

7 Summary of the Panel Discussion

During the workshop, the short presentations of the invited speakers were followed by a general discussion with the speakers and other participants on some key elements of the education and training of the scientist.

The first part of the discussion focused on identification of the persons that are actually responsible for the animal experiments and need education in laboratory animal science (LAS) and the 3Rs. Generally it is the principal investigator (PI) who designed the study, received the grant and who is often also legally responsible. This person generally does not perform the experimental procedures him- or herself, but, nevertheless, needs to be educated in animal use and care (laboratory animal science) in order to have insight into all aspects of animal experimentation. It was concluded that the PI should at least have an academic background in one of the biomedical disciplines and have completed a course in LAS. In some countries, animal ethics committees (AECs, IACUCs, ACCs, etc.) assess the required skills of the persons involved in the animal experimentation and may conclude that the complete team, rather than one individual, has the necessary knowledge and experience for a particular study involving animals. It was, however, generally agreed that also in this situation there should be one person who is finally responsible for the experiments, and this person should have been educated in one of the biomedical areas and, in addition, must have taken some form of LAS education.

It was acknowledged that there are great differences between countries and even between institutions within countries with respect to the length (from 2 hours up to 120 hours) and depth of courses. In Europe, scientists are considered qualified (but not yet competent) to design animal experiments after having graduated in one of the biomedical disciplines and having completed the 80-hour FELASA C course on laboratory animal science. In the USA, IACUC's usually decide on what aspects the scientist should be trained in. This effectively means that there is a great variation in the requirements between institutes. Often, several modules must be taken in order to be approved to do animal research. Which modules must be completed depends on the nature of the research. A minimum duration of the course is regarded acceptable as long as the students are aware of the fact that education and training never stops and is a continuous process. It is very important that through continuing education scientists are kept up-to-date with new developments, even when they become more distantly involved in the practical aspects of animal experiments.

Attitude training was identified as the most important aspect of the course, which, at the same time, is the most difficult aspect to assess. Persons need to be trained in reflective practice. Ideally, attitude training should not be restricted to the LAS course but be part of the basic training of MSc students. As part of the attitude formation towards animals it should become natural that the default starting point is research without animals. The question that scientists should ask themselves is whether an experimental animal is really needed to obtain the necessary results of a study. Attitude training should include discussions on the ethics of animal experiments, which can be stimulated by inviting representatives from animal welfare organisations. Although ethics is seen as a recurring thread throughout the course, it was mentioned that it is also advisable to have an ethicist discussing bio-ethics during the course who could present the students with case-studies and ethical questions for group discussions. Education on all aspects of the 3Rs is seen as an essential and intrinsic part of the basic course. The scientists should be aware of the 3Rs principles and should not only be able to perform effective literature searches for alternatives, but should also understand *why* they are performing those searches.

Furthermore, the need for hands-on training of experimental procedures was discussed. It was generally acknowledged that some practical training in handling and basic procedures should

be an essential part of the course. This training could also contribute to the attitude of the scientist, making him/her aware of the impact of procedures on the welfare of the animals and the results of the experiments. As some courses focus on the most commonly used laboratory animals, rats and mice, the question was raised whether it is ethical to use mice to train scientists that work only with fish, for example. On the other hand, not every technique with every animal can be covered during the course. For example, when persons perform wildlife studies, some of the courses will have no opportunities or knowledge available for such specific training. It was suggested that in these situations the scientists should be allowed to train "on the job" and have the team leader sign off and be responsible for the competence of the team members. Therefore it was concluded that specific skills for the performance of procedures should not be part of the general course but should be taught on an individual basis, depending on the requirements of the research. For that reason it was suggested that the ideal course should contain a basic theoretical component for every scientist involved in animal experiments. The practical training should then be given in separate modules or on the job, specifically focused on the animals and procedures the scientist is going to work with and apply.

To facilitate access to LAS courses, some modules could be supplied through the Internet (Web-based learning). On the other hand, it was also generally agreed that group discussions, student to student and student to teacher contact and discussions have a great impact on the students and their attitude development and are often appreciated by them. But the modern opportunity of group discussions through the Internet allows flexibility and may also stimulate interactions among students and between students and teachers. As was stated during the discussion, young students are used to the new media and to communicate by these modern means. What education method is preferred also depends on resources and teacher's availability. This is particularly crucial in Asian countries, where training in laboratory animal science is not yet common. Furthermore, it is not only the cost of the course that should be taken into account, but also the cost of the scientist who is not available for work when attending such a course. On the other hand, it was stated that it should be realised that education on laboratory animal science and the 3Rs not only helps to reduce animal numbers and their suffering, but will also contribute to better research.

To harmonise course content and length and facilitate exchange of scientists several organisations (FELASA, ILAR, CCAC and ICLAS) have developed guidelines documents and guiding principles. Future activities could possibly lead to harmonisation of the different guidelines documents.

8 Conclusion

The need for the education and training of scientists in the principles of laboratory animal science and the 3Rs is generally acknowledged. It was agreed that a major objective of education and training is the development of a proper attitude, based on respect for animals and consideration of 3Rs alternatives. In addition, some form of hands-on training is essential. Live contact between the scientist and laboratory animals in the form of animal handling and observations was mentioned as the minimum that courses should provide.

There are, however, major differences between countries on how competence can be achieved. Harmonisation of course content and the goals to achieve is important for the international exchange of scientists. The workshop revealed a fundamental difference between countries that require a general course for all types of experiments (with emphasis on attitude) and countries that tend to follow a system with modules designed to meet the specified needs of the scientist (with some emphasis on technical aspects). This discrepancy needs to be further discussed before recommendations on international harmonisation of requirements can be made. This workshop may provide a basis for further discussions on this process. Guidelines as proposed by ICLAS, ILAR, CCAC and FELASA can be used to develop generally accepted criteria for such courses.

The final goal will be to minimise the use and suffering of laboratory animals by raising awareness of the opportunities that 3Rs methods offer. This will not only benefit the animals but also improve the quality of science.

References

(Last access of websites was October 14th, 2009)

CCAC (1999). *CCAC guidelines on: institutional animal user training*, http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/NIAUT/NIAUTCOV.HTM

CCAC (1989). *CCAC policy statement on; ethics of animal investigation*. http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/POLICIES/ETHICS.HTM

CFR (1985). Code of Federal Regulations, Title 9 (Animals and Animal Products), Subchapter A (Animal Welfare). Washinton DC, Office of the Federal Register.

Dobrovolny, J., Stevens, J. and Medina, L. V. (2007). Training in the Laboratory Animal Science Community to Support Adult Learning. *ILAR Journal 48*, 75-89.

European Council (1986). Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. *Off. J. Eur. Comm. L 358*, 1-29. Brussels.

European Parliament (2009). European Parliament legislative resolution of 5 May 2009 on the proposal for a directive of the European Parliament and of the Council on the protection of animals used for scientific purposes. http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0343+0+DOC+XML+V0//EN&language=EN.

FELASA (2002). Recommendations for the accreditation of laboratory animal science education and training. *Lab. Animals 36*, 373-377

FELASA (1995). FELASA recommendations on the education and training of persons working with laboratory animals: Categories A and C. *Lab. Animals 29*, 121-31

ILAR (2007). Training and Adult Learning Strategies for the

Care and Use of Laboratory Animals. *ILAR 48*(2).
ILAR (1996). Guide for the Care and Use of Laboratory Animals. Institute of Laboratory Animal Resources, Commission on Life Sciences, National Research Council, National Academy Press.
ILAR (1991). Education and Training in Care and Use of Laboratory Animals, A Guide for Developing Institutional Programs. Institute of Laboratory Animal Resources, Commission on Life Sciences, National Research Council, National Academy Press.
Lloyd, M. (2009). Advancing Refinement Through Training: Is there a Role for Reflective Practice? *ATLA 37*, 167-171.
NHMRC (2008). *Guidelines to Promote the Wellbeing of Animals used for Scientific Purposes*, Commonwealth of Australia, Canberra. http://www.nhmrc.gov.au/publications/synopses/ea18syn.htm
NHMRC (2004). *Australian Code of Practice for the Care and Use of Animals for Scientific Purposes*. Commonwealth of Australia, Canberra. http://www.nhmrc.gov.au/publications/synopses/ea16syn.htm
PHS (Public Health Service) (1996). Public Health Service Policy on Humane Care and Use of Laboratory Animals, Washington DC; U.S. Department of Health and Human Services, 18 pp, (PL 99-158, Health Research Extension Act, 1985)
Russell, W. M. S. and Burch, R. L. (1959). *The principles of humane experimental technique*. London: Methuen.

Acknowledgements
We would like to thank the audience of the session on Education/Training during the 7th World Congress on Alternatives & Animals in the Life Sciences for its active participation in the discussion.

Correspondence to
Dr. Jan van der Valk
Netherlands Centre for Alternatives to Animal Use
Fac. of Veterinary Medicine
Utrecht University
PO Box 80.166
3508TD Utrecht
The Netherlands
e-mail: J.vanderValk@uu.nl

FELASA Category C Courses

Bryan Howard[1] and Patri Vergara[2]

[1]FELASA, Staffs, UK; [2]Facultat Veterinaria, Universitat Barcelona, Spain

Summary

Category C training is an agreed approach to the training of laboratory animal scientists who design and oversee animal experiments. Attendance of such training has become mandatory in many European countries as a means of assuring competence, thereby meeting the requirements of Directive 86/609. The recommended course content is discussed in relation to the role of raising awareness about the 3Rs, and reference is made to the impact of the FELASA Accreditation Scheme in promoting uniform high standards and harmonisation across Europe.

Keywords: competence, education, training, European Directive

The European Regulatory Framework

During the process of revising the European Directive 86/609/EEC (European Council, 1986), stakeholders within the European Community have been consulted about their attitudes concerning the use of animals for experimental purposes (European Commission, 2006). There was an extremely high response rate (42,655 replies from all 25 Member States, as well as third countries). The survey was carried out in the summer of 2006. The scientific rigour of the survey can be questioned, because it involved open solicitation of comments rather than direct questioning of citizens and therefore tends to represent disproportionately the views of people who hold strong opinions. However the results of this survey showed that almost three-quarters of the population felt that the level of welfare and protection of experimental animals within the European Union was either poor or very poor. Citizens also felt that the ethical acceptability of research depended on a number of factors, including its objectives (approximately half), the amount of pain, suffering and distress caused (three-quarters), whether or not there were alternative methods (70%), and the care taken with husbandry and welfare of the animals (approximately half). These issues relate directly to the 3Rs of Russell and Burch (1959) and the need for cost-benefit assessment in deciding whether the research is ethically justified. The European Directive has been seen as a framework for addressing these various issues and incorporates recommendations intended to help address them.

At the same time as the Directive is being developed, calls are being made for the introduction of new diagnostic approaches, therapies and prophylactic measures for both human and veterinary use, such as vaccination against and treatment of H5N1 influenza infection and addressing enzootic diseases in Third World countries, including scientific endeavours conducted under the widely acclaimed "Grand Challenges in Global Health" initiative through the Bill & Melinda Gates Foundation. Also, there is increased concern about environmental issues and calls for more rigorous safety testing of chemicals used within the environment, for example the REACH initiative (European Council, 2006).

Scientific research is thus faced with two opposing forces: public concern about the way in which science is conducted and animals are cared for and demands by society that testing should continue but that compliance with the 3Rs of Russell and Burch should have a high priority. A great deal of the responsibility for bridging this dilemma lies with those responsible for designing and overseeing the conduct of animal experiments. These persons identify programmes of work which address specific scientific questions, break these down into discrete hypotheses, design experiments necessary to test those hypotheses, and are also responsible for the way in which animals are used during the course of experiments. They can also have a major influence on the way in which animals are kept, by their relationships with those directly responsible for animal care routines. It is for this reason that Article 7 of the European Directive 86/609 stipulated that "Experiments shall be performed solely by competent authorised persons, or under the direct responsibility of such a person, or if the experimental or other scientific project concerned is authorised in accordance with the provisions of national legislation". The text currently under consideration for the new European Directive retains the requirement that those involved in the care and use of animals for scientific purposes must be competent (European Parliament, 2009). For example in the current draft, Article 20 requires that authorities should "have evidence of the requisite competence" and that all authorisations of persons should "be granted for a limited period, not exceeding five years…only granted on the basis of…the requisite competence".

Categories of competence

In recognising these requirements, FELASA identified four categories of individuals, the activities of which impact directly on the conditions under which animals are kept and used:

– Category A: Persons taking care of animals;
– Category B: Persons carrying out experimental procedures (FELASA, 2000);
– Category C: Persons responsible for directing or designing experiments (FELASA, 1995);
– Category D: Laboratory animal science specialists (FELASA, 1999).

Recommendations concerning basic competencies in animal care and husbandry for Category A were seen as having broad applicability in satisfying the requirement of the 1986 Directive that "persons carrying out or supervising the conduct of experiments" (Categories B and C) "shall be capable of handling and taking care of laboratory animals" (Article 14 of European Council, 1986). Recommendations for practical, theoretical and ethical training were applicable for all Categories and included strategies for utilising replacement techniques wherever possible, minimising the number of animals used and refining techniques.

The requirement for competence was originally met by a miscellany of national initiatives, such as compulsory courses introduced in the Netherlands in 1985, in the UK (1994 and 1995) and in Switzerland (1999). Many other countries now recognise the importance of conducting scientific programmes to standards acceptable to the principal global markets within Europe, the US and other advanced economies. In Europe this process is being accelerated by the accession of new countries to the EU, and imminent changes to the European Directive.

Oversight of research: FELASA Category C

It has long been recognised that the skills required of a competent researcher extend beyond the core scientific needs of the relevant academic discipline. Increasing concern for wider implementation of the 3Rs has extended the definition of competence to include better matching of scientific need with the impact of science on animals. Competence, then, contains elements of knowledge, understanding, skill and attitudes and involves a commitment to lifelong learning. Of these elements, knowledge is the easiest to acquire – formal taught courses, reading books, seminars, e-learning, etc. Understanding is more difficult to attain, but may involve interaction with situations or other people – perhaps partly achievable by involvement in discussion groups or tutorials and being mentored by experienced colleagues. The basis of practical skills – including underpinning theoretical aspects – can be taught, but the development of sufficient skill to fulfil the needs of competence requires considerable time working alongside and under the supervision of experienced colleagues. Appropriate attitudes can be learned and acquired, but possibly not formally taught – they involve the reception and processing of information and understanding its in-depth meaning rather than recall of memory; it requires a developed appreciation of the attitudes of self and others. In view of the foregoing, it is unlikely that rounded competence can be developed within a formal training course although many of the underlying principles and elements can be developed and a basis provided for the individual to become a successful and ethically conscious researcher.

– It can reasonably be argued that Category C persons require three principal characteristics:
– Possession of the right attitude toward the ethical complexities which their work entails.
– A willingness to work as part of the team, which continually strives to improve the quality of animal care, displays compassion towards the experimental subjects by application of the 3Rs and conducts high quality science which is relevant to the needs of the biomedical community.
– Possession of an appropriate level of competence to ensure that responsibilities are carried out effectively and efficiently.

Of these, the first two are possibly the most important for the supervisor of experiments, who can employ others more manually dextrous to conduct the research; however, these are the least amenable to training, because they depend not so much on the possession of factual knowledge or on practical skills but on the recognition, acceptance and display of appropriate attitude. This attitude is one which would be demonstrated by a person who strives to minimise animal suffering, not only in the design but also in the conduct of studies, who is sufficiently open-minded to seek and accept information and advice from others and who fully appreciates that scientific procedures impose costs on animals. Such a person would continually re-assess those characteristics and strive to maximise them. This same individual must recognise how his/her research relates to current perceptions of the scientific and broader societal community and remain mindful of the responsibilities which using animals entails.

In summary, training appropriate for Category C persons should perform three functions:

– Impart knowledge;
– Raise awareness;
– Influence practice.

It is the matter of practice, or possessing the right degree of competence, which forms the core of category C training. The remaining two requirements are expected to be delivered during the course of this training, but, as seen earlier, are difficult to assess and quantify, so they are not specifically identified here. Despite this, a full appreciation of the topics recommended by FELASA will provide the tools for someone with the appropriate attitude to become an effective and responsible category C person. Eight principal topics have been identified for inclusion within the teaching syllabus for Category C. These are:

Biology and husbandry of laboratory animals.
A broad consideration of the uses of animals, their behaviour and biology, issues relating to handling, transportation, nutrition, health status, the importance of genotype and the selection of appropriate genetically defined models, the ability to recognise, assess and control pain, suffering or distress.

Microbiology and disease.
In particular, this includes health monitoring and the prevention of disease by appropriate biosecurity and biocontainment

methods, an appreciation of the nature of diseases of laboratory animals and impact which these may have on animal welfare and scientific findings, and the importance of taking appropriate precautions when working with animals with infectious or other diseases.

Health hazards and safe practices in the animal houses.
This includes an awareness of issues relating to allergies, zoonoses, pathogens and chemical and physical hazards, including those posed by technical equipment and test substances, including radioactive materials.

Alternatives to animals.
Not only appreciating the importance of using alternatives to living animals when they exist and sources of information about them, but an understanding of the range of alternative methods available, their possibilities, limitations and implementation.

Skill in the design and conduct of animal experiments.
This a broad field that includes an appreciation of issues relating to experimental design, including strategies for minimising the number of animals required whilst ensuring that the groups selected and numbers in each are appropriate for the purpose of the experiment in hand; preparation of a protocol for experiments, including review of literature to identify the best animal model; specification of the source of animals, the technical procedures to be conducted, requirements for their care and the conduct of experiments in ways which represent best practice and are scientifically rigorous.

Anaesthesia, analgesia and experimental procedures.
These topics relate to the conduct of experiments and include an appreciation of the importance of minimising pain by selecting the most appropriate way of conducting procedures, including use of anaesthetic and analgesic agents in such a way that animals will experience the minimum of harm. The syllabus includes instruction about the choice of anaesthetic and analgesic agents in relation to the animal species and the nature of the experiment, an understanding of how to manage an anaesthetic protocol, including dealing with emergencies and possible complications, monitoring and effective recovery post-anaesthesia. It also includes the conduct of non-surgical experimental procedures (including euthanasia) and the principles of surgical interventions, including the need for aseptic technique.

Ethical aspects and legislation.
In this section, delegates are taught about the importance of possessing appropriate attitudes and come to appreciate the intrinsic value of animals including an understanding of the nature of societal concerns about the use of animals in science and a full understanding of the ethical implications which arise from this. It is also important that Category C persons have a clear understanding of the law which regulates their activities and of their legal responsibilities under that.

Analysis of scientific literature.
This includes advice on search strategies for implementing the 3Rs, how to rigorously scrutinise scientific publications with a view of avoiding pitfalls when designing studies and how to search for and understand the significance of cognate findings to one's own scientific field, so that more appropriate experimental objectives identified with higher scientific value and less impact on the animals are used.

Going forwards

It is clear from this that the Category C syllabus emphasises the importance of the 3Rs in the planning, design and conduct of scientific experiments. Courses usually include practical guidance on searching for alternatives, selection of models and selecting an experimental design that minimises the number of animals required and refinement by reviewing strategies for conducting experimental procedures so as to minimise their impact on animals' well-being. However, no programme of study can be guaranteed to create a researcher who is fully competent to design and oversee scientific experiments; rather it can provide sufficient of the fundamentals to enable somebody to perform in this way. It is anticipated that those who have completed a Category C course will develop their skills in conjunction with colleagues and within a supportive local regulatory climate, including the ethical review process. This will enable further understanding as the individual progresses in his or her career. It is also anticipated that learning will continue beyond the end of the course and throughout one's career.

FELASA's recommendations recognise that additional specialised training may be necessary before an individual is competent to carry out advanced techniques such as surgery, the use of specialised equipment or working with genetically abnormal animals. Such skills can be obtained by working in collaboration with other experienced investigators and animal technicians, or by attending specialist courses. FELASA recommended that at the end of the Category C course there will be an examination which tests the knowledge, skills and, so far as possible, the attitude of those who have completed training. It also recognises the need for collaboration between teaches and instructors who deliver training to ensure the availability of suitable teaching materials and the dissemination of expertise between those concerned with the development and delivery of training programmes in different establishments.

The Category C syllabus has formed the basis of training delivered to senior scientists in many European and some other countries. However, recently it has become apparent that the recommendations are being seen to have wider relevance and, in some cases, courses are presented to animal research technicians and others whose duties do not include the formulation of research programmes but rather their implementation. Courses based on the same syllabus have also been presented to undergraduate students as a way of raising their awareness of the key ethical and legal issues and providing them with basic knowledge relating to the scientific use of animals.

The original FELASA Category C recommendations addressed neither the depth of study nor the process and quality of training, and courses established by different training providers varied considerably in content, emphasis and length. Consequently, FELASA later introduced an accreditation scheme designed to assure the quality of education and training, promote responsible and high quality science, facilitate free movement of personnel between countries, and to assist with further harmonisation (FELASA, 2002). This scheme further promotes refinement by insistence on low ratios of student to tutor in practical classes, and reduction by promoting the use of replacement strategies in training and focuses on the content and effectiveness of training and assessment so as to avoid unnecessary use of animals or the need to repeat training. Category C training has proven to be very popular, and many establishments in Europe have adopted it as the standard for training laboratory animal scientists.

References

European Commission (2006). Public consultation on the revision of Directive 86/609/EEC on the protection of animals used for experimental and other scientific purposes. http://ec.europa.eu/environment/chemicals/lab_animals/background_en.htm. Accessed 23 August 2009.

European Parliament (2009). European Parliament legislative resolution of 5 May 2009 on the proposal for a directive of the European Parliament and of the Council on the protection of animals used for scientific purposes. http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0343+0+DOC+XML+V0//EN&language=EN. Accessed 23 August 2009.

European Council (2006). Regulation (EC) 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Off. J. Eur. Comm. L 396*, 49.

European Council (1986). Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. *Off. J. Eur. Comm. L 358*, 1-29. Brussels.

FELASA (2002). FELASA recommendations for the accreditation of laboratory animal science education and training. *Lab. Anim. 36*, 373-377.

FELASA (2000). FELASA recommendations for the education and training of persons carrying out animal experiments (Category B). *Lab. Anim. 34*, 229-235.

FELASA (1999). FELASA guidelines for education of specialists in laboratory animal science (Category D). *Lab. Anim. 33*, 1-15.

FELASA (1995). FELASA recommendations on the education and training of persons working with laboratory animals: Categories A and C. *Lab. Anim. 29*, 121-131.

Russell, W. M. S. and Burch, R. L. (1959). The *Principles of Humane Experimental Technique*, London: Methuen.

Correspondence to

Prof. Patri Vergara
Unitat Fisiologia
Facultat Veterinaria
Universitat autonoma de Barcelona
08193-Bellaterra (Barcelona)
Spain
e-mail: patri.vergara@uab.es

BS20: Animal use policies

Animal Experiments – the Principle of Proportionality as Key Principle of Ethical Evaluation

Joerg Luy
Veterinary Faculty, Free University, Berlin, Germany

Summary
Since 1986 animal experiments in Germany may only be approved by the authorities if they are indispensable for a permissible purpose and the inflicted suffering is not classified as disproportionate to the benefits thereof. In order to verify these conditions, an ethical evaluation procedure has been developed, which – based on the Principle of Proportionality – fulfils not only bioethical but also legal requirements. The ethical evaluation procedure is performed first by the scientist applying for the authorisation, second by the animal experimentation committee of each competent authority, and third by the competent authority itself. Based on the outcome of the ethical evaluation procedure every year several applications are rejected in Germany.

Keywords: animal experiments, animal protection, ethical evaluation, ethics, Principle of Proportionality

1 Introduction

Since 1986, animal experiments in Germany may only be approved by the authorities if they are indispensable for a permissible purpose (*unerlässlich*) and the inflicted suffering is not classified as disproportionate to the benefits thereof (*ethisch vertretbar*, Article 8, paragraph 3 German Animal Protection Act, TierSchG). In order to verify these conditions, each competent authority is required to convene one or more committees consisting of scientists experienced in the assessment of animal experiments and of selected members of proposed lists of animal welfare organisations (Article 15 TierSchG). These Animal Experimentation Committees have been gaining experience in the ethical review of applications for animal experiments since 1986. Because rejected applications of animal experiments are often resolved judicially, an ethical evaluation procedure has been developed, which – based on the Principle of Proportionality – fulfils not only bioethical but also legal requirements.

2 The Principle of Proportionality

The Proportionality Principle is an ethical tool mostly used to regulate ethical dilemmas by law. In ethical dilemmas, every feasible option leads to an outcome which must be judged as ethically problematic or even immoral. Ethical dilemmas have therefore only ethically problematic solutions. In such situations, the Proportionality Principle helps to find the least problematic way out. The Principle of Proportionality is a political maxim accepted worldwide and used for the solution of a great variety of social problems (including animal welfare), and it is a fundamental principle of EU law in the allocation of legislative competencies within the Union. In animal protection legislation the Proportionality Principle works worldwide as the key principle.

Although often reduced to one or two steps, the Proportionality Principle consists, for logical reasons, of not less than four steps. When using it to find the least problematic solution in an ethical dilemma one must test every alternative action (including omission, i.e. doing nothing) by four test steps. The Proportionality Principle is designed in such a way that only one alternative is able to fulfil the requirements of all four test steps: this will be the ethically least problematic and therefore the an ethically justifiable option. The four test steps are:

1. Check whether the stated purpose of the intended action (or omission) is *permissible*!
2. Check whether the action (or omission) is in fact fit to promote the stated purpose! In addition, the action (or omission) should, compared to alternatives, be the most likely to achieve the benefits of the stated purpose.

3. Check all adverse effects of the action (or omission) on humans, animals and the environment to see if each is *indispensable* (i.e. necessary, essential) to achieve the stated purpose! This includes checking for and considering all possibilities of compensating adverse effects.
4. Check whether the action (or omission) is proportionate compared with the hoped-for benefits of the action (or omission) in regard to every adverse effect! To be proportionate, the action (or omission) must fulfil the following conditions:
   a. The action (or omission) must lead to a less adverse result than the omission (or any action).
   b. Every action (or omission) must be in itself permissible, i.e. may not harm absolute (deontological) moral rights which have been proposed, to be protected against weighing, e.g. human rights; in animal ethics at least the right not to undergo severe suffering, i.e. suffering that would be judged "unbearable" by a human proband (test person).
   c. The action (or omission) must, in spite of its moral shortcomings, be judged adequate, proper and fair in its context by the majority of citizens of the corresponding territory who have developed a sound moral attitude.

In Germany, animal experiments have been restricted by the Proportionality Principle for several decades. Today, its four test steps are incorporated into the Animal Welfare Act (Article 7 and the following). While any person wishing to conduct experiments on vertebrates must obtain authorisation of the planned experiment from the competent authority, authorisation is to be granted only if and when scientific evidence is produced by the scientist a) that the experiment is *indispensable* for and *fit* to promote a *permissible* purpose (Article 7 paragraphs 2, 4 and 5 define "*permissible* purposes") and b) that the pain, suffering or harm that can be expected is "ethically justifiable" (*proportionate*) compared with the hoped-for benefits of the experiment. Thus, a precondition of the authorisation is the fulfilment of the requirements of all four test steps. The test is performed first by the scientist applying for the authorisation (these results are set out in the application), second by the animal experimentation committee of each competent authority (their results are to assist the authority in deciding whether to authorise the experiment), and third by the competent authority itself (those results ultimately decide whether the authorisation is granted or not).

Step 1: Check whether the stated purpose of the intended procedure is permissible!

The following are regarded as "permissible purpose" (in Germany):

1. the prevention, diagnosis or treatment of diseases, suffering, bodily defects or other abnormalities or the detection or exertion of influence of physiological conditions or functions in human beings or animals;
2. the detection of environmental hazards;
3. the testing of substances or products to ensure that they are safe in terms of human or animal health or that they are effective against animal pests;
4. basic research.

Experiments on animals to develop or test weapons, ammunition and related equipment are prohibited, as are, subject to an exemption clause, experiments on animals to develop tobacco products, detergents and cosmetics.

A problematic shortcoming with regard to the first test step is, however, that in practice only experiments subject to authorisation must be checked. Since in Germany no "authorisation from", but only a "notification to" the authority is required for planned experiments expressly required by a statute, an ordinance or the *Pharmacopoeia*, or another binding provision, these experiments continue to be performed for many years even if their ethical justification is more than dubious (e.g. the mouse bioassay as gold standard for detecting biotoxins in shellfish for human consumption, or the LD_{50} test for every batch of Botox produced even when applied for cosmetic reasons). Animal testing of luxury goods like shellfish or those Botox batches that are used for cosmetic purposes cannot comply with the Principle of Proportionality and is, therefore, "ethically unjustifiable". For testing shellfish there an alternative method already exists, and an alternative is under development for Botox testing. Consequently the use of Botox for cosmetic reasons should be provisionally banned (since it would also be unjustifiable to use untested batches).

Step 2: Check whether the intended procedure is in fact fit to promote the stated purpose!

Only protocols that are *fit* to promote the stated purpose may be authorised. Therefore it is necessary to check, among other things:

1. whether the number of animals to be used is the *minimum* necessary to ensure a meaningful interpretation of the findings and the statistical validity of the findings;
2. whether *species*, *sex* and *age* of the experimental animals are fit to promote the stated purpose;
3. whether there are doubts that the results might not be transferable to the species of the stated purpose;
4. whether protocols and endpoints are the most *likely* to produce satisfactory scientific results.

Step 3: Check deterioration of the quality of life at all stages of the animal's life to see if each is indispensable to achieve the stated purpose! – The Three Rs principle.

Only experiments with animals that have a good life quality during their whole life span except for those cutbacks that are each indispensable to achieve the stated purpose can fulfil the requirement of this test step. This includes checking for and considering not only all the possibilities to avoid adverse effects, but also those to compensate them. Thus, within the third test step one has to check compliance with the the Three Rs principle (Replacement, Reduction and Refinement).

1. *Replacement*: means the attempt to replace animals by non-living or at least non-sentient alternatives. – Therefore, a check must be made to see:

- whether the purpose of the specified programme of work

could be achieved satisfactorily, in whole or in part, by any other reasonable and practicable method that does not require the conduct of procedures on animals (in particular, whether in advance performed *in vitro* screening would influence the ethical test results of the protocols);
- whether there are chances to replace animals by non-sentient (e.g. early stages of development) or nonliving alternatives, or by a protocol in which the animals are gently made unconscious prior to the experiment and are then killed before they regain consciousness.

2. *Reduction*: means the attempt to reduce the number of animals to the minimum necessary without compromising the quality of scientific results. The UK Home Office rightly advises: "It is recognised, however, that the number of animals that need to be used can sometimes be reduced if *additional suffering* is allowed to be caused to fewer animals. The method licensed will be the one judged to cause *the least suffering* or distress." (Guidance on the Operation of the Animals (Scientific Procedures) Act 1986, Chapter 5, No. 5.15) – Therefore, a check must be made to see:
- whether the number of animals is reduced to the minimum (e.g. check group size, indispensability of each group, in particular, of identical but not contemporaneous groups, sequence of protocols, and used means of statistical analysis), subject to the exemption that a higher number of animals allows for less suffering or distress of individual animals.

3. *Refinement*: means the employment of methods to ensure that any possible pain, suffering, distress and harm is reduced to the minimum, as well as to improve the care, treatment and living conditions of the animals to enhance their well-being. – Therefore, a check must be made to see:
- whether protocols and endpoints, compared with feasible alternatives, cause the least pain, suffering, distress or harm (only *unavoidable* pain, suffering, distress or harm may be authorised; in Germany pain, suffering, distress or harm may not be inflicted "to save work, time or costs" (Article 9 (2) No.3); to be considered are, in particular, inspection schedules, opportunities for general or local anaesthesia, analgesics and today's *most* humane way of killing);
- whether there are sufficient measures proposed to prevent or minimise the extent, duration and incidence of the adverse effects (measures should include reasonably designed control measures, because it is indispensable for the laboratory animal to be monitored closely, and as an individual animal);
- whether there are chances to refine techniques, management or housing in a way that the harm caused to the animals is minimised (to be considered are "*environmental enrichment*" and sheltered positions for the animals to retreat to);
- whether protocols incorporate best practice and will be applied competently (only by skilled personnel);
- whether all sick or injured animals get treatments that alleviate pain or suffering (in Germany the applicant must produce scientific evidence of incompatibility with the objective of the experiment for each untreated but suffering animal);
- whether all reasonable steps are taken to ensure that the physical, health and behavioural needs of the animals are met in accordance with both good practice and scientific knowledge (the Commission Recommendations 2007/526/EC on guidelines for the accommodation and care of animals used for experimental and other scientific purposes define minimum standards);
- whether choice of species and stage of development is made in order to use animals that have the lowest degree of neurophysiological sensitivity (i.e. have the least capacity to experience pain, suffering or distress).

Step 4: Check whether the deterioration of the quality of life is proportionate compared with the hoped-for benefits of the experiment!

Scientists are often not familiar with the requirement of the fourth step of the Proportionality Principle to let the majority of the (morally developed) citizens (of the corresponding territory) decide whether a scientific procedure on animals is "proportionate" (adequate, proper and fair) in its context. Since in Germany the whole Animal Welfare Act refers to the Proportionality Principle, even its central idea of justifiability, the "good reason" (*vernünftiger Grund*), is bound to what the majority of the (morally developed) German citizens believe to be a "good reason".

Step 4a: Check for protocols that are *absolutely prohibited*, for they are likely to conflict with absolute (deontological) moral rights, such as, in particular, the animal's moral right not to undergo severe suffering (i.e. suffering that would be judged "unbearable" by a human proband). Absolute (deontological) moral rights (e.g. human rights, the bans on slavery, torture etc.) have been proposed to be protected *against weighing*. Exemptions are not intended; it is, therefore, not possible to argue e.g. on the basis of the eminent importance of the experiment. While "severe suffering" is expressly prohibited by UK law (experiments must include the specification of "*humane endpoints*", i.e. animals are to be humanely killed before the procedure has finished, and some of the expected gain in knowledge is waived), in Germany procedures will not be authorised if pain, suffering, distress or harm are considered "ethically unjustifiable" by the authorities. The local authorities decide without federal guidelines (but often after consultation with each other or with external experts), what (parts of) protocols must be considered as "ethically unjustifiable". Experiments on non-human hominids (great apes) are regarded as "ethically unjustifiable" too. According to the results of a questionnaire of the EU Commission for the general public on the revision of Directive 86/609/EC, 80% of the (European) participants consider all primate experiments "not acceptable". In Germany, every year several applications are rejected because the authorities have good reasons to assume that the public judge some protocols to be "ethically unjustifiable" (e.g. just recently the case of the neurophysiological procedures on primates in the city of Bremen, which has, as in many cases, been linked to considerable and balanced news coverage by the media).

Ethical Principles and Guidelines for Experiments on Animals (Switzerland)

"Certain experimental set-ups can be expected to cause such severe suffering for animals that the weighing up of ethical concerns will always fall in favour of the animals. If it is not possible to find less harmful and more ethically acceptable test arrangements by changing the research hypothesis, it will be necessary to refrain from carrying out the experiment and to forgo the expected gain in knowledge."

Swiss Academy of Medical Sciences (SAMS) & Swiss Academy of Sciences (SCNAT): Ethical Principles and Guidelines for Experiments on Animals, 3rd edition 2005, Paragraph 3.5: http://www.samw.ch/en/Ethics/Guidelines/Currently-valid-guidelines.html

Step 4b: *Cost/Benefit Assessment*: Check whether the hoped-for benefits really outweigh the likely harm to the animals (performing the procedure must globally, in the medium term, result in less pain, suffering, distress or harm than its omission). In ethics the pain, suffering, distress or harm of the experimental animals ("cost") can only be balanced against the contribution to future reduction of pain, suffering, distress and harm, which counts as a "benefit" (identical units are required on both sides of the equation).

In order to weigh the potential benefit against the likely adverse effects, one must find the "value" and the "severity level" of each procedure. The value of a procedure is defined by the contribution of the specific outcomes of the programme of work to future reduction of pain, suffering, distress and harm, rather than by the importance of the general area of activity. The severity level of a procedure, on the other hand, is mostly defined as the upper limit of the expected adverse effects that may be encountered by an animal, taking into account the measures for avoiding and controlling adverse effects. The severity level of a procedure represents the worst potential outcome for any animal, even if it may only be experienced by a small number of the animals to be used. If several procedures are combined, the stress level for the animal will most likely increase, and the *cumulative effect* must be taken into account (e.g. water deprivation + fixation); therefore, in some cases, the severity level of combined protocols must be set higher than the level of the protocols alone. Furthermore, the cumulative effect of the repetition of identical protocols within a procedure can, in some cases, justify setting the severity level higher than the level of a single protocol. What are useful are official lists with examples for severity levels (e.g. the Swiss "Classification of Animal Experiments according to Grades of Severity prior to the Experiment": http://www.bvet.admin.ch/themen/tierschutz/00777/00778/index.html?lang=de, or the German "Guidelines to help evaluate the stress factor for laboratory animals during authorised animal experiments", released by the Berlin Work Group of Animal Welfare Officers: http://www.charite.de/tierschutz/download/Orientierungshilfe-englisch.pdf). Anyhow, much experience is needed to draw the line between "moderate" and "severe" suffering. And it is particularly difficult to assign severity categories when adverse effects are uncertain or unpredictable, for example in the production of genetically modified animals or in toxicity testing.

Correspondence to
Prof. Dr. Joerg Luy
Veterinary Faculty/Free University Berlin
Member of the Animal Experimentation Committee of Berlin
Oertzenweg 19b
14163 Berlin
Germany
e-mail: luy@vetmed.fu-berlin.de

Special Protection for Primates – the Need for Faster Progress

Maggy Jennings
RSPCA, Research Animals Department, Horsham, UK

Summary
The use of non-human primates in scientific research raises serious ethical and welfare concerns, which has resulted in legislation in some countries requiring special justification for the use of primate species. Furthermore, in the last ten years, there have been a number of authoritative reports urging tighter controls on primate use, higher standards of husbandry and care, and an end to the capture and use of wild animals. However, there appears to be great resistance to changing the status quo, and this paper will review whether progress has been made on any of these issues in the light of proposals for the revision of the EU Directive regulating the use of animals in experiments.

Keywords: primates, legislation, Three Rs

1 Introduction

The use of non-human primates in experiments is currently a hot topic of debate – something that could have been said at every World Congress since these began in 1993. However, it is particularly pertinent in Europe now, because Directive 86/609, which regulates animal experiments (European Community, 1986), is under revision and there is a good opportunity to incorporate provisions that would make a real difference to whether and how primates are used. Indeed, the notes in the introductory "Context Section" of the Commission's draft proposal (European Commission, 2008a) state that: "*specific provisions have been incorporated to reduce the use of primates to a minimum*" and that "*there are ambitious requirements on the origins of animals and specific monitoring mechanisms to ensure the effectiveness of the proposed measures ultimately facilitating the move toward abolishing the use of non-human primates in scientific procedures*."

This follows on from statements, reports and recommendations from authoritative national and international bodies, which highlight the need for further application of *all* 3Rs to primate use. The issue is firmly on the agenda, but the key question is whether anything is actually happening as a result.

There is progress in some scientific establishments in some countries, although mainly with the "R" of Refinement, but any progress in legislation seems extraordinarily slow and proposals for change encounter a high level of resistance. This paper therefore:

- reiterates why primates need special protection and what experimental use means for them – there is so much political wrangling this is sometimes forgotten;
- considers primate use in recent years and whether the trend in numbers is up or down;
- looks at some of the statements and recommendations made by European and national bodies that have an input into legislation to see whether these are taken into account in the proposed Directive; and
- makes a plea for more action on such recommendations now, not in the next one hundred years.

2 Why primates deserve special protection

Primates are highly intelligent, social animals. They occupy and interact with a large and diverse home range and have a complex range of behaviours and physical and emotional needs that are very difficult to satisfy in a laboratory environment. There is no question that they have the capacity to experience pain, and it should be assumed that procedures likely to cause pain and distress in humans are likely to have similar effects in other primates (e.g. Organisation of Economic Co-operation and Development, 2001; National Institutes of Health, 2002). It is also generally accepted that they can experience both negative emotions (fear, anxiety, boredom, frustration and, in some species, grief) and positive emotions (interest, pleasure, happiness, excitement) (see Jennings and Prescott, 2009).

The impact of a laboratory environment

It seems something of an anomaly that researchers who study primates in the wild emphasise their amazing capabilities, which are brought to television screens around the world for people to marvel at, and admire. Yet in the laboratory, these animals are confined in small cages, subjected to invasive procedures and treated in a reductionist way as research "tools".

Compare and contrast the natural and laboratory environments for macaques and marmosets, for example. The natural

habitat of long-tailed (cynomolgous) macaques is forest, commonly near water, a complex three dimensional environment. Their home range is around 2 km^2, reaching from ground level to tree-top height. They live in social groups of 10 to 50 animals, with sometimes as many as 100, and are very active – walking, leaping, climbing, foraging and swimming. They have extensive sensory and communication abilities with a perceptual world similar to humans.

Contrast this with a laboratory environment. The Guidelines for Accommodation and Care of experimental animals in the recently revised *Annex II* of EU Directive 86/609 require a minimum cage area of 2 m^2 with a height of 1.8 m. The standards are intended to encourage pair housing but some places still keep animals singly. In any case, this only allows animals to take a few steps in each direction and can never offer much complexity; it is barren and boring. In effect, animals that have evolved to live in a forest are consigned to a small metal box.

Marmosets also live in family groups of 2 to15 animals in complex forest habitats with a range of .006 to 0.06 km^2. They, too, are very active animals with a diverse sensory world. In the laboratory they may fare slightly better, being kept in pairs or family groups, but they are still very confined with the standards for cage size being a mere 0.5 m^2 in area and 1.5 m in height.

Cumulative harms

Given the nature of primates and the difficulty of providing for their needs, it is generally considered that the cumulative harms for these animals are greater than for other animals. As well as restricted housing and the effects of experimental procedures, there are significant welfare issues with respect to capture and use of wild animals, early weaning, transport, pre-transport "conditioning" in tiny cages for long periods in some supplying centres, and handling and restraint. In fact the animals' whole lifetime experience is seriously compromised.

Public and political concern

There is a high level of public and political concern about the use of primates – people have a special affinity with these animals as fellow primates. This was illustrated in the European Commission survey of public opinion in 2008 when over 93% of respondents believed it was important to improve the current level of welfare/protection for primates in research and testing (European Commission, 2006).

3 Trends in primate use

Given the level of concern about primate use it might be expected that this would have had an effect on the numbers used, but this is not so. Figures for primate use over the last ten years do not show a decrease. For example, in the UK, numbers fluctuate around 3500 per year. Figures for Europe show an increase in primate use from 7284 in 1999 to 10,443 in 2007[1], as do figures in the USA where numbers have risen from around 55,000 in 1999 to nearly 70,000 in 2007[2]. Furthermore, some scientists are talking about a likely increase in the need for primate use for "biological" pharmaceutical products and research into diseases of ageing. Even the use of chimpanzees has not been ruled out, and countries in Asia are known to be increasing their primate research capacity.

4 Legislation and relevant reports and statements

Legislation is a major driver of change as are, or should be, recommendations in reports commissioned or developed by legislative or regulatory bodies. There are many of these covering issues such as: the capture and use of wild primates; their accommodation and care; and the necessity and justification for using them. Some of the key recommendations made over the last fifteen years or so by formal European and national bodies are summarised below with a consideration of whether and how these are being taken forward in the revision of Directive 86/609.

Capture and use of wild primates for breeding or experiments

It is accepted that capture of animals from the wild and their use in experiments is a major stressor for primates (e.g. Johnson et al., 1973; Laudenslager et al., 1999; Suleman et al., 2004). The issue was mentioned in the report of the 1993 Berlin Workshop on the accommodation of laboratory animals. The workshop was organised by the German Federal Ministry of Food, Agriculture and Forestry, in conjunction with the Federal Health Office supported by Directorate-General XI of the European Commission (O'Donoghue, 1994). The aim was "*to review critically*" the standards of accommodation in the European Convention and Directive "*in the light of knowledge gained since they were formulated*". The report recommended that the use of wild-caught primates should be phased out completely within five years.

Then, in 1997, the Council of Europe issued a Declaration of Intent: "*to require precise information on the origin and the provenance of the animals (primates) with the objective of limiting the use of animals to those which are purpose-bred*"; and "*to encourage initiatives and measures to end the use of wild-caught primates*" (Council of Europe, 1997). The Declaration was signed by sixteen Member States and twelve stakeholder organisations, including Eurogroup for Animal Welfare, Federation of European Laboratory Animals Science Associations (FELASA), Federation of European Laboratory Animal Breeders Associations (FELABA), European Federation of Pharmaceutical Industries Associations (EFPIA) and European Biomedical Research Association (EBRA). The following year the issue of capture of wild primates was mentioned in the recitals to the Council of the European Union's decision to approve

[1] Figures for the EU are available from the tri-annually produced official reports: http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm

[2] Figures for the US are available from the Annual Reports of Enforcement (Animal Welfare Act) produced by the USDA (see Appendix 5 of 2007 report): http://www.aphis.usda.gov/animal_welfare/publications_and_reports.shtml

the European Convention on animal experiments. There then followed a number of authoritative national and international reports, which re-emphasised concerns over the capture of wild animals and argued the need to use only purpose-bred primates and move to the use of F2[3] animals (e.g. Animal Procedures Committee, 2002; Scientific Committee on Animal Health and Welfare, 2002).

How does this translate to the revision of the Directive? The draft proposal to revise 86/609 stated in its Article 10.1 that there should be a "move to exclusive use of F2 macaques, seven years after transposition of the Directive". Seven years is a long time, but this at least threw down a definite marker. However, despite the level of concern over primate capture, the European Parliament's report on the draft modified the text and introduced the need for a feasibility study for moving to exclusive use of F2 animals. Subsequent discussions allowed up to five years for such a study to be carried out. If this procedure is agreed, and assuming the Directive is accepted in 2010 with an immediate start to the feasibility study, then in 2015, eighteen years after the 1997 Declaration of Intent, there *may* be a decision about whether a move to F2 can be achieved – at some stage! This is surely an inexcusable amount of time to deal with an issue that all stakeholders have agreed is a serious problem.

Accommodation and care

Turning to husbandry standards, again there is a long history of reports and comments on the inadequacy of housing and care for primates in the laboratory environment. The Berlin Workshop report clearly stated that the "*existing [1986] recommendations for primate housing in the Council of Europe Convention are inadequate*". It went on to state that: "*the sizes of cages are too small to meet the behavioural needs of the animals; they provide neither adequate space for exercise nor room for environmental enrichment*". The report made recommendations to the Council of Europe and the European Commission for improvement.

Five years later, in 1998, the Council of Europe began the process of revising Appendix A to the Council of Europe Convention (Council of Europe, 2005 on the use of animals for scientific purposes in which standards of accommodation and care for laboratory animals are set out. (The minimum cage size for one or two 15-25 kg macaque at the time was 1.5 m^2 x 1.25 m). The process was an extended one with revised standards developed through working groups of experts from the stakeholder groups of animal welfare, academia, industry and regulation. It was finally concluded in 2006. The revised standards for cage sizes have since been translated into the Annex to Directive 86/609, but with a date of 2017 allowed for their adoption. Thus, *twenty-four years* after it was agreed that existing standards were inadequate, primates may have 0.5 m^2 extra space! However, at the time of writing, the European Parliament's report on the draft Directive allows exemptions and reduces the status of the standards to guidelines instead of mandatory requirements. Fortunately, some establishments already use the new standards and improve on them further, but this is by no means universal, and many macaques are still kept in small inadequate housing.

Necessity and justification for primate use

In some countries, such as the UK, concern over primate use is reflected in legislation with "additional justification" required before such animals can be used, together with exclusions on the use of wild-caught animals and Great Apes. A number of national and European reports have urged more critical scrutiny of the necessity and justification for primate use including the Scientific Committee on Animal Health and Welfare Report (SCAHAW, 2002), the UK Animal Procedures Committee (APC) report on primate use in regulatory toxicology (Animal Procedures Committee, 2006) and most recently the Scientific Committee on Health and Environmental Risks report on primate use (SCHER, 2009).

Particularly relevant in the context of the proposed EU Directive is a 2007 European Parliamentary Declaration, which called for a timetable for replacing all use of primates (European Parliament, 2007). The European Commission responded to this with reference to the Directive, which it said could "*incorporate strong incentives together with a specific review clause which would provide a mechanism to move towards the ultimate goal of phasing out primate use*" (European Commission, 2008b). The Commission went on to say that it was convinced this goal could only be achieved with a "*vision, close co-operation and combined effort of all concerned*". Animal protection organisations had said something very similar in a "Resolution" on primate use passed four years earlier at the World Congress in Berlin. The Resolution urged "*governments, regulators, industry, scientists and research funders worldwide to accept the need to end primate use as a legitimate goal; to make achieving this a high priority; and to work together to facilitate this*".

Has any of this been carried through into the revision of the Directive? As quoted earlier in this paper, there are strong statements in the introductory "Context" section of the Commission Proposal, but these are not carried through into the articles. This makes no progress at all for primates and does not reflect the spirit of the "Background" and "Recitals" text. However, it is positive to see a ban on the use of Great Apes, although extremely disappointing that it was deemed necessary to insert a "safeguard clause" to allow their use in an "emergency". Quite what emergency would warrant subjecting chimpanzees to confinement in a metal box is beyond imagining, even if the practical difficulties of acquiring, housing and using such animals within a rapid response scenario could be overcome.

There were restrictions on the use of other species of primate; use had to be undertaken "*with a view to the avoidance, prevention, diagnosis or treatment of life-threatening or debilitating clinical conditions*". Despite the fact that spokespersons in the research community argue in public that all primate research

[3] F2: Second generation animals whose parents were captive-bred.

is undertaken for serious medical purposes – and would therefore be allowed to continue even with the proposed restriction – there was a great deal of lobbying against this requirement. As a result, the parliamentary report removed the restriction on use but added in a "two year review" of primate use to examine the impact of developments in technological, scientific and animal welfare knowledge and set targets for implementation of validated replacement methods. What form this should take was not specified.

5 What is now needed

Looking at the timelines on the way these issues have been dealt with, the idea of progress is a misnomer. What is needed is fewer words and more action. It is true that some people, at some establishments, in some countries, have put a lot of effort into refinement and improving housing and care. However, much more immediate action is necessary if this is to be more universal and if the many good recommendations that have been published in reports and the scientific literature are to be translated into real achievements in Reduction and Replacement (or avoidance) of primate use as well. In Europe, the Directive can help if the requirements for an authorisation process and ethical review remain, but specific restrictions on primate use are also necessary.

Legislation, of course, will not work alone, and there are some useful recommendations that need translating into practice. The SCHER Report, for example, provides a useful starting point with its recommendations for:

– regular meetings to stimulate scientific discussions and exchange of information;
– development of databases and collaborative user networks covering data sharing, tissue sharing, exchange of knowledge and information;
– global networks to exchange information on the 3Rs, including providing clear and consistent guidance on the criteria for use of primates; and
– further negotiations between the EU, USA and Japan on harmonisation of regulatory test requirements

A UK committee comprising the major research funders in 2006 also made useful recommendations including: to undertake a systematic review of research funded in the last ten years; and to regularly collate and disseminate information about evolving research technology (Anon, 2006). Two years later there is no information in the public domain as to whether this is being taken forward, but it could be done through:

– focussed reviews of the validity and need for primate use in individual research and testing fields carried out by teams of open-minded multidisciplinary experts
– a co-ordinated, well-funded international effort to support a transition from animal-based toxicology to a mechanism-based paradigm; and
– a fundamental shift in attitude away from the defensive "why primates are essential" to a constructive exploration of "how the research goals could be achieved without them".

Finally, the essential point missing from most reports is a properly structured implementation plan for the recommendations they contain, with clearly defined responsibilities and some challenging timelines. This is what is desperately needed to achieve progress for these animals sooner rather than later, i.e. now, not in 100 years time.

References

Animal Procedures Committee (2006). *Acceptance of Overseas Centres Supplying Non-human Primates to UK Laboratories: A Report by the Primates Sub-Committee of The Animal Procedures Committee*. http://www.apc.gov.uk/reference/primate-sources-report.pdf

Animal Procedures Committee (2002). *The use of primates under the Animal (Scientific Procedures) Act 1986. Analysis of Current Trends with Particular Reference to Regulatory Toxicology*. Home Office, Communications Directorate, UK.

Anon (2006). *The use of non-human primates in research*. A Working Group report chaired by Sir David Weatherall. ISBN No. 1-903401-13-5. Printed in UK.

Council of Europe (2005). *European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (ETS123)*. Strasbourg. 18.III.1986 with Appendix A text amended according to the provisions of the Protocol (ETS No. 170) as of its entry into force on 2 December 2005. http://conventions.coe.int/Treaty/en/Treaties/Htlm/123.htlm (accessed 18.06.2009)

Council of Europe (1997). *Declaration of Intent Concerning Animals Used for Scientific Procedures*. Strasbourg. See http://www.coe.int/t/e/legal_affairs/legal_co-operation/biological_safety,_use_of_animals/laboratory_animals/Declaration%20of%20Intent%20E%201997.asp#TopOfPage (last assessed 12 December 2008)

European Commission (2008a). *Proposal for a Directive of the European Parliament and of the Council on the protection of animals used for scientific purposes and repealing Directive 86/609/EEC on the protection of animals used for experimental and other scientific purposes*. Brussels, COM (2008) 543/5

European Commission (2008b). *Response by the European Commission to European Parliament declaration on non-human primates in scientific experiments (includes a response to petitions received by the Commission on this declaration)*. EC: Brussels. http://ec.europa.eu/environment/chemicals/lab_animals/pdf/fische_suite_en.pdf

European Commission (2006). *Public consultation on the revision of Directive 86/609/EEC on the protection of animals used for experimental and other scientific purposes* [Survey period: 16 June - 18 August 2006]. EC - Environment DG: Brussels. http://ec.europa.eu/environment/chemicals/lab_animals/questionnaire1.htm

European Community (1986). *Council Directive of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scien-*

tific purposes (86/609/EEC) Official Journal of the European Union. Available at http://ec.europa.eu/food/fs/aw/aw_legislation/scientific/86-609-eec_en.pdf

European Parliament (2007). *Written Declaration 0040/2007 on non-human primates in scientific experiments*. EP: Strasbourg. http://ec.europa.eu/environment/chemicals/lab_animals/pdf/declaration_nhp_en.pdf

Jennings, M., Prescott M. J, (Joint eds.) Buchanan-Smith, H. M., Gamble, M. R., Gore, M. et al. (2009). Refinements in husbandry, care and common procedures for non-human primates. *Lab. Animals 43, Suppl. 1*, ISSN 0023-6772.

Johnson, P. T., Valerio, D. A. and Thompson, G. E. (1973). Breeding the African green monkey, Cercopithecus aethiops, in a laboratory environment. *Lab. Anim. Sci. 23*, 355-359.

Laudenslager, M. L., Rasmussen, K. L., Berman, C. M. et al. (1999). A preliminary description of responses of free-ranging rhesus monkeys to brief capture experiences: behaviour, endocrine, immune, and health relationships. *Brain Behav. Immun. 13*, 124-137.

National Institutes of Health (2002). Public Health Service Policy on Humane Care and Use of Laboratory Animals. Bethesda. See http://grants.nih.gov/grants/olaw/references/phspol.htm (last accessed 12 December 2008)

O'Donoghue, P. N. (ed.) (1994). The *Accommodation of Laboratory Animals in Accordance with Animal Welfare Requirements*. Bonn, Germany: Bundesministerium fur Ernaehrung, Landwirtschaft und Forsten.

Organisation for Economic Co-operation and Development (2001). *Guidance Document on the Recognition Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation*. Paris: ENV/JM/MONO(2000)7.

Scientific Committee on Health and Environmental Risks, SCHER (2009). The need for non-human primates in biomedical research, production and testing of products and devices. http://ec.europa.oie/environment/chemicals/lab_animals/pdf/scher_0_110.pdf

Scientific Committee on Animal Health and Welfare, SCAHAW (2002). *The Welfare of Non-Human Primates Used in Research*. Brussels: Health and Consumer Protection Directorate-General, European Commission. See http://europa.eu.int/comm/food/fs/sc/scah/out83_en.pdf (last accessed 12 December 2008)

Suleman, M. A., Wango, E., Sapolsky, R. M. et al. (2004). Physiologic manifestations of stress from capture and restraint of free-ranging male African green monkeys (Cercopithecus aethiops). *J. Zoo Wildl. Med. 35*, 20-24.

Correspondence to
Dr Maggy Jennings BSc PhD
RSPCA, Research Animals Department,
Wilberforce Way, Southwater,
Horsham, West Sussex,
RH13 9RS, UK
e-mail: mjennings@rspca.org.uk

Applying Information Gained from Animal Impact Categorization Systems – Case Study of Canada's Categories of Invasiveness

Gilly Griffin, Nicole Fenwick, Elisabeth Ormandy and Clément Gauthier

Canadian Council on Animal Care, Ottawa, Ontario, Canada

Summary

The Canadian animal impact categorization system[1] known as Categories of Invasiveness (CI) serves four main purposes: 1) it educates animal users about concepts for humane animal experimentation; 2) it alerts investigators and animal care committees (ACCs) to the degree of pain and distress that experimental procedures may cause animals and thereby serves as a tool for ethical review of animal based studies; 3) it informs the public about the numbers of animals that potentially experience the adverse impact associated with each category; and 4) it provides data that inform national policies on animal use in science. The Canadian system provides important input to the cost-benefit analysis of animal use because it assists ACCs in assessing the potential costs and provides some national consistency in ethical review. ACCs may view classification in a high CI as a cue to more closely scrutinize the justification and benefits of a protocol. CI data are published annually online and can be accessed by the media and humane organizations, although the extent of wider public access is not known. To identify trends in animal use and to inform national policy, CI statistics are reviewed annually. This information has been used to evaluate the effect of guidelines in establishing humane endpoints and to identify problems in the categorization of genetically-engineered animals (GEAs). Emerging challenges to the current Canadian system include: categorizing impacts on animals in breeding colonies, especially animals bred specifically to have a disease condition; categorizing GEAs at different stages of new animal line generation; and ensuring that publicly reported information more accurately reflects the actual adverse impacts experienced by animals. Future policy goals include using CI data to drive Replacement[2] efforts in Canada.

Keywords: Categories of Invasiveness, animal impact categorization, Canadian Council on Animal Care, ethical review, animal use data

1 Introduction to Canadian Categories of Invasiveness (CIs)

Animal impact categorization systems are classifications – typically going from least invasive, or harmful, to most invasive – that are used to describe and sort the various adverse impacts that animals are exposed to when they are used for scientific purposes. Such systems provide a way of categorizing the scientific use of live animals according to the experiences of the animals. This information can be beneficial to researchers, policy makers, and the public.

Canada is one of 11 countries with a mandated, national animal impact categorization system in place. The other countries are Australia, Finland, Germany, Ireland, The Netherlands, New Zealand, Poland, Sweden, Switzerland, and the United Kingdom (UK). Beginning approximately 20 years ago, countries have sought to find ways of identifying the level of pain and distress experienced by animals used in research, teaching, and testing. The implementation of systems to categorize adverse animal impact was influenced to a large extent by the 1978 publication "Alternatives to Animal Experiments" in which David H. Smyth, then president of the Research Defence Society in the UK, sought to categorize experimental procedures as a means to prioritize efforts for replacing the use of animals (arguing that attention should be placed first and foremost on those procedures that cause the most pain and distress).

In North America in the early 1980s, F. Barbara Orlans began writing and speaking about the potential utility of categorizing

[1] Categorization systems have been variously referred to as pain scales, harm scales, categories of invasiveness, severity scales, impact scales, and degrees of animal welfare compromise. Here we are using the term "animal impact categorization system" to more accurately represent the different types of classifications and to avoid the impression that pain is the sole impact on animals used in science. Adverse impacts include (but are not limited to) discomfort, pain, distress, fear, nutritional deprivation, and behavioural deprivation.

[2] Replacement, Reduction and Refinement of animal use, commonly known as the Three Rs, and first described by Russell & Burch (1959) provide the basis for the ethic of animal experimentation that underpins the Canadian system of oversight of animals used in science.

the adverse effects of procedures on animals to regulate the use of animals for teaching purposes. She described a "classification of biological studies based on the degree of invasiveness of the procedure to an animal" with the goal of illustrating the progression of "ethical costs" (i.e. increasing adverse impacts on animals) (Orlans, 1979, 1980). Her nine-level scale began with no use of living organisms (with least ethical costs) to use involving severe and/or protracted vertebrate animal pain (highest ethical cost).

During this period, both the United States and Canada began to establish ethical review and animal use oversight procedures. In 1982, a Swedish Professor, Karl Obrink, visited the United States to present an overview of Swedish law on laboratory animals to the Scientists' Center for Animal Welfare (SCAW), an association that aims to improve animal welfare outcomes for animals used in science (Orlans, 1993). Inspired by the Swedish example, SCAW created its own categorization system with the view that the newly legislated US institutional animal care and use committees could use it (voluntarily) as a tool when reviewing protocols. Published in the mid-eighties, the "Categories of Biomedical Experiments Based on Increasing Ethical Concerns for Nonhuman Species" sought to provide a "conceptual basis for assessing the degree of animal harm, and, therefore, for assessing ethical costs of an experiment" (Orlans, 1987, p256).

The development in 1987 of Canada's Categories of Invasiveness flowed directly from the SCAW system. Harry Rowsell (the Canadian Council on Animal Care's (CCAC's) first Executive Director and emeritus professor at the University of Ottawa) was a member of SCAW and a contributor to the development of the SCAW categorization system. The University of Ottawa adopted use of the SCAW scale early on, and then, under Dr. Rowsell's guidance, the CCAC developed a slightly modified version of the SCAW scale as national policy.

The CCAC policy statement on: the categories of invasiveness in animal experiments (CCAC, 1991) underwent considerable discussion and peer-review prior to its implementation. The Canadian system differed from the original SCAW system in two main ways. First, it included a wider category for experiments on vertebrate animals that involve moderate to severe distress or discomfort (referred to as CI-D; see below). This was a result of criticism from Canadian scientists that the SCAW categories jumped too quickly from minor stress to significant but unavoidable stress (CCAC, personaly communication 2009)

Second, while all animal impact categorization systems provide examples of procedures for each category, the Canadian system also included (and still includes) examples of outcomes to animals that should not occur. For example, animals undergoing CI-C procedures (see below) "must not show self-mutilation, anorexia, dehydration, hyperactivity, increased recumbency or dormancy, increased vocalization, aggressive-defensive behavior or demonstrate social withdrawal and self-isolation." When implemented in 1987, this was a fundamental difference between the Canadian Categories of Invasiveness and other categorization systems. It marked the beginning of inclusion of specific animal outcomes, in addition to types of procedures, and it introduced the use of the generic animal outcome of "pain and distress"

The "Categories of Invasiveness" were subsequently revised in 1989, and the severity of some procedures was upgraded, further differentiating it from the SCAW system. For example, "procedures that produce pain in which anesthetics are not used, such as toxicity testing with death as an endpoint," was moved from CI-D to CI-E (see below). A third revision to the "Categories of Invasiveness" was undertaken in 1991, adding more detailed examples to adapt the categories to non-biomedical use of animals, such as agricultural research and regulatory toxicology studies (CCAC, 1991). The need to provide specific direction on how to assign a CI to newly created genetically-engineered animals (GEAs) was identified, and guidelines requiring that GEAS to be assigned to CI-D were published in 1997 (CCAC, 1997 and Griffin et al., 2007). The need for specific direction continued as further interpretation of CIs was again added in 2003 to address wildlife research (CCAC, 2003).

In 1996 Canada began publishing annual data on the number of animals per CI (http://www.ccac.ca/en/Publications/New_Facts_Figures/trends/figure6.htm).

Today the Canadian Categories of Invasiveness range from A to E as follows:

- CI-A: Experiments on most invertebrates, or on live isolates
- CI-B: Experiments which may cause little or no discomfort or stress
- CI-C: Experiments which may cause minor stress or pain of short duration
- CI-D: Experiments which may cause moderate to severe distress or discomfort
- CI-E: Procedures which cause severe pain near, at, or above the pain threshold of unanesthetized conscious animals

As discussed, categorization is based both on procedures and animal outcomes. The list includes categories for procedures that do not use vertebrate animals (or cephalopods) and, as such, are intended to cause no adverse impacts (CI-A) and for those that cause severe unrelieved pain and are highly questionable (CI-E). Canadian investigators are required to assign proposed experiments to a CI prior to submitting the experimental protocol to an ACC for ethical review (including a verification of the assigned CI).

Investigators also are required to determine prospectively the number of animals to be used for each protocol. Then, on a yearly basis, ACCs report institutional animal use to the CCAC, including both the number of animals authorized by the ACC and the number of animals actually used. Therefore, the CCAC yearly statistics on animal use in Canada are based on the actual number of animals used; however, each animal's assignment to CI is reported from prospective data.

2 Applying information gained from the Categories of Invasiveness

It has been proposed that the use of mandated, national animal impact categorization systems serve four main purposes (e.g. Orlans, 2000; Griffin et al., 2007). These are: 1) education of animal users on concepts for humane animal experimentation (i.e. Three Rs); 2) provision of a practical tool to assist animal care

committees (ACCs) in cost-benefit review of animal use protocols; 3) provision of a system for collecting data to inform the public about the numbers of animals that potentially experience the adverse impact associated with each category to improve public accountability; and 4) provision of a system to collect data that will inform the development of national policies on the use of animals in science. However, to determine whether mandated national categorization systems are successfully fulfilling the four "main" purposes (or others), and whether the systems are meeting emerging animal welfare challenges, requires a comprehensive multi-nation review. As a first step we conducted a case-study of the Canadian system. The following sections examine the effectiveness of the Categories of Invasiveness in meeting each purpose and identify future challenges to animal impact categorization.

2.1 Are CIs educating animal users about concepts for humane animal experimentation?

In Canada, the required process of assessing protocols and assigning them to a CI is perceived by the CCAC to help educate, sensitize, and alert investigators and (CC) to the degree of pain and distress that experimental procedures cause animals. In addition, the existence of a category for procedures that do not use vertebrate animals (or cephalopods) and, as such, are intended to cause no adverse impacts (CI-A) is thought to convey the concept of Replacement (Orlans, 1993). Similarly, CI-E (procedures that cause severe unrelieved pain and are highly questionable) can be used to convey the concept of procedures that are unacceptable, regardless of scientific merit (Orlans, 1993). Although it is intuitive that education about humane animal experimentation is occurring, it is difficult to obtain empirical evidence because of the difficulty in separating the educational effect from the entire ethical review process. However, there is some information available that may help assess, in a general way, whether the categorization process educates animal users about humane experimentation.

A few published studies have touched upon the role of CIs in educating animal users in Canada. Bowd (1997) studied the effectiveness of one Canadian ACC in educating investigators and identified several inconsistencies between the beliefs of the investigators and CCAC-endorsed ethical positions and policies (which presumably include CIs). Although not specifically analyzing the role that categorization of animal impacts plays in the education of animal users, Bowd wrote that study results suggested there was a need for greater education of investigators on the ethical principles behind guidelines and policies. The CCAC guidelines on: institutional animal user training (CCAC, 1999) mandated training for all investigators wanting to use animals in their studies. The Recommended Syllabus for an Institutional Animal User Training Program (CCAC, 1999), published alongside the guidelines document, includes a section on the Three Rs with the expressed emphasis placed on achieving expertise in identification of pain and distress, limiting pain and distress, and preventing the wastage of animals. Furthermore, the CCAC online training modules for the core topics of the recommended syllabus, developed to support the implementation of the guidelines, include an entire module on pain and distress. A similar analysis to Bowd (1997) would be useful in indicating to what extent the guidelines, and above all the mandatory theoretical and practical training received by all users since 2003, have improved investigator awareness of the impact of procedures on study animals.

2.2 Are CIs providing practical tools for the ethical evaluation of protocols?

When preparing an animal use protocol, investigators must assign a CI to their study. This assignment is based on a precautionary approach that reflects the maximum level of pain and distress animals might experience. Protocols that involve the use of vertebrates (or cephalopods) in Categories B through E must be submitted to an ACC for approval to proceed. In some institutions ACCs also review CI-A protocols, usually because of an institutional policy to foster a respect for all life forms and to ensure that adequate facilities and care are available for organisms that would otherwise not be covered under CCAC guidelines.

Animal impact categorization systems, in general, appear to have originated from the desire to contribute practical tools to assist in cost-benefit based review. In Canada, use of the CI system is believed to provide ACCs with a tool to use when evaluating adverse effects caused by scientific procedures and to help achieve some national consistency in local institution-based ethical review (Griffin et al., 2007). Research shows that during cost-benefit analysis, ACC members use CIs to assess whether protocol justification needs additional scrutiny or whether protocol benefits sufficiently outweigh the costs to the animals. In one study of Canadian ACCs, Schuppli (2004) found that in addition to using CIs in the expected way, ACC members also use them to 1) indicate not only whether the justification needs to be scrutinized more but also whether the benefits need to be higher and 2) to "test" the accuracy of the protocols (e.g. do the procedures outlined match the CI indicated). Schuppli (2004, p90) wrote that "some members viewed inconsistencies in the way that the application form was filled out as reason for a negative recommendation because investigators were not portraying information accurately. Inconsistencies included whether the listed drugs matched the drugs described in procedures or whether the Category of Invasiveness matched what was expected from procedures."

Based on analysis of protocols and their subsequent related publications, Bowd (1997) observed that a large proportion of the more invasive studies were not published and suggested that "within the present Canadian system it is possible for an investigator to continue to receive ACC endorsement for invasive research despite having failed to publish the results of similar studies over a lengthy period". Although publication is not necessarily a reliable indicator of the value of animal-based work, the ability to track outcomes of protocols through post-approval monitoring, annual renewal and re-evaluation (including reassignment of CIs) affords the opportunity for ACCs to place additional scrutiny on those protocols where there is likely to be higher levels of pain and distress. In addition, there continues to be discussion by CCAC constituents relating to the acceptance of CI-E protocols. Some academic institutions prohibit any ani-

mal protocols that would be placed in CI-E from being carried out by their principal investigators. At least for the foreseeable future, however, there are highly invasive tests that have to be carried out as a requirement for animal-based data for human or environmental health risk assessment purposes. The continued use of CI-E in these instances acknowledges the potential for animals to experience severe adverse effects and signals the need for extra vigilance on behalf of study directors/investigators, animal health professionals, and ACCs. It also is an indication of where the most efforts should be placed to implement the Three Rs (Smyth, 1978).

2.3 Are CIs contributing to public accountability?

In Canada animal use statistics are reported annually by the CCAC. These statistics are collated from the animal use data from all institutions participating in the CCAC Program. Animal data are reported by species, by purpose of animal use, and by permitting a considerable level of detail to be made available to the public. Each year, upon publication of the animal use data, there are questions from media representatives and from members of the public relating to various aspects of animal use. Typically these focus on animal use in the highest category of invasiveness and the use of some large animal species, especially non-human primates and dogs. These requests for information permit the CCAC to further interpret the data and are also used as an indication of areas of public concern. However, ensuring that the public understands that CCAC data only reflect the number of animals that have potentially experienced a particular level of pain and distress remains an on-going challenge to ensuring accountability to the Canadian public and will be discussed in greater detail below.

To our knowledge, only six countries publicly report animal impact categorization data broken down into number of animals per category (Australia, Canada, Finland, The Netherlands, New Zealand and Switzerland). Comparisons can be made between countries through accessing their animal use statistics. Comparisons are difficult, however, due to the variety of ways in which different countries categorize and report the various levels of impact (e.g. countries vary in the number of categories of animal impact they use). In the future, more uniform international data are expected to become publicly available if the European Union continues with plans to establish classification based on four categories of severity and to require submission of this data by member countries (EU, 2009).

Here we report animal use data for Canada (see Fig. 1).

2.4 Are CIs informing Canada's national policies on animal use?

Analysis of the trends of categories of invasiveness has proved useful to establish whether the number of animals in the most severe categories is changing (i.e. increasing or decreasing). It follows that such analysis can give us valuable information about how the Three Rs are being implemented, and indeed, where to target future Three Rs efforts. Investigators follow the Three Rs as ethical guidelines for humane animal research, thus it is expected that the number of animals in the categories describing the most severe experimental harms should go down as the number of animals used in research is reduced and tech-

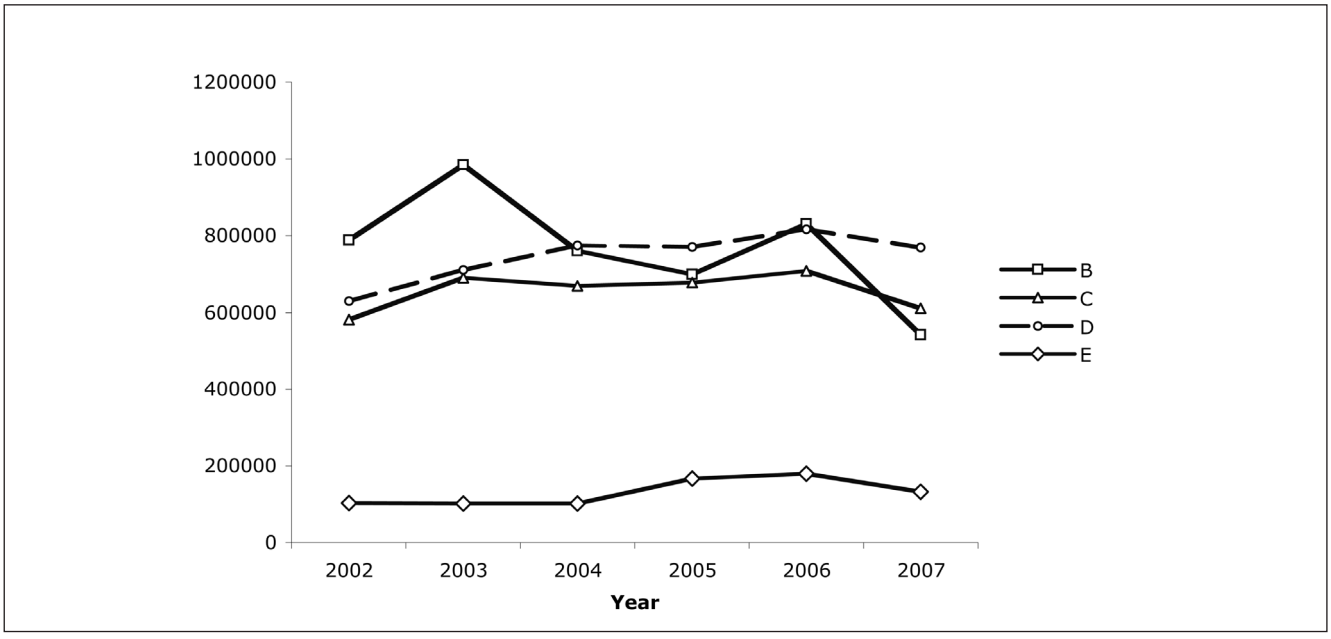


Fig. 1: Canadian animal use by categories of invasiveness from 2002-2007

Between the years 2002-2007 an average of 2.3 million animals were used each year. The numbers of animals in the most severe category (E) fluctuate between 55,121 at the lowest point (in 2003) and 179,781 at the highest point (in 2006). However, category E remains the category with the lowest number of animals. There has also been an increase in the numbers of animals in Category D since 2002. This may be attributed to the fact that in Canada all newly created genetically engineered animals are put into Category D until a stable phenotype has been established.

niques are refined to minimize harmful impacts on animals.

Even before publication of animal use numbers by CI, beginning in 1996 in Canada, there was a growing concern for the large number of animals that faced the potential for significant pain and distress. This has been addressed through a variety of mechanisms, for example the development of score sheets to assist in the early identification of clinical signs of pain and disease as a means of encouraging intervention before animals experienced extreme pain and or distress (Morton and Griffiths, 1985), encouragement to use anesthetics and analgesics to relieve pain and distress (Flecknell, 1994), as well as efforts to refine procedures known to cause considerable pain (see for example Hem et al., 1998 – alternatives to retro-orbital bleeding). In Canada the considerable number of animals reported in CI-D & CI-E prompted the development of CCAC guidelines on: choosing an appropriate endpoint in experiments using animals in research, teaching and testing (1998). The years following the publication of this guidelines document did see a reduction in the number of animals reported in CI-E, reflecting the increased level of monitoring of animals and ACCs confidence that, due to this closer monitoring and use of clinical signs, animals would no longer experience extreme pain and distress before the experiment was terminated (Gauthier, 2004)

It is important to note that the Canadian system of requiring investigators and ACCs to take a precautionary approach in assigning CIs prospectively gives enhanced protection to the animals but has the unfortunate corollary of skewing animal use data to suggest that the worst case scenario has actually been experienced. A case in point is the assignment of CIs for genetically engineered animals. The CCAC guidelines on: transgenic animals (CCAC, 1997) require investigators to assign protocols where GEAs are to be created to CI-D. Although the guidelines offer the opportunity for ACCs to reassign the CI after a year, based on an animal welfare assessment of the new animal line, in reality ACCs have continued to count and report the number of animals on GEA protocols as CI-D. This has had the effect of overinflating numbers of animals in CI-D (Fig. 1), creating the false impression that there is a substantial increase in the number of animals experiencing considerable levels of pain and distress in Canada. Efforts are currently underway with the development of new CCAC guidelines on: genetically engineered animals used in science (in preparation) to address this inaccuracy.

Although the CCAC guidelines on: choosing an appropriate endpoint in experiments using animals for research, teaching and testing (CCAC, 1998) had an immediate impact on the number of animals in CI-E, this decline did not continue and, furthermore, there was an actual increase in numbers of animals in CI-E in the years 2004-2006. Closer examination revealed that most of the animals used in this category were for the purpose of regulatory testing, prompting the CCAC to focus on the implementation of the Three Rs in regulatory testing as one area for study by the CCAC research fellows in animal policy development. Guy and Griffin (2009) examined the opportunities and obstacles for implementation of the Three Rs in shellfish toxin testing, identifying several areas where federal government policy (including the implementation of validated non-animal methods) could have a significant impact on the numbers of animals potentially experiencing the highest levels of pain and distress. This study follows the intention first expressed by Smyth (1978) that categorization of animal experiments into various impact categories could be used as a driver for Replacement.

3 Emerging challenges to the current Canadian animal impact lower case for categorization system

3.1 The need to accurately reflect the number of animals that have actually experienced the various levels of pain and/or distress

There is no doubt that retrospective reporting of CI would increase public accountability and ensure that publicly reported information more accurately reflects the pain and distress experienced by animals (i.e. actual impacts vs. expected impacts); however, this needs to be weighed against the administrative burden placed on institutions, investigators, and ACCs. This burden may prove to be acceptable, however, as a recent study by the Laboratory Animal Science Association in the UK proposed that administrative burden could be acceptably minimized if all statistics on animal use were reported retrospectively and if appropriate training initiatives were in place to aid investigators in transitioning to the new system (Animal Procedures Committee 2008). It is important to note that in Canada some ACCs do now report CI data based on actual observation of the animals as a result of enhanced post-approval monitoring and subsequent re-evaluation of assigned CIs.

3.2 Categorizing based on animal-outcomes rather than solely procedure-based

Currently, the Canadian categorization system is focused mainly on procedures, with indications of animal outcomes; for example: "CI-C *Experiments which cause minor stress or pain of short duration*. Possible examples: cannulation or catheterization of blood vessels or body cavities under anesthesia; minor surgical procedures under anesthesia, such as biopsies, laparoscopy; short periods of restraint beyond that for simple observation or examination… Such procedures should not cause significant changes in the animal's appearance, in physiological parameters such as respiratory or cardiac rate, or fecal or urinary output, or in social responses." (CCAC, 1991)

Retrospective reporting of CIs might also assist in further understanding the animal outcomes for particular procedures. For procedures where the outcome is unknown, the inability to assign a CI accurately could be used as an indication that a pilot study is needed in order for an ACC to be able to ascertain the likely impact on the animal.

3.3 Categorizing harms to animals in breeding colonies, especially animals bred specifically to have a disease condition

Revising the CCAC guidelines on: transgenic animals (as CCAC guidelines on: genetically engineered animals used in science; CCAC, in preparation) has triggered consideration of how to categorize animals that are bred to have a particular con-

dition that affects their welfare. Once a GEA has been created, the impact of the modification of its genome on its phenotype potentially affects the manner in which it is able to interact with its environment. For many GEA lines, there may be no obvious difference from the wild-type. However, for others there may be outcomes (either intended, because they are designed to mimic a disease, or unintended as a result interaction with other genes), that interfere with welfare. For the animal it does not matter whether the negative welfare state is a result of genetic engineering or of in-breeding; what is important is the level of impairment. There is currently discussion about how this might be reflected in the Canadian categorization system, with the ultimate goal to ensure that these adverse impacts are recognized, minimized, and mitigated.

3.4 Categorizing impacts to animals that occur at different times in animals' life cycle, especially categorizing GEAs at different stages of new animal line generation

Attempts to accurately identify the pain and distress experienced by an animal have led to discussions about whether impact categorization should more appropriately reflect the life-time experience, or at least the protocol-life experience of the animal. Nowhere is this of greater relevance than for GEAs. In the UK, being born a GEA is considered to be a procedure, and assigned to a particular severity band, recognizing that the animal may experience negative welfare due to having a harmful mutation. If the animal is then used in a protocol – for example to elucidate the impact of the genetic modification on its phenotype, these procedures may impose an added burden on the welfare of the animal, over and above what might be experienced by a "conventional" animal of the species. In addition, there could be instances, particularly for larger animals where the need to maintain a defined health status dictates housing the animal in an impoverished environment, where a certain level of distress is experienced by the animal due to the inability to carry out important behaviors. A more accurate reflection of these life-time experiences within the current Canadian system is emerging as a particular challenge, where guidance for ACC members is badly needed.

Finally, with respect to GEAs, it remains to be seen how the Canadian CI system can address animal experiences where there has been a change to the animal's telos (i.e. the manner in which it expresses its interactions with its environment (Rollin, 1998).

4 Conclusions and animal welfare implications

The Canadian system of animal impact categorization has been in place for twenty years. During that period it has undergone a number of changes in an attempt to better reflect the outcomes of procedures for the animals. But has it achieved any success? Based on our case study, we believe that it has. We believe that, assigning CIs to animal use protocols has, over this 20-year period, sensitized Canadian scientists to the likely pain and distress experienced by the animals involved in their work.

CIs also have provided an important tool for ACCs in evaluating the ethical acceptability of animal use protocols. It is generally recognized that protocols with a higher CI require additional levels of scrutiny, both during the approval itself and during the conduct of the protocol. A combination of assigned CIs, post approval monitoring, and annual reviews of protocols are tools that ACCs can use to improve animal welfare outcomes.

In providing good governance for the ethical use of animals in science, the CCAC is accountable to the Canadian public (Schuppli and McDonald, 2005). This requires attention to providing accurate, meaningful animal use statistics, including information about CIs. In addition, the opportunity to reflect on animal use policy as a result of analysis of animal use data relating to the various CIs has led to guideline documents and other initiatives that are focused on improving animal welfare.

The ongoing evolution of science continues to present new challenges to old policies. The Canadian CI system is undergoing evaluation to better reflect the impact of animal-based procedures that were not envisioned when the CI scale was implemented. Part of this evaluation will include an examination of how best to reflect animal outcomes, as well as consideration of how the Canadian system can harmonize with other animal impact categorization systems worldwide.

References

Animal Procedures Committee (APC) (2008). Final report of a LASA/APA Working Group to examine the feasibility of reporting data on the severity of scientific procedures on animals. http://www.apc.gov.uk/reference/lasa_apc_final_report.pdf

Bowd, A. D. (1997). The educative role of an animal care committee in Canada: A case study. *Animal Welfare 6*, 309-315.

Canadian Council on Animal Care (CCAC) (in prep). CCAC guidelines on: genetically engineered animals used in science. Ottawa ON: CCAC.

Canadian Council on Animal Care (CCAC) (2003). CCAC guidelines on: the care and use of wildlife 70pp. Ottawa ON: CCAC. Available at http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/Guidelis.htm

Canadian Council on Animal Care (CCAC) (1999). CCAC guidelines on: institutional animal user training program and a Recommended Syllabus for an Institutional Animal User Training Program 12pp. Ottawa ON: CCAC. Available at http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/Guidelis.htm

Canadian Council on Animal Care (CCAC) (1998). CCAC guidelines on: choosing an appropriate endpoint in experiments using animals for research, teaching and testing 9pp. Ottawa ON: CCAC. Available at http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/GDLINES/Guidelis.htm

Canadian Council on Animal Care (CCAC) (1997). CCAC guidelines on: transgenic animals 9pp. Ottawa ON: CCAC. Available at http://www.ccac.ca/en/CCAC_Programs/Guide-

lines_Policies/GDLINES/Guidelis.htm
Canadian Council on Animal Care (CCAC) (1991). CCAC policy statement on: categories of invasiveness in animal experiments. Ottawa ON: CCAC. Available at http://www.ccac.ca/en/CCAC_Programs/Guidelines_Policies/POLICIES/CATEG.HTM
European Union (2009). Expert working group on severity classification of scientific procedures performed on animals. Final Report Conducted in support of the revision of Directive 86/609/EEC on the protection of animals used for scientific purposes pp15. European Commission: Brussels http://ec.europa.eu/environment/chemicals/lab_animals/pdf/report_ewg.pdf.
Flecknell, P. A. (1994). Refinement of animal use – assessment and alleviation of pain and distress. *Laboratory Animals 28*, 222-231.
Gauthier, C. (2004). Overview and analysis of animal use in North America. *Alternatives to Laboratory Animals 32, Suppl. 1*, 275-285.
Griffin, G., Dansereau, M. and Gauthier, C. (2007). Categories of invasiveness – A precautionary approach. *AATEX 14*, 715-720. Available at http://altweb.jhsph.edu/wc6/paper715.pdf.
Guy, A. L. and Griffin, G. (2009). Adopting alternatives for the regulatory monitoring of shellfish for paralytic shellfish poisoning in Canada: Interface between federal regulators, science and ethics. *Regulatory Toxicology Pharmacology 54*, 256-263.
Hem, A., Smith A. J. and Solberg, P. (1998). Saphenous vein puncture for blood sampling of the mouse, rat, hamster, gerbil, guinea pig, ferret and mink. *Laboratory Animals 32*, 364-368.
Morton, D. B. and Griffiths, P. H. M. (1985). Guidelines on the recognition of pain and discomfort in experimental animals and an hypothesis for assessment. *Veterinary Record 116*, 431-436.
Obrink, K. J. (1982). Swedish law on laboratory animals. In W. J. Dodds and F. B. Orlans (eds.), *Scientific Perspectives on Animal Welfare* (55-58). New York: Academic Press.
Orlans, F. B. (2000). Public policies on assessing and reporting degrees of animal harm; international perspectives. In M. Balls, A. M. van Zeller, and M. E. Halder (eds), *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (1075-1082). Amsterdam: Elsevier Science B.V.
Orlans, F. B. (1993). Animal pain scales in public policy: an international perspective. *In In the Name of Science: Issues in Responsible Animal Experimentation* (118-127). New York: Oxford University Press.
Orlans, F. B. (1987). Research protocol review for animal welfare. *Investigative Radiolog 22*, 253-258.
Orlans, F. B. (1980). Animal welfare. *BioScience 30*, 144-145.
Orlans, F. B. (1979). Animals in Education: Use of Animals in High School Biology Classes and Science Fairs. In H. McGiffin, and N. Brownley (eds), *The Use of Animals in High School Biology Classes and Science Fairs*. September of 1979. Washington, DC: The Institute for the Study of Animal Problems.
Rollin, B. (1998). On telos and genetic engineering. In Ed. B. Rollin, A. Holland & A. Johnson (eds), *Animal Biotechnology and Ethics* (156-171). New York: Chapman & Hall.
Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. 238 pp. London, UK: Methuen. (Reissued: 1992, by Universities Federation for Animal Welfare, Potters Bar, Herts, UK.) Available at: http://altweb.jhsph.edu/publications/humane_exp/het-toc.htm
Schuppli, C. A. and McDonald, M. (2005). Contrasting modes of governance for the protection of humans and animals in Canada: lessons for reform. *Health Law Review 13*, 97-106.
Schuppli, C. A. (2004). The Role of the Animal Ethics Committee in Achieving Humane Animal Experimentation. Ph.D.Thesis. *The University of British Columbia*, 1-162.
Smyth, D. H. (1978). *Alternatives to Animal Experiments*. London UK: Scolar Press.

Acknowledgements

The authors gratefully acknowledge the following people for their assistance in the preparation of this paper: Wendy Clarence, Julie Dale, Catherine Schuppli, and Anne Thibaudeau.

Correspondence to

Gilly Griffin
Canadian Council on Animal Care
1510-130 Albert St.
Ottawa, Ontario
Canada, K1P 5G4
e-mail: ggriffin@ccac.ca

Should Mammalian Fetuses be Excluded from Regulations Protecting Animals during Experiments?

David J. Mellor[1,2], Tamara J. Diesch[1] and Craig B. Johnson[1,3]

[1]Animal Welfare Science and Bioethics Centre, [2]Institute of Food, Nutrition and Human Health and [3]Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand

Summary

The definition of "animal" in welfare regulations would usually include life stages where suffering may occur through the conscious perception of noxious stimuli (e.g. feeling pain), and would exclude those stages when, at the time the regulations were framed, the young were presumed to lack the capacity for conscious perception. Recent re-evaluations of the literature and our related research on mammalian young suggest that states of unconsciousness, which preclude suffering, persist to developmental stages beyond those where the young are protected under such regulations. These observations challenge the basis of regulations framed in these terms and raise several questions. Should such regulations now be changed to better reflect our new understanding of developmental neurobiology? Would an innate drive or desire to care for and protect vulnerable young override such scientific knowledge? Is the precautionary principle of ensuring that mammalian young are protected well before there is any likelihood that they have reached a developmental stage when suffering could occur sufficient to maintain the status quo?

Keywords: fetuses, newborns, pain, regulatory protection

1 Introduction

During the recent past the extent to which mammalian fetuses may experience sensations, including pain, has been re-evaluated. The conclusion, based on integrating previously uncollated high quality scientific evidence widely distributed in the literature and on our own more recent research, is that conscious experience of sensations does not normally occur until after birth in any mammalian young observed to date (Mellor and Gregory, 2003; Lee et al., 2005; Mellor et al., 2005, 2009; Mellor and Diesch, 2006; Diesch et al., 2008). A rather similar, but qualified, conclusion has been drawn regarding the possibility that consciousness may not occur until after hatching in domestic chickens (Mellor and Diesch, 2007), but this will not be addressed here. Such conclusions have relevance to the definition of "animal" in those animal welfare laws that focus on the possibility of an organism suffering as a result of the conscious perception of noxious stimuli (e.g. feeling pain). That is because those definitions usually exclude developmental stages when, at the time the legislation was framed, the young were presumed to lack the capacity for conscious perception. For instance, mammalian fetuses are protected under the New Zealand Animal Welfare Act 1999 after the first half of pregnancy has elapsed, and marsupial pouch young are protected from immediately after birth.

The purposes of the present paper are, first, to very briefly summarise the main conclusions about the onset of consciousness during development of mammalian young and the timing of that in relation to birth. The second is to evaluate the possibility that invasive stimuli applied before the onset of conscious perception could subsequently increase an animal's sensitivity to pain and suffering after the capacity for consciousness has developed. The third is to consider evidence for a greater drive or desire among both women and men to care for and protect vulnerable young, and to highlight the proposition that this may influence their attitude to regulatory protection of fetuses independently of scientific evidence suggesting that such protections may not be necessary. Finally, consideration will be given to the potential impact of these insights on the utility and appropriateness of some facets of the current definition of "animal" in welfare legislation, using the specific details of the New Zealand Animal Welfare Act 1999 (Anonymous, 1999) as an example.

As most of the scientific background to the notion that the onset of consciousness occurs only after birth in mammalian young has been published in the comprehensive reviews noted above and in other publications referred to by Mellor et al. (present proceedings) in the present proceedings, the reader is referred to these articles for more details.

2 Neurological development and the onset of consciousness – a summary

The general pattern of neurological development appears to be rather similar in most mammals, irrespective of when the capacities for sentience and conscious perception first appear in relation to the timing of birth (see Mellor et al., present proceedings). Major features are as follows. Anatomically, initially rudimentary neural tissue develops progressively in its size, differentiation and interconnectivity into the complex, sophisticated and multi-structured nervous system of the adult. Paralleling this are changes in behaviour such that initial "startles" or

jerky whole-body movements progress through poorly coordinated individual limb, neck or head movements to increasingly well-coordinated and controlled movements or changes in body position or activity.

Accompanying these anatomical and behavioural changes is a progression of electrical states in the brain, those in the cerebral cortex (measured as the electroencephalogram or EEG) being of particular note because functional maturation of the cortex is considered to be an essential prerequisite of conscious perception. Pre-cortical and cortical structures are electrically "silent" initially – i.e. there is no activity in the EEG. The EEG then exhibits sporadic very short epochs of activity, which evolve into progressively longer periods of sustained activity against correspondingly shorter intervening periods of electrical "silence". Continuous mixed sleep-like EEG activity then appears and this subsequently matures into differentiated and alternating rapid-eye-movement (REM) and non-REM sleep-like patterns. Around the time of REM-non-REM differentiation, neural connections, which are considered to be essential for consciousness, become well established between sub-cortical brain structures and the cerebral cortex. Finally, EEG patterns indicating repetitive sleep-wake cycles are the last to appear, and, as noted below, these usually appear only after birth.

Mammals may be assigned to three groups on the basis of their stage of neurological maturity at birth; namely those that are exceptionally immature, moderately immature or mature. Newborn marsupial joeys are neurologically exceptionally immature at birth, as they have rudimentary precortical structures that are electrically silent. Changes in their EEG and behaviour support the view that they have no capacity to consciously experience any sensations before or during birth and for several months after birth. Newborn rat and mouse pups, rabbit kits, kittens and puppies are neurologically moderately immature at birth. Immediately after birth they exhibit "silent", intermittent or continuous but undifferentiated EEGs, and thereafter it takes several days for REM-non-REM differentiation to occur and sleep-wake cycles to become established. Using as a basis the *first* appearance of differentiated REM-non-REM EEG patterns, it is conservatively concluded that the earliest onset of conscious perception in these species would occur at 4-14 days after birth. Neurologically mature newborns include calves, fawns, kids, lambs, foals, piglets and guinea-pig pups, and indeed human infants. They evidently develop the capacity for sentience by about 75-80% of the way through pregnancy, but do not exhibit conscious perception before or during birth because of initial neurological immaturity and subsequently because of the operation of a set of *in utero* neuroinhibitors that are unique to fetal life. After birth, withdrawal or reduction of this neuroinhibition and an onset of potent neuroactivation together lead to the gradual appearance of consciousness, which, as judged behaviourally, occurs over several minutes or hours.

Thus, neuroanatomical, neurophysiological and behavioural evidence suggests that all mammalian young examined to date remain in unconscious states before and during birth, and that the onset of conscious perceptions only occurs postnatally, taking months, days, or minutes to hours to appear, depending on the species (see Mellor et al., present proceedings).

3 Potential for pre-conscious induction of greater post-conscious pain sensitivity

Although the pre-conscious organism is not able to *experience* pain, invasive procedures will stimulate nociceptors (pain receptors) and thereby will cause impulse barrages in those nociceptive nerve tracts that have developed at the time the procedure is conducted. These impulse barrages cause withdrawal reflexes and other behavioural responses, stress hormone release and changes in the rates of blood flow to the brain and other organs during and shortly after the invasive procedures are conducted (Lee et al., 2005; Mellor et al., 2005). However, none of these responses requires an intact cerebral cortex as they are elicited by neural mechanisms below the level of the cerebral cortex (Lee et al., 2005; Mellor et al., 2005; Mellor and Diesch, 2006, 2007). Nevertheless, they raise the possibility that early noxious stimulation might initiate responses in the developing nervous system resulting in long-term changes that persist until after the onset of consciousness. There is an increasing body of evidence in the human literature that children subjected to noxious stimulation as pre-term infants, or during the early postnatal period in full-term infants, may demonstrate increased perception of pain lasting for many months or years (Taddio et al., 1997; Grunau, 2000, 2006; Buskila et al., 2003). A recent study in our laboratory has suggested that such lasting hyperalgesia might also occur in sheep exposed to the noxious stimulus of castration within 12 hours of birth (McCracken et al., 2006). Whilst as yet unproven, the possibility of such long-term hyperalgesia warrants further study.

4 Emotional responses to fetuses and newborns

Notwithstanding the above scientific considerations suggesting that fetuses remain unconscious and therefore cannot suffer by experiencing noxious sensory inputs such as pain, another factor may give impetus to a desire to provide regulatory protection to mammalian fetuses used for scientific purposes. It relates to the emotional impact that vulnerable young have on human beings.

Darwin (1872) recognised that something about human infants causes adults to respond to and care for them in order to increase survivorship of their own offspring. This innate drive or desire to care for and protect the newborn is likely based on our ability to recognise certain physical and behavioural features of infants that are considered to trigger "innate releasing mechanisms" for affection and nurturing responses in adult humans (Lorenz, 1943; Fullard and Reiling, 1976; Gould, 1980; Morreall, 1991; Morris et al., 1995). These features include a large rounded forehead, large, low-lying and wide-set eyes, shortened nose or muzzle, short and thick limbs, clumsy move-

ments and playfulness (Lorenz, 1943, 1971). Infant helplessness can therefore be instinctively recognised ensuring care for offspring (Bowlby, 1957, 1969) and ultimately survival of the species. Recently, a neurophysiological basis for this has been recognised in adult humans (both female and male) who evidently exhibit highly specific and salient brain activity within a seventh of a second in response to infant faces but not to adult faces (Kringelbach et al., 2008). It is apparent, however, that our emotional responses to such neonatal features are not restricted to human offspring, as the young of some other animal species also elicit them. Hence, kittens, puppies and other newborn animals evoke in us an urge to care and nurture, as do our own children. Furthermore, we have selectively bred many cat and dog breeds to retain juvenile characteristics into adulthood (neoteny), making them apparently more attractive as a result (Lawrence, 1989; Archer, 1997; Stafford, 2006). In like manner, anecdotally, some animals appear to be more tolerant of interference from very young human infants than from older children, although, if true, it is not clear whether this would be due to an impact on the animal of the infants' juvenile features, or to pheromones or other factors.

5 Implications for legal definitions of "animal"

The New Zealand Animal Welfare Act 1999 has as one of its major purposes the protection of animals from human actions that cause unreasonable or unnecessary pain, distress or other harm (Anonymous, 1999). The definition of "animal" in the Act is framed around this purpose and includes, among other animals, any mammal, bird or reptile at any age apart from the earliest developmental stages. Thus, the definition excludes the first half of gestation or development in mammalian fetuses or avian or reptilian pre-hatched young, respectively, and excludes the developmental stage up to and including birth, before the marsupial joey enters the maternal pouch. This was because of the presumption, at the time the Act was framed, that young beyond these stages have the potential to be conscious and experience pain and distress. The observations outlined here cast serious doubt on this presumption with regard to mammalian young. The situation is clearly more complex than was previously envisaged.

This raises the question of whether or not the regulations should be changed to reflect this new understanding. The answer, as recommended elsewhere (Mellor et al., 2009, and present proceedings), is that more time should probably be allowed to see if the conclusion that unconsciousness persists throughout fetal life in all mammals, and for months or days after birth in some of them, will survive direct experimental challenge, especially as this conclusion is based on newly configured evidence. Moreover, the current restrictions have merit, in a precautionary sense, in ensuring that mammalian young are protected before there is any likelihood that they have reached a developmental stage when suffering could occur, or where treatment during pre-conscious stages might eventually be shown to result in greater sensitivity to painful stimuli subsequently, once the capacity for consciousness has developed. They also harmonise with the apparent greater human desire to care for and protect vulnerable young, an emotional response, which may have impact quite independently of any scientific knowledge suggesting that such regulatory protections may not be required.

References

Anonymous (1999). *Animal Welfare Act 1999* (5-6, 48-67). Wellington: New Zealand Government Printer.

Archer, J. (1997). Why do people love their pets? *Evol. Hum. Behav. 18*, 237-259.

Bowlby, J. (1969). Attachment and loss. In *Vol 1: Attachment*. London: Hogarth Press.

Bowlby, J. (1957). An ethological approach to research in child development. *Brit. J. Med. Psychol. 30*, 230-240.

Buskila, D., Neumann, L., Zmora, E. et al. (2003). Pain sensitivity in prematurely born adolescents. *Arch. Pediat. Adol. Med. 157*, 1079-1082.

Darwin, C. (1872). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

Diesch, T. J., Mellor, D. J., Johnson, C. B. and Lentle, R. G. (2008). Responsiveness to painful stimuli in anaesthetised newborn and young animals of varying neurological maturity (wallaby joeys, rat pups and lambs). *AATEX 14, Spec. Issue*, 549-552.

Fullard, W. and Reiling, A. M. (1976). An investigation of Lorenz's "Babyness". *Child Dev. 47*, 1191-1193.

Gould, S. J. (1980). *The Panda's Thumb*. New York: WW Norton and Company.

Grunau, R. E. (2000). Long-term consequences of pain in neonates. In K. J. S. Anand, B. J. Stevens and P. J. McGrath (eds), *Pain in Neonates*. Amsterdam: Elsevier.

Grunau, R. E. (2006). Long-term consequences of pain in human neonates. *Semin. Fetal Neonat. Med. 11*, 268.

Kringelbach, M. L., Lehtonen, A., Squire, S. et al. (2008). A specific and rapid neural signature for parental instinct. *PLoS ONE 3*, e1664, 1-7.

Lawrence, E. A. (1989). Neoteny in American perceptions of animals. In R. J. Hoage (ed), *Perceptions of animals in American culture*. Washington DC: Smithsonian Institution Press.

Lee, S. J., Peter Ralston, H. J., Drey, E. A. et al. (2005). Fetal pain: a systematic multidisciplinary review of the evidence. *J. Am. Med. Assoc. 294*, 947-954.

Lorenz, K. (1943). Die angeborenen Formen möglicher Erfahrung. *Z. Tierpsychol. 5*, 235-409.

Lorenz, K. (1971). *Studies in Animal and Human Behavior, vol. II*. London: Methuen.

McCracken, L., Waran, N., Mitchinson, S. and Johnson, C. B. (2006). Does age of castration influence post-castration hyperalgesia? (Abstract). *Proceedings of IX World Congress of Veterinary Anaesthesiology*, p165.

Mellor, D. J. and Diesch, T. J. (2006). Onset of sentience: the potential for suffering in fetal and newborn farm animals. *Appl. Anim. Behav. Sci. 100*, 48-57.

Mellor, D. J. and Diesch, T. J. (2007). Birth and hatching: key events in the onset of "awareness" in lambs and chicks. *New Zeal. Vet. J. 55*, 51-60.

Mellor, D. J. and Gregory, N. G. (2003). Responsiveness, behavioural arousal and awareness in fetal and newborn lambs: experimental, practical and therapeutic implications. *New Zeal. Vet. J. 51*, 2-13.

Mellor, D. J., Diesch, T. J. and Johnson, C. B. (2010). When do mammalian young become sentient? *ALTEX*, present proceedings.

Mellor, D. J., Patterson-Kane, E. and Stafford, K. J. (2009). Integrated perspectives: sleep, developmental stage and animal welfare. In *The Sciences of Animal Welfare* (161-185). Oxford: Wiley-Blackwell Publishing.

Mellor, D. J., Diesch, T. J., Gunn, A. J. and Bennet, L. (2005). The importance of "awareness" for understanding fetal pain. *Brain Res. Rev. 49*, 455-471.

Morreall, J. (1991). Cuteness. *Brit. J. Aesthet. 31*, 39-47.

Morris, P. H., Reddy, V. and Bunting, R. C. (1995). The survival of the cutest: who's responsible for the evolution of the teddy bear? *Anim. Behav. 50*, 1697-1700.

Stafford, K. (2006). *The Welfare of Dogs*. Dordrecht: Springer.

Taddio, A., Katz, J., Ilersich, A. L. and Koren, G. (1997). Effect of neonatal circumcision on pain response during subsequent routine vaccination. *Lancet 349*, 599-603.

Acknowledgements

We are grateful to many colleagues for their helpful input during the development of these ideas. We also thank the Agricultural and Marketing Research and Development Trust for providing PhD funding for one of us (TJD), was well as the Geoffrey Gardiner Foundation (Australia), MAF Biosecurity New Zealand, and the Animal Welfare Science and Bioethics Centre and the Riddet Institute at Massey University for providing financial support for our related research.

Correspondence to

Prof. David J. Mellor, PhD
Animal Welfare Science and Bioethics Centre, IFNHH
College of Sciences, Massey University
Palmerston North 4442, New Zealand
e-mail: D.J.Mellor@massey.ac.nz

Reviewing the Reviews: An Analysis of the Process of Ensuring Regulatory Compliance in the Use of Animals in Science in New Zealand

Virginia M. Williams[1] and Linda A. Carsons[2]

[1]Mt Admiral, Cheviot, New Zealand; [2]Animal Welfare Directorate, MAF Biosecurity, New Zealand

Summary

In New Zealand, the Animal Welfare Act 1999 requires that organisations using animals in research, testing and teaching be audited at least every five years for compliance with both the Act and the organisations' individually approved codes of ethical conduct. This paper looks at the results of this ongoing review process since the introduction of the Act at the beginning of 2000, during which time most of the organisations have been reviewed at least twice. It includes an analysis of the non-compliance issues that have arisen.

Keywords: animal use in science, regulation, New Zealand, ensuring compliance

1 Introduction

In New Zealand, the use of animals in research, testing and teaching (RTT) is governed by a self-contained set of provisions contained within Part 6 of the Animal Welfare Act 1999 (the Act). While the central focus of the Act imposes obligations on owners or those in charge of animals to ensure that the physical, health and behavioural needs of the animal are met and, where practicable, ensure that when the animal is ill or injured it receives treatment that alleviates any unreasonable or unnecessary pain or distress, Part 6 stands separate from the rest of the Act, because it provides a process that, in some circumstances, sanctions manipulations that have the potential to cause suffering, distress, or compromised care. This recognises that the manipulation of a small number of animals may result in significant benefits to a wider group of people or animals, to society generally or to the environment.

However, society has required that the legislation include adequate safeguards governing such animal use. Any individual or organisation wanting to manipulate animals is subject to a comprehensive set of requirements. Firstly, they must hold a code of ethical conduct (CEC) approved by the Director-General of Agriculture (D-G) or be employed by a person or organisation that holds an approved CEC, or be employed by a person or organisation that has an approved arrangement to use another organisation's animal ethics committee (AEC); and secondly, each individual project must first be approved by an AEC appointed by the code holder.

Currently, there are 30 code holders in New Zealand, two of which each have three AECs, giving a total of 34 such committees. An additional 78 organisations do not have their own code or AEC but are "parented" by one of the 17 code-holding organisations that have chosen to take on this responsibility. The types of organisations holding CECs are given in the following table:

Tab. 1

Institution type	Number	Number of AECs
Crown Research Institutes (CRIs)	4	6
Universities	7	9
Commercial organisations	10	10
Institutes of Technology	6	6
Government departments	1	1
Schools	1	1
Other	1	1

The review process

As a further layer of scrutiny of the animal ethics system in New Zealand, code holders and their AECs must undergo periodic reviews by independently accredited reviewers, with the aim of assessing the extent to which the code holder and the AEC are both complying with the Act and the CEC as well as implementing the policies, procedures, and requirements set out in the Act and the CEC. A satisfactory review report is a prerequisite to obtaining approval of a CEC for a second or subsequent period.

Where a code holder holds an approved CEC for the first time, or where a person did not carry out RTT in the 2 years prior to obtaining their current CEC approval, the first independent review must take place within 2 years. Subsequent reviews must be completed before the term of approval of the current CEC has expired. For example, where the CEC was approved for the maximum term of 5 years, an independent review must be carried out within the first 2 years and again 3 years later. After that, expiry reviews take place at an interval determined as part of the approval – usually every 5 years.

Reviewers are accredited by the D-G, with due regard to their competency, their character or reputation, and their ability to

maintain an appropriate degree of impartiality and independence when conducting reviews. There are currently five reviewers, although there have been up to eight. All are veterinarians except one, who is a research scientist and Chair of an AEC.

The reviewer's draft report is initially sent to the code holder for comment, with the final report, along with any code holder comments, going back to the code holder, to the D-G and to the National Animal Ethics Advisory Committee (NAEAC), an advisory committee to the Minister of Agriculture on matters relating to the use of animals in RTT.

The report may notify a critical situation, key issues, key topics and/or recommendations. A critical situation is any situation which, in the judgement of the reviewer, places the code holder's, the AEC's or the Director-General's credibility at risk. If a critical situation is identified during a review, this must be immediately notified to the Director General, and could potentially lead to the suspension or revocation of the approval of a code of ethical conduct.

A key issue is defined as a non-compliance that, in the opinion of the reviewer, demonstrates a major non-compliance of the institution's procedures and policies with the Act, its regulations or the CEC. It may be a specific non-compliance or a system with multiple non-compliances having a cumulative effect. Key issues may be created by escalation of outstanding issues from previous reviews. Any key issues found will be discussed during the review and will appear in the review report. The Director General will require these to be resolved within an agreed time frame. The effectiveness of corrective actions will be measured in subsequent reviews.

A key topic is defined as a point of interest, which is discussed in the report. It may include positive and negative findings and may give rise to key issues or recommendations.

A recommendation is a suggestion aimed at improving the procedures and policies. Recommendations are non-binding.

The D-G, having received the final report, must then inform the code holder in writing whether the review indicates that a satisfactory level of compliance has been achieved. If compliance is unsatisfactory, the D-G must inform the code holder of the actions that must be taken in order to achieve a satisfactory level of compliance. If the code holder does not subsequently comply, or if the response is unsatisfactory, the D-G can decline approval of a new CEC or can revoke the existing CEC.

2 Method

This study analysed all reviews carried out since the introduction of the Act at the beginning of 2000 until the end of 2008, with emphasis on the results of such reviews, on the resultant actions required where non-compliance was identified and on the main areas of non-compliance.

3 Results

Sixty reviews of 34 AECs were undertaken during the assessment period. Of these, 45 were expiry reviews, including 16 second term expiry reviews. Six reviews were undertaken of institutions with new CECs and AECs, and nine follow-up reviews were required as the result of unsatisfactory levels of compliance. The results are given in Table 2.

The compliance of twenty-one out of 29 institutions (72%) was judged satisfactory at their first expiry review, and of 13 out of 16 (81%) at their second expiry review. Five out of six institutions (83%) with new codes of ethical conduct and CECs had satisfactory reviews, while all reviews that were required following identification of areas of non-compliance were satisfactory.

During the period surveyed, there were 10 institutions whose level of compliance was found to be unsatisfactory, one of these on two occasions. Of these, three institutions were not required to undergo re-review, one because it was amalgamating with another institution, one because it chose not to renew its CEC but to be parented by another organisation, and the third because

Tab. 2

Year	1st expiry	Result		2nd expiry	Result		New	Result		Follow-up	Result	
		✔[1]	✖[2]		✔	✖		✔	✖		✔	✖
2002	8	6	2									
2003	10	7	3				2	2		2	2	
2004	9	6	3				1		1	3	3	
2005										2	2	
2006										2	2	
2007	1	1		7	5	2	2	2				
2008	1	1		9	8	1	1	1				
Totals	**29**	21	8	**16**	13	3	**6**	5	1	**9**	9	

[1] Satisfactory

[2] Unsatisfactory

the key issue, although serious enough to warrant mention, had already been resolved by the time of the review. One institution was required to be reviewed annually for three years.

Of the institutions with an unsatisfactory level of compliance, two were universities (one twice), four were Institutes of Technology, three were commercial organisations and one was a CRI.

Two institutions that were found satisfactory on their first expiry review were found unsatisfactory on their second expiry review.

Critical situations identified

There were no critical situations identified during the survey period.

Key issues identified

Key issues were identified in 12 of the 60 reviews and fell into three main groups – AEC procedures, monitoring and animal facilities.

1. AEC procedures
   a. AEC not appointed in accordance with the code
   b. Lack of documented AEC procedures and approvals
   c. Lack of adherence to maximum approval periods
   d. Projects proceeding without AEC approval
   e. Lack of compliance with CEC and AWA in criteria for approval
   f. Lack of compliance for stated standards for meeting frequency, quorum, minute taking
2. Monitoring of compliance
   a. Failure to monitor compliance with approved protocols
3. Facilities
   a. Inadequate ventilation
   b. Temperature controls inadequate resulting in higher than acceptable temperatures.

Key topics identified

Forty six of the 60 reviews identified key topics. In 28 reviews, positive key topics were identified. For nine of these, only positive topics were raised, with five of the nine being assessments at follow-up reviews. Examples of positive key topics include:

– Commendation of animal care and enrichment;
– Commendation of the focus on animal welfare;
– Commendation of excellent recording and quality control systems;
– Commendation for inclusion of a biometrician on the AEC.

Negative key topics fell into seven categories:

1. AEC processes complying with some but not all CEC requirements (noted in 26 reviews). Examples:
   a. Procedures for dealing with complaints not included;
   b. Endpoints not well defined.
2. Lack of documentation of AEC processes (noted in 14 reviews). Examples:
   a. No formal detailing of AEC processes;
   b. No clear process for managing records.
3. Inadequacy of the monitoring process (noted in 15 reviews). Examples:
   a. Lack of a formal monitoring process;
   b. Lack of visits to view procedures by external members.
4. Inadequate documentation of monitoring by the AEC (noted in three reviews). Examples:
   a. Applicants not required to report back at the end of their study;
   b. Monitoring reports not written or filed.
5. Inadequate facilities (noted in five reviews). Examples:
   a. Rodent cages inappropriate;
   b. Lack of hygiene on panels round ventilation vents;
   c. Inadequate cage washer resulting in an unacceptable hygiene risk.
6. A lack of facility and animal care documentation in the form of standard operating procedures (noted in four reviews).
7. 11 Miscellaneous issues (noted in 11 reviews). Examples:
   a. A lack of staff familiarity with the CEC;
   b. Inadequate attendance at AEC meetings of some members.

Recommendations

As recommendations are non-binding, they tend to be suggestions from the reviewer for increasing quality and efficiency in the AEC process. Examples include:

1. The animal programme manager should be a member of the AEC;
2. Acclimatisation of indoor sheep is recommended;
3. A review of meeting frequency is recommended;
4. An internal audit process is suggested;
5. The formalisation of monitoring procedures as an SOP is suggested;
6. An improvement in the clarity of the application form is suggested.

Rereviews

The terms of reference for follow-up reviews are set by the MAF Director of Animal Welfare and vary depending on the initial review result. Outstanding issues are required to be remedied within a given timeframe, with time limits set for follow-up reviews. Follow-up reviews are required to be performed by the same reviewer who did the initial non-compliant review. In all bar one case, a single follow-up review was required within at most a year of the unsatisfactory review, with the focus being on those areas of non-compliance. In one case,

Tab. 3

Reviewer	No. of reviews (excluding follow-ups)	No. satisfactory	No. unsatisfactory
1	2	2	-
2	2	1	1
3	1	1	-
4	23	19	4
5	10	10	-
6	1	-	1
7	11	7	4
1 + 4	1	-	1

three follow-up reviews were required at yearly intervals. The first and third of these were focused only on identified non-compliant areas, while the second was a full review.

Reviewers

Reviews were carried out by seven reviewers during the sample period, with one review being conducted jointly by two of the seven.

The majority of reviews were performed by three reviewers – 23 (38% of total), 11 (18%) and 10 (17%) respectively. Two of the seven reviewers had no unsatisfactory reviews – one had done only one review; the other had done 10.

4 Discussion

With the first round of expiry reviews being considered an educative as well as an auditing process, given the relatively recent introduction of the Act, it would be expected that a higher proportion of second expiry reviews would show satisfactory compliance. This was the case: 13 from 16 (81%) compared to the 21 from 29 (72%) at the first expiry review. However, three reviews are singled out for comment: in two cases (A & B) organisations were found satisfactory at the first review but not the second, and in another case (C), neither the first nor the second was deemed satisfactory.

A. A small organisation covering a limited number of animal-based activities, mainly in the training area. This organisation was reviewed by the same reviewer at first and second expiry reviews, with the cumulative effect of key topics from both reviews leading to the unsatisfactory result. As a result of the unsatisfactory review, this organisation has given up its CEC and is now parented by another institution.
B. An Institute of Technology training veterinary nurses, where once again the care of animals was commended. However, due to changing personnel combined with an inadequate transfer of knowledge of the animal ethics system, the current approval for use of animals had expired. This was rectified when identified and no further action was required.
C. In this case once again, attention to the welfare of the animals and to AEC processes was commended, but the non-compliances on both occasions related to structural problems with older buildings, with a relatively long timeframe allowed for rectifying the problems because of the considerable financial outlay required.

A point arising from B above is the relatively high proportion of Institutes of Technology where compliance has been unsatisfactory. Four from six (67%) have had unsatisfactory reviews compared to universities (27%), commercial organisations (33%) and CRIs (25%). There are several reasons that might account for the higher rate of unsatisfactory reviews:

1. The animal manipulations covered by these institutions are mainly concerned with the handling of animals by students in animal care or veterinary nursing. Such manipulations are regarded as having a relatively low impact on the welfare of animals, and for this reason, may be seen as less important to tutors who, from the very nature of their jobs, would see themselves as having a focus on animal welfare.
2. There is often a high turnover of staff in such institutions, with inadequate transfer of information, resulting in a poor understanding of the system.

MAF is currently reviewing the situation in relation to Institutes of Technology to find ways of circumventing these problems.

With a small number of reviewers, some of whom have performed very few reviews, an emphasis is placed on ensuring consistency of the review process. Procedures to improve consistency include:

– Provision of performance standards for reviews;
– Provision of a template for a checklist based on information provided to code holders in two documents: *the Guide to the Preparation of Codes of Ethical Conduct and the Good Practice Guide for the Use of Animals in Research, Testing and Teaching*;
– An annual teleconference where reviewers "meet" with MAF staff and members of the National Animal Ethics Advisory Committee to discuss the previous year's reviews including any new issues raised as a result of those reviews;
– Regular update of reviewer documents as a result of the annual meetings as well as any other issues that have arisen;
– Three-yearly audit of the performance of the reviewers themselves.

In conclusion, the analysis of reviews between 2002 and 2008 shows that in the majority of cases (24 from 34 organisations) a satisfactory level of compliance with the Act and CEC was achieved. For those where compliance was adjudged unsatisfactory, subsequent follow-up reviews showed that a satisfactory level of compliance had been achieved. Given these results, the review system would appear to be working well and should add confidence to the overall regulatory system for the use of animals in RTT in New Zealand.

Correspondence to
Dr. V. M. Williams
624 Conway Flat Rd
R.D. 4, Cheviot
North Canterbury 7384
New Zealand
e-mail: vwilliams@xtra.co.nz

Session EB4: Status report on ICATM

The International Cooperation on Alternative Test Methods (ICATM)

Marilyn Wind[1], David Blakey[2], Hajime Kojima[3], Joachim Kreysa[4] and William Stokes[5]

[1]US Consumer Product Safety Commission, Bethesda, Maryland, USA; [2]Environmental Health-, Science- and Research Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Canada; [3]JaCVAM, National Institute of Health Sciences, Tokyo, Japan; [4]ECVAM, IHCP, Joint Research Center, European Commission, Ispra, Italy; [5]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA

Summary

On April 27, 2009, Canada, the European Union, Japan, and the United States signed a Memorandum of Cooperation (MOC) on International Cooperation on Alternative Test Methods (ICATM). The agreement provides for enhanced cooperation and collaboration between four national validation organizations: the U.S. National Toxicology Program Interagency Center for the Evaluation of Alternative Methods and its Interagency Coordinating Committee on the Validation of Alternative Methods, the Japanese Centre for Validation of Alternative Methods, the European Centre for the Validation of Alternative Methods, and the Environmental Health Science and Research Bureau within Health Canada. The validation organizations developed a framework to promote harmonization of scientific recommendations on alternative toxicity testing methods in response to a 2007 charge from the International Cooperation on Cosmetics Regulation (ICCR). The MOC, which implements the framework adopted by ICCR in 2008, provides for enhanced cooperation in three critical areas: validation studies, independent scientific peer review, and development of harmonized test method recommendations for regulatory consideration. By communicating and working together, the ICATM validation organizations will identify and embrace scientifically sound and robust test methods that will protect human and animal health and the environment and that will also reduce, refine, and replace the use of animals in testing.

Keywords: ICATM, validation, regulation, international harmonization, alternative test methods

1 Introduction

On April 27, 2009, representatives from four international agencies signed a Memorandum of Cooperation (MOC) establishing the International Cooperation on Alternative Test Methods (ICATM). The agreement promotes enhanced international cooperation and coordination on the scientific validation of non- and reduced-animal toxicity testing methods. The test methods evaluated under this agreement are expected to be more readily accepted by regulatory agencies by assuring international agreement on the scientific information demonstrating that the methods are reproducible and able to accurately identify product related health hazards. This paper describes the development of the MOC and the three major areas of cooperation covered by the agreement.

2 ICATM validation organizations

ICATM is a voluntary international cooperation of four validation organizations from the United States, Japan, the European Union, and Canada. The four initial ICATM Validation Organizations are:

- Japanese Center for the Validation of Alternative Methods (JaCVAM)
- European Centre for the Validation of Alternative Methods (ECVAM)
- U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)
- Environmental Health Science and Research Bureau, Canada

The inclusion of other participants and their appropriate status will be decided by consensus by the participating validation organizations.

3 ICATM development

The impetus for the development of ICATM occurred during the first meeting of the ICCR in September 2007. The ICCR recognized the importance of reducing, refining, and replacing animal use in toxicity testing and recommended the strengthening of international collaboration and communication. In response to the ICCR recommendation, ICCVAM, ECVAM, JaCVAM, and Health Canada developed a framework to ensure a collaborative approach to this issue. ICCR noted that such efforts should be supported by scientific experts from the regulatory bodies. Figure 1 shows a timeline of the development and adoption of the ICATM framework leading to the signing of the MOC in April of 2009. The ICATM Memorandum of Cooperation was signed at the National Institutes of Health on April 27, 2009. Signers included Masahiro Nishijima, Ph.D., Director General, National Institute of Health Sciences, Ministry of Health, Labour, and Welfare of Japan; Linda S. Birnbaum, Ph.D., D.A.B.T., A.T.S., Director, National Toxicology Program and the National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services of the United States; Elke Anklam, Ph.D., Director, Institute for Health and Consumer Protection, Joint Research Centre, European Commission of the European Union; and David H. Blakey, D.Phil., Director, Environmental Health Science and Research Bureau, Safe Environments Programme, Healthy Environments and Consumer Safety Branch, Health Canada.

4 Coordination of alternative test method validation and evaluation activities

Collaborations between ICCVAM, ECVAM, and JaCVAM have existed and have steadily increased during the past ten years. However, coordination of interactions had been on an ad hoc informal basis. The lack of consistent coordination and the different processes used by the different validation organizations sometimes resulted in validation studies, peer reviews, and development of formal recommendations by one organization without adequate consultation and input from the others. This often led to test method recommendations by one organization that could not be considered by another organization without extensive additional review efforts, and wide variations in transparency and outcomes of peer review processes.

5 ICATM Memorandum of Cooperation

The purpose of the MOC is to promote consistent and enhanced voluntary international cooperation, collaboration, and communication among national validation organizations in order to:

- Further the optimal design and conduct of validation studies to support national and international regulatory decisions on the usefulness and limitations of alternative methods
- Further high quality independent scientific peer reviews of alternative test methods that incorporate transparency and opportunity for stakeholder involvement
- Enhance the likelihood of harmonized recommendations by validation organizations on the usefulness and limitations of alternative test methods for regulatory testing purposes
- Achieve greater efficiency and effectiveness by avoiding duplication of effort and leveraging limited resources
- Support the timely international adoption of alternative methods

The Memorandum of Cooperation (MOC) was signed at the National Institutes of Health on April 27, 2009 (ICATM, 2009; U.S. NIH, 2009). The goals of the MOC are to:

- Establish international cooperation in the critical areas of validation studies, independent peer review, and development of harmonized recommendations to ensure that alternative methods/strategies are more readily accepted worldwide; and
- Establish international cooperation necessary to ensure that new alternative test methods/strategies adopted for regulatory use will provide equivalent or improved protection for people, animals, and the environment, while replacing, reducing or refining (causing less pain and distress) animal use whenever scientifically feasible.

Development of ICATM Framework and Memorandum of Cooperation	
Sep 2007	• ICCR Initial Meeting – Brussels
Feb/Mar 2008	• Validation Organizations Develop Draft Framework
Apr 2008	• Validation Organizations Present Draft Framework to ICCR Working Group
Jul 2008	• Revised Draft Discussed at Second ICCR Meeting – Rockville, Maryland, U.S.A.
Oct 2008	• ICATM Framework Adopted by ICCR
Apr 2009	• ICATM Memorandum of Cooperation Signed by Canada, European Union, Japan, and the United States.

Fig. 1: The development of the ICATM Framework was an interactive process involving consensus development among all four validation organizations.

The MOC addresses three critical areas of cooperation: validation studies, independent peer review of the validation status of test methods, and the development of formal test method recommendations on alternative testing methods. By coordinating the validation, peer review, and development of recommendations, ICATM will reduce differences between the organizations in recommendations on the usefulness of alternative methods for regulatory purposes. This should accelerate international adoption of scientifically valid alternative test methods.

Critical area #1 – Test method validation studies
The objective of ICATM cooperation on test method validation studies is to share information and develop consensus on critical scientific aspects of validation studies prior to their initiation. Participants will collaborate and seek consensus on the proposed validation study design, study protocol, and selection of reference substances to be tested. Data developed in such studies are more likely to be usable by all members and meet the needs of their regulatory authorities, reducing the cost and time wasted in duplication of efforts. (Fig. 2)

Critical area #2 – Independent Scientific Peer Review Meetings and Reports
The organization and conduct of Independent Scientific Peer Review Meetings is a critical, but timely and costly component of evaluating alternative test methods. ICATM members have agreed to consider the needs of all ICATM Validation Organizations when organizing and conducting such meetings.

Specifically, member organizations have agreed to share and seek input from each other during the preparation of background review documents and draft recommendations and to make these publically available when provided to peer review panel(s). Peer review panels will have international representation including solicitation of nominations from other ICATM Validation Organizations.

ICATM is also committed to making the review process as open and transparent as possible by holding public peer review meetings or providing other opportunities for stakeholder and public comment and by making peer review panel reports available to the public and to ICATM Validation Organizations to consider in developing final recommendations.

Critical area #3 – Development of harmonized test method recommendations
The last critical area of ICATM cooperation is the development of harmonized test method recommendations to forward to regulatory authorities for acceptance decisions. Ultimately, the most expeditious international adoption of new alternative test methods can be accomplished when there is agreement on test method recommendations by all of the national validation organizations. In order to achieve such harmonization, the ICATM framework provides for the cooperation between ICATM Validation Organizations in the preparation of both draft and final recommendations. Draft final recommendations will be shared within ICATM and will be considered along with the peer review panel report(s) and other supporting documents. Members will notify each other of their respective draft position. In cases where all of the ICATM Validation Organizations mutually agree, each organization will finalize and forward their recommendations to their respective regulatory authorities as authorized by applicable law. If there are unresolved disagreements among the ICATM Validation Organizations, the scientific rationale for these disagreements will be documented and provided by each validation organization with recommendations.

Harmonization of recommendations prior to regulatory consideration is expected to reduce differences in test methods adopted by various countries. This in turn will reduce issues created by differing regulatory guidelines, thereby facilitating more rapid adoption of new alternative test methods internationally.

6 Conclusions

The development of the ICATM framework and signing of the MOC has set the stage for a high level of transparency and the opportunity for broad stakeholder involvement during validation, peer review meetings, and preparation of final recommendations. ICATM participants are committed to consist-

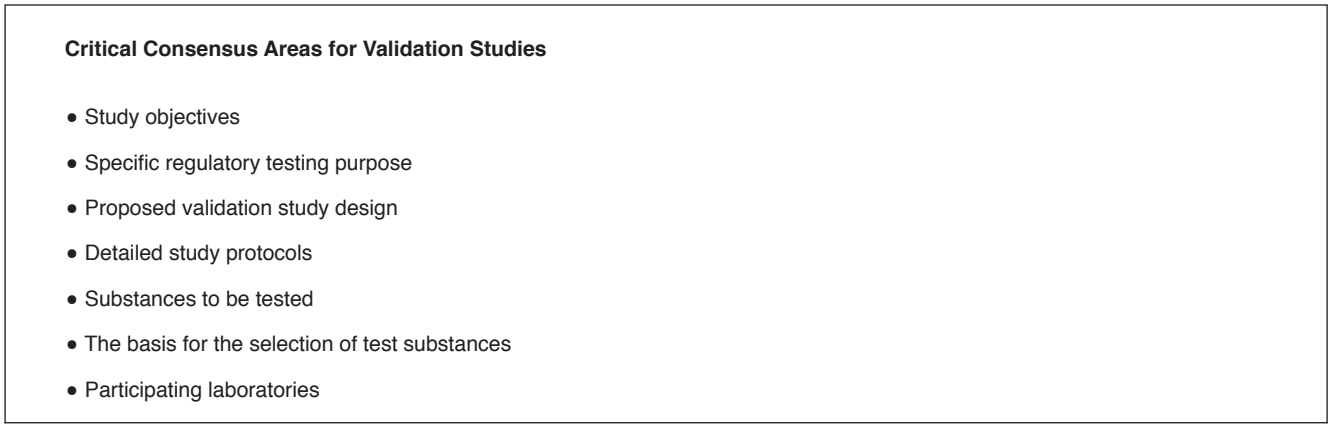


Fig. 2: Critical areas for ICATM Validation Organizations to reach consensus prior to initiation of validation studies.

ent coordination, cooperation, and communication to achieve success in the adoption of scientifically valid alternative test methods. Success will be indicated by consensus on the usefulness and limitations of new alternative methods, followed by rapid national and international acceptance.

We are encouraged that the new level of cooperation and coordination outlined in the ICATM agreement will ensure that new test methods provide for equivalent or better protection of people, animals, and the environment, and the reduction, refinement, and replacement of animals where scientifically feasible.

References

ECVAM (2009). About ECVAM. Available at: http://ecvam.jrc.it/ (accessed 20 August 2009).

Health Canada (2007). Environmental Health, Science and Research Bureau. July 26, 2007. Available at: http://www.hc-sc.gc.ca/ahc-asc/branch-dirgen/hecs-dgsesc/sep-psm/ehsrb-bsser-eng.php (accessed 20 August 2009).

ICATM (2009). Memorandum of Cooperation on International Cooperation on Alternative Test Methods (ICATM). April 27, 2009. Available: http://iccvam.niehs.nih.gov/docs/about_docs/ICATM-MOC.pdf (accessed 20 August 2009).

ICCVAM (2009). The NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). July 24, 2009. Available at: http://iccvam.niehs.nih.gov/ (accessed 20 August 2009).

JaCVAM (2009). Japanese Center for the Validation of Alternative Methods. Available at: http://jacvam.jp/ (accessed 20 August 2009).

U.S. FDA (2008). Framework for International Cooperation on Alternative Test Methods (ICATM). April 30, 2009. Available at: http://www.fda.gov/InternationalPrograms/HarmonizationInitiatives/ucm114518.htm (accessed 20 August 2009).

U.S. NIH (2009). Countries Unite to Reduce Animal Use in Product Toxicity Testing Worldwide: U.S., Canada, Japan and European Union Sign International Agreement. Press Release, April 27, 2009. Available at: http://www.nih.gov/news/health/apr2009/niehs-27.htm (accessed 20 August 2009).

Acknowledgements

The authors wish to acknowledge the representatives from ICCVAM, ECVAM, JaCVAM, and Health Canada who contributed to the development of the ICATM framework. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

Dr. William S. Stokes
Director, NICEATM
National Toxicology Program
National Institute of Environmental Health Sciences
P.O. Box 12233, Mailstop K2-16
Research Triangle Park, North Carolina
27709, U.S.A.
e-mail: stokes@niehs.nih.gov

NICEATM and ICCVAM Participation in the International Cooperation on Alternative Test Methods

William S. Stokes[1] and Marilyn Wind[2]

[1]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, Research Triangle Park, North Carolina, USA; [2]U.S. Consumer Product Safety Commission, Bethesda, Maryland, USA

Summary

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) work to promote the validation and regulatory acceptance of new, revised, and alternative test methods that will provide continued or improved protection of people, animals and the environment while reducing, refining, and replacing the use of animals. On April 27, 2009, the United States, Canada, European Union, and Japan signed a Memorandum of Cooperation on International Cooperation on Alternative Test Methods (ICATM). The agreement provides for regular collaborations of NICEATM and ICCVAM with the Japanese Center for the Validation of Alternative Methods (JaCVAM), the European Centre for the Validation of Alternative Methods (ECVAM), and Health Canada's Environmental Health Science and Research Bureau during validation studies, scientific peer reviews, and development of harmonized recommendations for new test methods. NICEATM and ICCVAM have implemented processes to support efficient international collaborations in order to achieve more rapid international acceptance of harmonized scientifically valid test methods.

Keywords: NICEATM-ICCVAM, ICATM, international harmonization, alternative test methods

1 Introduction

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is an interagency committee consisting of 15 U.S. Federal regulatory and research agencies that use, generate, require, or disseminate safety-testing information (Fig. 1).

The committee was established in 1997 to coordinate the interagency evaluation of the scientific validity of new, revised, and alternative methods proposed for regulatory safety testing. The ICCVAM Authorization Act of 2000 (USC, 2000) established ICCVAM as a permanent interagency committee of the NIEHS under NICEATM with specific purposes and duties (Fig. 2). NICEATM and ICCVAM work collaboratively to evaluate new and improved test methods applicable to the needs of U.S. Federal agencies (Stokes and Schechtman, 2007). Based on comprehensive scientific evaluations, ICCVAM and NICEATM forward formal recommendations to Federal agencies on test method usefulness and limitations for regulatory testing (Stokes and Hill, 2000, 2002). ICCVAM and NICEATM

ICCVAM Member Agencies	
Regulatory/Research	**Non-Regulatory/Research**
• Consumer Product Safety Commission	• Department of Defense
• Department of Agriculture	• Department of Energy
• Department of Interior	• National Cancer Institute
• Department of Transportation	• National Institute of Environmental Health Sciences
• Environmental Protection Agency	• National Institute for Occupational Safety and Health
• Food and Drug Administration	• National Library of Medicine
• Occupational Safety and Health Administration	• National Institutes of Health, Office of the Director
• Agency for Toxic Substances and Disease Registry	

Fig. 1: ICCVAM Member Agencies

work collaboratively to promote the validation and regulatory acceptance of new, revised, and alternative test methods that are based on sound science and that will provide continued or improved protection of people, animals and the environment while reducing, refining, and replacing the use of animals where scientifically feasible.

NICEATM administers ICCVAM and provides scientific and technical support for ICCVAM activities. NICEATM organizes test method peer reviews and workshops in conjunction with ICCVAM and carries out independent validation studies on high priority test methods (Stokes, 2003).

Since its establishment, ICCVAM has contributed to the evaluation of 27 alternative test methods that have been accepted or endorsed by national and international authorities. These methods provide alternatives that reduce, refine, or replace the use of animals for the hazard assessment of acute systemic toxicity (ICCVAM, 2001, 2006b), dermal corrosivity and irritation (ICCVAM, 2002), dermal phototoxicity, allergic contact dermatitis (ICCVAM, 2009a), ocular corrosivity and irritation (ICCVAM, 2006a), vaccine potency and safety testing, and pyrogenicity testing (ICCVAM, 2008a). ICCVAM completed additional evaluations for new alternatives for allergic contact dermatitis and ocular irritation in 2009 (ICCVAM, 2009b, 2009c). All of the materials associated with the ICCVAM evaluations, including background review documents, peer review panel reports, and ICCVAM test methods recommendations can be found on the NICEATM-ICCVAM website (http://iccvam.niehs.nih.gov).

2 International cooperation on alternative test methods

ICCVAM has long history of international collaborations with the European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM) (Fig. 3), and these collaborations have been steadily increasing (Schechtman and Stokes, 2002; Stokes et al., 2002).

On April 27, 2009, the United States, Japan, European Union, and Canada signed a Memorandum of Cooperation (MoC) establishing the International Cooperation on Alternative Test Methods (ICATM). The MoC provides for cooperation and collaboration in three critical areas: 1) the design and conduct of validation studies, 2) independent scientific peer review meetings and reports, and 3) development of harmonized test method recommendations for regulatory acceptance consideration (ICATM, 2009). Figure 4 summarizes the ICATM framework, progressing from validation studies through national and international regulatory acceptance of new alternative test methods.

ICCVAM's Purpose
• Increase the efficiency and effectiveness of U.S. Federal agency test method review
• Eliminate unnecessary duplicative efforts and share experience between U.S. Federal regulatory agencies
• Optimize utilization of scientific expertise outside the U.S. Federal government
• Ensure that new and revised test methods are validated to meet the needs of U.S. Federal agencies
• Replace, reduce, or refine the use of animals in testing, where feasible
ICCVAM's Duties
• Review and evaluate new, revised, or alternative test methods
• Facilitate interagency and international harmonization of test methods
• Facilitate and provide guidance on test method development, validation criteria, and validation processes
• Facilitate acceptance of scientifically valid test methods
• Submit test recommendations to U.S. Federal agencies
• Consider petitions from the public for review and evaluation of validated test methods

Fig. 2: The ICCVAM Authorization Act of 2000 defines ICCVAM's purpose and duties.
(Adapted from USC, 2000)

1993	• NIEHS Director serves as keynote speaker at First World Congress on Alternatives and Animal Use in the Life Sciences
1995	• Ad Hoc ICCVAM International Workshop on Validation and Regulatory Acceptance Criteria
1998	• ICCVAM holds its first international independent scientific peer review panel on the Murine Local Lymph Node Assay
2001	• First joint NICEATM – ECVAM International Validation Study (Cytotoxicity Assays estimate acute systemic toxicity
2002	• ECVAM liaison invited to participate in SACATM meetings
2003	• NICEATM and ICCVAM liaisons designated for ESAC
2007	• JaCVAM liaisons invited to participate in SACATM meetings
	• First NICEATM-ECVAM-JaCVAM International Validation Study (ER STTA Assay)
2009	• ICATM Memorandum of Cooperation signed

Fig. 3: Selected Highlights: NICEATM and ICCVAM international cooperation.

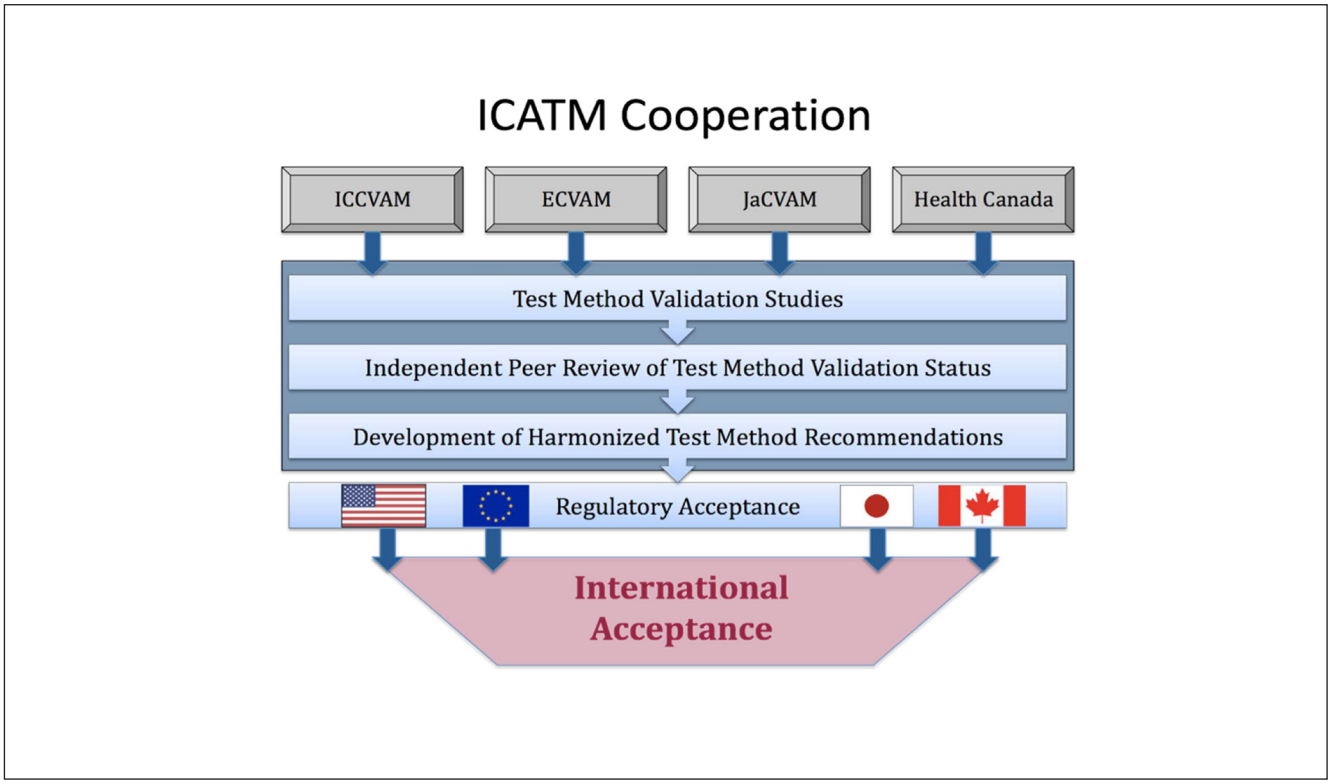


Fig. 4: ICATM framework for cooperation and collaboration between participating organizations.

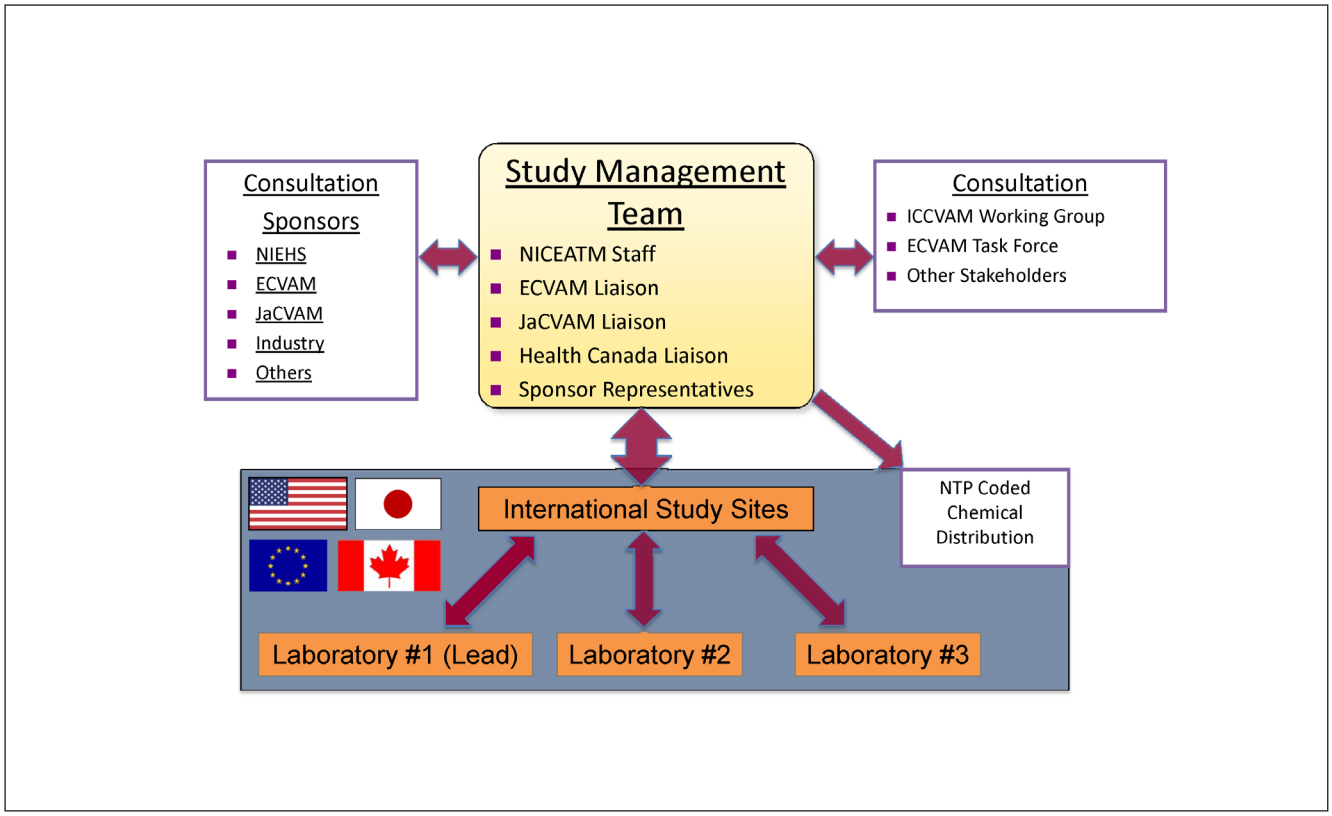


Fig. 5: The NICEATM-ICCVAM model for cooperative international validation studies includes active involvement and consultation with ICATM participants and stakeholders.

Prior to the development of the ICATM framework, collaborations among the national validation organizations were on an *ad hoc* informal basis. The lack of consistent coordination and the different processes used by each of the validation organizations often contributed to validation studies, peer reviews, and development of formal recommendations by one organization without adequate consultation and input from the others. This sometimes led to test method recommendations by one organization that could not be considered by another organization without extensive additional review efforts, and wide variations in transparency and outcomes of peer review processes.

The new cooperation agreement provides for communication and cooperation throughout the test method validation and evaluation processes. The ICATM agreement seeks to develop harmonized recommendations for scientifically sound test methods that will maintain or enhance protection of human and animal health and the environment, while reducing, refining, and replacing the use of animals where scientifically feasible.

3 International validation study cooperation

NICEATM and ICCVAM have developed a model for managing validation studies based on experience gained in managing two international validation studies and participating on the validation management teams for validation studies led by ECVAM and JaCVAM (ICCVAM, 2006b, 2006c; Stokes et al., 2008; Tice et al., 2009). In order to ensure early and consistent contributions from other ICATM participants, the Study Management Team (SMT) for NICEATM-led validation studies includes liaisons from ECVAM, JaCVAM, and Health Canada. (Fig. 5)

The SMT is charged with the development and approval of the validation study design, test method protocol, and selection of appropriate reference substances. The SMT consults with the appropriate ICCVAM Test Method Working Group to obtain regulatory and scientific input on these critical aspects. SMT liaison members from ECVAM, JaCVAM, and Health Canada

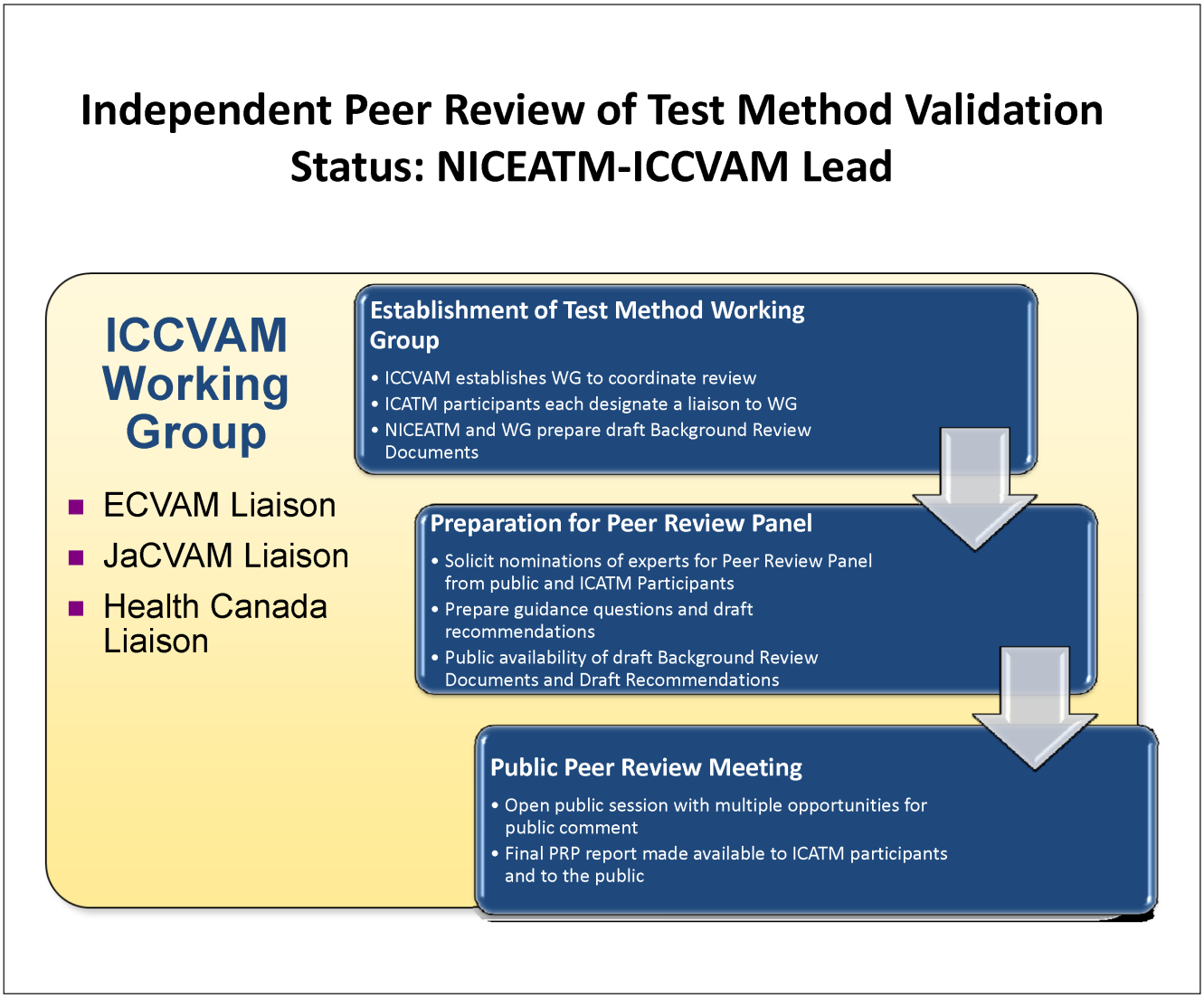


Fig. 6: NICEATM and ICCVAM Working Group Test Method Evaluation Process: Independent Scientific Peer Review.

are responsible for obtaining input from their appropriate regulatory and scientific stakeholders.

The NICEATM SMT is responsible for overall scientific coordination and management of the study in conjunction with the participating labs. The SMT evaluates the performance and results of the study at key points and determines any protocol optimization or modifications needed between study phases, and approves progression to the next phase of the study. ICATM liaisons on the SMT serve as a key conduit for information exchange throughout the study. After completion of the study, the SMT is responsible for data evaluation and interpretation, including review and approval of all reports resulting from the study.

NICEATM is currently using this model to manage an international validation study evaluating the LUMI-CELL® ER stably transfected transcriptional activation assay (ICCVAM, 2008b). ICCVAM and NICEATM liaisons also actively participated in the design and conduct of a recently completed ECVAM-led validation study on *in vitro* reconstructed human epidermis test methods and are participating in a similar role on three additional ECVAM-led validation studies: an *in vitro* testing strategy for skin sensitization (h-CLAT; DPRA; MUSST), an *in vitro* hepatic biotransformation enzyme induction assay, and human reconstituted tissue models for ocular irritation (i.e., EpiOcular™ and SkinEthic HCE™). ICCVAM and NICEATM also have liaisons contributing to two JaCVAM-led validation studies: the *in vivo/in vitro* comet assay and the BHAS cell transformation assay.

4 Independent scientific peer review of the validation status of test methods

NICEATM-ICCVAM evaluations of test method scientific validity are coordinated by a specific ICCVAM Working Group (WG) assisting NICEATM. The evaluation includes independ-

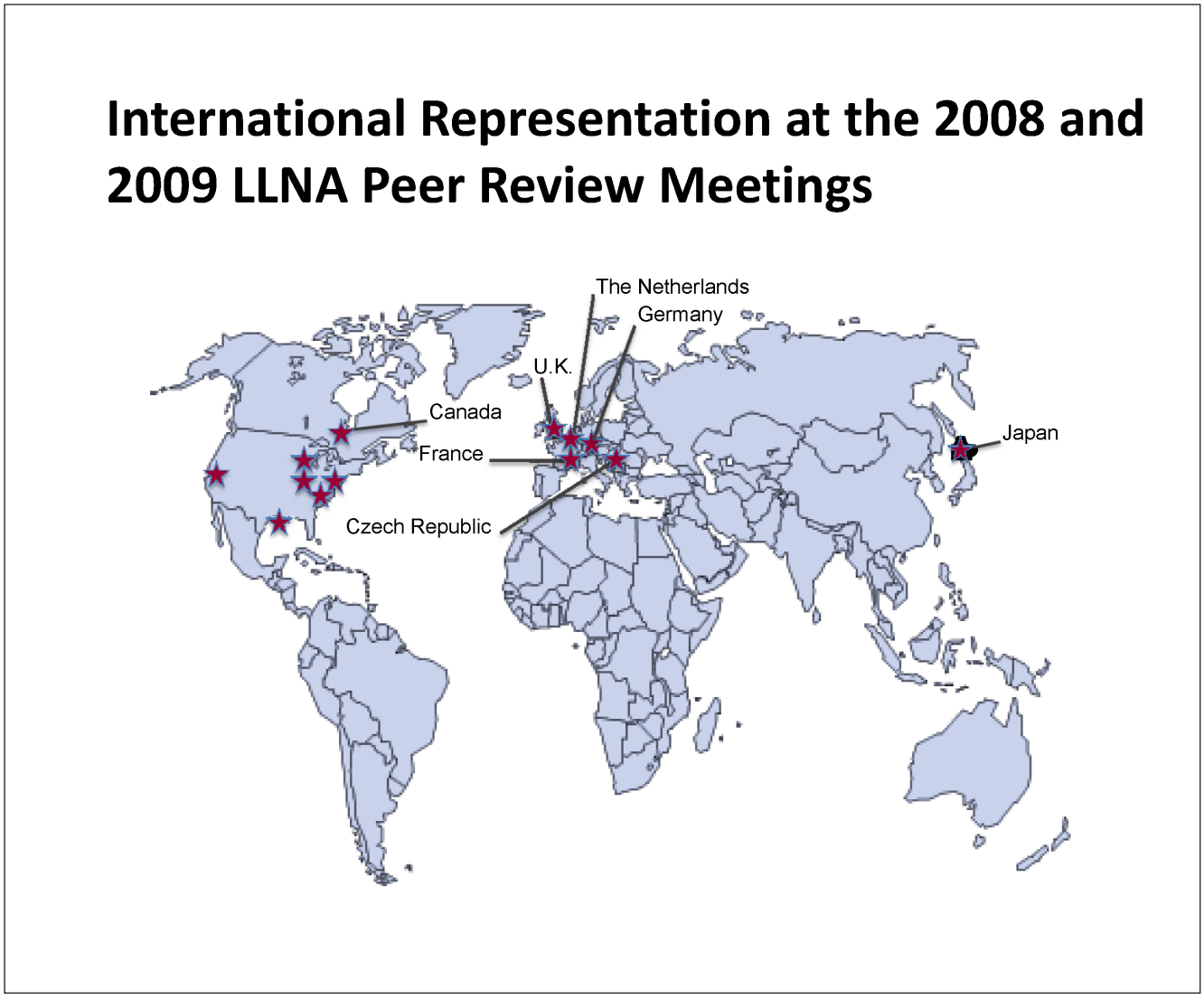


Fig. 7: The 2008 and 2009 International LLNA Peer Review Panels included members from eight countries.

ent scientific peer by an international panel of experts convened as an NIH Special Emphasis Panel in accordance with provisions of the Federal Advisory Committee Act (General Services Administration, 2001). Working groups are established for each toxicity area and consist of scientists from ICCVAM Federal Agencies and liaisons from the other ICATM participating organizations. The peer review panel consists of outside subject matter experts from industry, academia, other stakeholder groups, and the international scientific community. An important role of the ICCVAM Working Group is to recommend appropriate experts for the panel. ICATM liaisons are asked to identify appropriate international experts.

Figure 6 outlines the role of NICEATM and the ICCVAM Working Group in coordinating test method evaluations and independent scientific peer reviews. Key aspects of the process include public availability of all materials (i.e., draft background review documents and draft test method recommendations) made available to the peer review panel for review, public meetings of the peer review panel, the opportunity for public comments at the peer review panel meeting, and public availability of the panel's independent peer review report.

Two recent NICEATM-ICCVAM coordinated international peer reviews exemplify the commitment to international collaboration and transparency. On April 28-29, 2009, NICEATM-ICCVAM convened a Peer Review Panel Meeting on *Updated Validation Status of New Versions and Applications of the Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals and Products* (ICCVAM, 2009c). The ICCVAM Immunotoxicity Working Group included liaison members from both ECVAM and JaCVAM. In addition to other collaborations, the liaisons provided recommendations for panel membership. The combined panel membership from the 2009 peer review meeting and a similar peer review meeting held in 2008 included 19 members from eight countries: U.S., Canada, Europe (Czech Republic, France, Germany, The Netherlands, and U.K.), and Japan (Fig. 7).

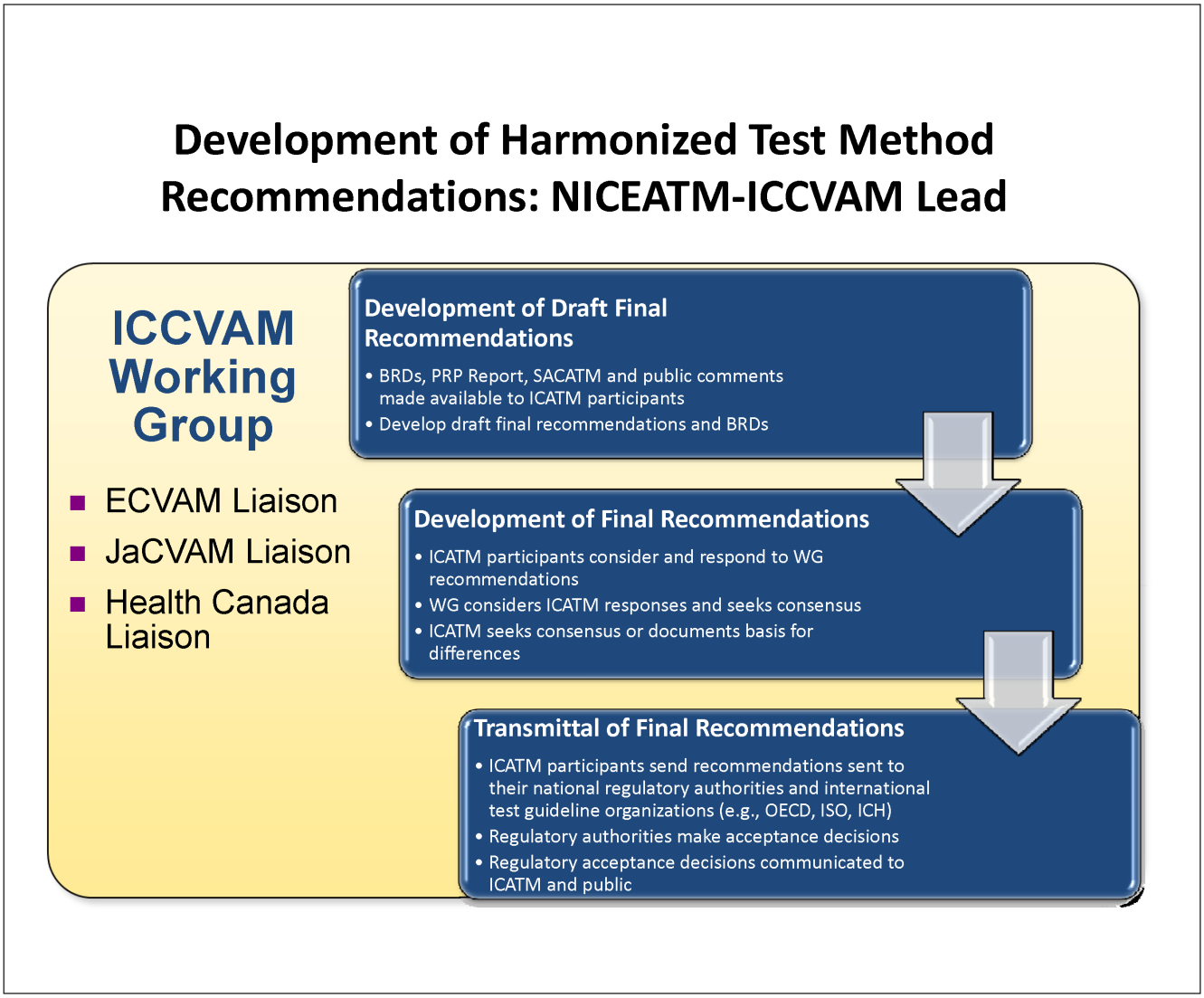


Fig. 8: ICCVAM's process for development of ICATM harmonized test method recommendations.

Similarly, on May 19-21, 2009, NICEATM-ICCVAM convened an international Peer Review Panel Meeting on *Alternative Ocular Safety Testing Methods and Approaches* (ICCVAM, 2009b). The ICCVAM Ocular Toxicity Working Group also included liaisons from ECVAM and JaCVAM. This panel was composed of 22 members, including members from six countries: Belgium, Canada, Japan, the Netherlands, Spain and the United States.

5 Development of harmonized test method recommendations

The third critical area of ICATM cooperation is the development of harmonized test method recommendations to forward to regulatory authorities for their acceptance consideration. Figure 8 summarizes the process for developing final test method recommendations. The ICCVAM WG, which includes the liaisons from the other ICATM participants, develops draft final test method recommendations after considering the peer review panel report and all comments received during the evaluation process from the public and from the NICEATM-ICCVAM Scientific Advisory Committee on Alternative Toxicological Methods (SACATM). The ICATM participants then consider and respond to the WG recommendations. The WG considers the ICATM participants' responses and works to develop a consensus on the final recommendations. If a consensus cannot be reached, then the scientific rationale for any areas where the validation organizations differ on their recommendations are documented in writing and included as part of the final test method evaluation report. The ICATM participants then forward the final test method recommendations and reports to their respective national regulatory authorities for acceptance decisions. Each ICATM participant then communicates to the other ICATM participants the regulatory authority acceptance decisions. The lead ICATM organization coordinating the evaluation then forwards recommendations, on behalf of all of the ICATM organizations, to appropriate international test guideline organizations, such as the Organization for Economic Coordination and Development (OECD), International Standards Organization (ISO), and the International Conference on the Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

Examples of recent harmonized test method recommendations include performance standards for the murine local lymph node assays (LLNA) and two *in vitro* ocular safety-testing methods (ICCVAM, 2006a; OECD, 2009a, 2009b). ICCVAM's recommendations on non-radioactive LLNA methods (LLNA: DA and LLNA: BrdU-ELISA) both resulted in new draft OECD Test Guidelines (OECD, 2009c, 2009d). The harmonized recommendations on LLNA performance standards and the reduced LLNA were incorporated into a draft update to OECD Test Guideline 429 submitted in June of 2009 (ICCVAM 2009a, 2009d; OECD, 2009e). Recommendations on the bovine corneal opacity and permeability and isolated chicken eye tests resulted in new OECD Test Guidelines 437 and 438, respectively, that were adopted by the OECD Council more rapidly than any previous test guideline (OECD, 2009a, 2009b).

6 Conclusions

The establishment of ICATM promotes enhanced international cooperation and collaborations on the scientific validation and evaluation of alternative test methods that are expected to further reduce, refine, and replace the use of animals in regulatory safety testing. As a result of this international cooperation, test methods determined to be sufficiently accurate and reproducible for identifying health hazards or safety of chemicals and products are expected to be more readily and rapidly accepted for regulatory testing by national and international organizations.

This international cooperation will serve an important role in translating research advances into more effective public health prevention tools. It will speed the validation and adoption of new test methods based on advances in science and technology that will provide more accurate predictions of safety or hazard. Animal welfare will also be improved by the national and international acceptance of alternative test methods that reduce, refine, and replace the use of animals.

References

General Services Administration (2001). Federal Advisory Committee Management; Final Rule. 41 CFR Parts 101-6 and 102-3. *Federal Register 66(139),* 37728-37750. July 19, 2001. Available: http://frwebgate4.access.gpo.gov/cgi-bin/PDFgate.cgi?WAISdocID=692694114379+0+2+0&WAISaction=retrieve (accessed 20 August 2009).

ICATM (2009). Memorandum of Cooperation between: the Japanese Center for the Validation of Alternative Methods, National Institute of Health Sciences, Ministry of Health, Labour, and Welfare of Japan; the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health Department of Health and Human Services, The United States of America; the European Centre for the Validation of Alternative Methods, Institute for Health and Consumer Protection, Joint Research Centre, European Commission of the European Union and the Environmental Health Science and Research Bureau, Safe Environments Programme, Healthy Environments and Consumer Safety Branch, Health Canada of Canada, Regarding International Cooperation on Alternative Test Methods (ICATM). April 27, 2009. Available: http://iccvam.niehs.nih.gov/docs/about_docs/ICATM-MOC.pdf (accessed 20 August 2009).

ICCVAM (2009a). ICCVAM Test Method Evaluation Report: The reduced murine local lymph node assay: an alternative test method using fewer animals to assess the allergic contact dermatitis potential of chemicals and products (NIH Publication 09-6439). Research Triangle Park, NC: National Institute of Environmental Health Sciences. March 1, 2009.

Available: http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-LD/TMER.htm (accessed 20 August 2009). (Referenced on page 3)

ICCVAM (2009b). Independent Scientific Peer Review Panel Report: Evaluation of the validation status of alternative ocular safety testing methods and approaches. Research Triangle Park, NC: National Institute of Environmental Health Sciences. July 2009. Available: http://iccvam.niehs.nih.gov/docs/ocutox_docs/OcularPRPRept2009.pdf (accessed 20 August 2009). (Referenced on page 3 and page 9)

ICCVAM (2009c). Independent Scientific Peer Review Panel Report: Updated validation status of new versions and applications of the murine local lymph node assay: A test method for assessing the allergic contact dermatitis potential of chemical and products. Research Triangle Park, NC: National Institute of Environmental Health Sciences. June 2009. Available: http://iccvam.niehs.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf (accessed 20 August 2009). (Referenced on page 3 and page 8)

ICCVAM (2009d). Recommended Performance Standards: Murine Local Lymph Node Assay. (NIH Publication 09-7357). Research Triangle Park, NC: National Institute of Environmental Health Sciences. January 1, 2009. Available: http://iccvam.niehs.nih.gov/methods/immunotox/PerfStds/llna-ps.htm (accessed 20 August 2009).

ICCVAM (2008a). ICCVAM Test Method Evaluation Report: Validation Status of five in vitro test methods proposed for assessing potential pyrogenicity of pharmaceuticals and other products (NIH Publication 08-6392). Research Triangle Park, NC: National Institute of Environmental Health Sciences. May 1, 2008. Available: http://iccvam.niehs.nih.gov/methods/pyrogen/pyr_tmer.htm (accessed 20 August 2009). (Referenced on page 3)

ICCVAM (2008b). Validation of Xenobiotic Detection Systems, Inc. LUMI-CELL® Assay of estrogenic activity. Research Triangle Park, NC: National Institute of Environmental Health Sciences. November 20, 2008. Available: http://iccvam.niehs.nih.gov/methods/endocrine/end_eval.htm (accessed 20 August 2009). Referenced on page 7)

ICCVAM (2006a). ICCVAM Test Method Evaluation Report: In vitro ocular toxicity test methods for identifying severe irritants and corrosives (NIH Publication 07-4517). Research Triangle Park, NC: National Institute of Environmental Health Sciences. November 11, 2006. Available: http://iccvam.niehs.nih.gov/docs/ocutox_docs/OTeval/OTevalrpt.pdf (accessed 20 August 2009).

ICCVAM (2006b). ICCVAM Test Method Evaluation Report: In vitro cytotoxicity test for estimating starting doses for acute oral systemic toxicity tests (NIH Publication 07-4519) Research Triangle Park, NC: National Institute of Environmental Health Sciences. November 1, 2006. Available: http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_tmer.htm (accessed 20 August 2009).

ICCVAM (2006c). Background Review Document: In Vitro cytotoxicity test methods for estimating acute oral systemic toxicity (NIH Publication 07-4518). November 1, 2006. Available: http://iccvam.niehs.nih.gov/methods/invitro.htm (accessed 20 August 2009).

ICCVAM (2002). Evaluation of EPISKIN™ , EpiDerm™ , and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: In vitro test methods for assessing dermal corrosivity potential of chemicals (NIH Publication 02-4502). Research Triangle Park, NC: National Institute of Environmental Health Sciences. July 22, 2002. Available: http://iccvam.niehs.nih.gov/docs/dermal_docs/cwgfinal02/cwgfinal.pdf (accessed 20 August 2009).

ICCVAM (2001). ICCVAM Test Method Recommendations on acute systemic toxicity (Appendix I of Workshop Report, NIH Publication 01-4499). Research Triangle Park, NC: National Institute of Environmental Health Sciences. August 12, 2001. Available: http://iccvam.niehs.nih.gov/docs/acutetox_docs/finalrpt/finappi2.pdf (accessed 20 August 2009).

OECD (2009a) OECD Guideline for the testing of chemicals: Test No. 437: Bovine corneal opacity and permeability test method for identifying ocular corrosives and severe irritants. September 7, 2009. Available: http://lysander.sourceoecd.org/vl=9145314/cl=24/nw=1/rpsv/ij/oecdjournals/16 07310x/v1n4/s54/p1 (accessed 15 September 2009).

OECD (2009b). OECD Guideline for the Testing of Chemicals: Test No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants. September 7, 2009. Available: http://oberon.sourceoecd.org/vl=2119939/cl=33/nw=1/rpsv/ij/oecdjournals/1607 310x/v1n4/s55/p1 (accessed 15 September 2009).

OECD (2009c). OECD Guideline for the Testing of Chemicals: Proposal for a new test guideline: Skin sensitisation: Local Lymph Node Assay: BrdU-ELISA. June 26, 2009. Available: http://www.oecd.org/dataoecd/63/7/43302529.pdf (accessed 20 August 2009). (Referenced on page 10)

OCED (2009d). OECD Guideline for the Testing of Chemicals: Proposal for a New Test Guideline: Skin Sensitisation: Local Lymph Node Assay: DA Version. June 26, 2009. Available: http://www.oecd.org/dataoecd/63/5/43302184.pdf (accessed 20 August 2009). (Referenced on page 10)

OECD (2009e). Draft proposal for an update to Test Guideline 429 Skin Sensitisation: Local Lymph Node Assay. June 26, 2009. Available: http://www.oecd.org/dataoecd/62/48/43302600.pdf (accessed 20 August 2009).

Schechtman, L. M. and Stokes, W. S. (2002). ECVAM-ICCVAM: Prospects for future collaboration, *ATLA 30, Suppl. 2*, 227-236. (Referenced on page 4)

Stokes, W., Casati, S., Strickland, J. and Paris, M. (2008). neutral red uptake cytotoxicity tests for estimating starting doses for acute oral toxicity tests. *Current Protocols in Toxicology, Suppl. 36*, 20.4.1-20.4.20.

Stokes, W. S. and Schechtman, L. M. (2007). Validation and regulatory acceptance of new, revised, and alternative toxicological methods: In A. W. Hayes (ed.), *Principles and Methods of Toxicology* (1103-1128). Philadelphia, Pennsylvania: Taylor and Francis. (Referenced on page 2)

Stokes, W. S. (2003). The interagency coordinating committee on the validation of alternative methods (ICCVAM): Recent progress in the evaluation of alternative toxicity testing methods. In H. Salem, S. A. Katz (eds.), *Alternative Toxicological Methods* (15-30). Washington, DC: CRC Press. (Referenced on page 3)

Stokes, W. S. and Hill, R. N. (2002). The Role of ICCVAM in Evaluating New and Alternative Test Methods. *Lab. Anim. (NY) 31, 26-32*. (Referenced on page 2)

Stokes, W. S., Schechtman, L. M. and Hill, R. N. (2002). The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM): A review of the ICCVAM Test Method Evaluation Process and International Collaborations with the European Centre for the Validation of Alternative Methods (ECVAM), *ATLA 30, Suppl. 2*, 23-32. (Referenced on page 4)

Stokes, W. S. and Hill, R. (2000). The role of the interagency coordinating committee on the validation of alternative methods in the evaluation of new toxicological testing methods. In Balls, M., van Zeller, A. M. and Halder, M. (eds.), *Progress in the reduction, refinement, and replacement of animal experimentation* (385-394). Amsterdam: Elsevier Science. (Referenced on page 2)

Tice, R., Deal, F., Ceger, P. et al. (2009). establishing a historical database for a multi-phased international validation study of a stably-transfected Estrogen Receptor (ER) Transcriptional Activation (TA) test method. 1806. *The Toxicologist 108(S-1),* 375.

USC (2000). Public Law 106-545, December 19, 2000. Available: http://iccvam.niehs.nih.gov/docs/about_docs/PL106545.pdf (accessed 20 August 2009).

Acknowledgements

The authors acknowledge David Allen, Steven Morefield, Cathy Sprankle, and Debbie McCarley for their expert assistance in the preparation of this manuscript. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

Dr. William S. Stokes
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park, North Carolina
USA 27709
e-mail: stokes@niehs.nih.gov

Lunch Sessions SL4

Recent Progress and Future Directions at NICEATM-ICCVAM: Validation and Regulatory Acceptance of Alternative Test Methods that Reduce, Refine, and Replace Animal Use

William S. Stokes[1] and Marilyn Wind[2]

[1]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA; [2]US Consumer Product Safety Commission, Bethesda, Maryland, USA

Summary

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) promote the validation and regulatory acceptance of new, revised, and alternative test methods that are based on sound science and that will provide continued or improved protection of people, animals, and the environment while reducing, refining, and replacing the use of animals where scientifically feasible. NICEATM administers ICCVAM and provides scientific and technical support, including the conduct of high priority validation studies. ICCVAM has coordinated or contributed to the evaluation of 27 alternative test methods that have been accepted or endorsed by national and international authorities. In 2008, NICEATM-ICCVAM developed a Five-Year Plan in conjunction with Federal agencies that promotes the use of new science and technologies to develop alternative methods and expands international collaborations. Implementation of the Five-Year Plan is expected to advance alternative methods that will support improved safety assessments while reducing, refining, and replacing animal use.

Keywords: ICCVAM, NICEATM, regulatory acceptance, alternative test methods, safety testing

1 Introduction

In the United States, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) coordinates interagency issues on alternative toxicological test method development, validation, regulatory acceptance, and national and international harmonization. The committee is responsible both for coordinating Federal interagency technical evaluations of new, revised, and alternative safety testing methods and for forwarding test method recommendations to U.S. Federal agencies (USC, 2000). The National Institute of Environmental Health Sciences originally established ICCVAM in 1997 (NIEHS, 1997). In 2000, ICCVAM was established by law as a permanent interagency committee of the National Institute of Environmental Health Sciences under the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) (USC, 2000). ICCVAM is composed of representatives from the 15 U.S. Federal regulatory and research agencies that require, use, or generate toxicological testing data and information (Tab. 1).

NICEATM administers ICCVAM and provides scientific and operational support for ICCVAM activities. NICEATM also conducts and coordinates international validation studies to evaluate the usefulness and limitations of potential new alternative test methods. NICEATM and ICCVAM work together to evaluate and promote the validation and regulatory acceptance of scientifically valid toxicological test methods based on sound science that will protect human and animal health and the environment while reducing, refining (decreasing pain and distress), and replacing animal use where scientifically feasible (Stokes and Schechtman, 2007). NICEATM and ICCVAM provide an essential framework for translating advances in science and

Tab. 1: ICCVAM Member Agencies

ICCVAM Member Agencies	
Regulatory/Research	**Non-Regulatory/Research**
• Consumer Product Safety Commission • Department of Agriculture • Department of Interior • Department of Transportation • EnvironICCVAM Member Agenciesmental Protection Agency • Food and Drug Administration • Occupational Safety and Health Administration • Agency for Toxic Substances and Disease Registry	• Department of Defense • Department of Energy • National Cancer Institute • National Institute of Environmental Health Sciences • National Institute for Occupational Safety and Health • National Library of Medicine • National Institutes of Health, Office of the Director

technology into standardized and adequately validated safety testing methods that can be used to protect and advance public health. This paper summarizes recent progress and future plans of NICEATM and ICCVAM in advancing new, revised, and alternative safety testing methods.

2 Progress in the regulatory acceptance of alternative methods

Since its establishment in 1997, ICCVAM has coordinated or contributed to the evaluation of 27 alternative test methods that have been accepted or endorsed by national and international authorities. The types of safety testing and available alternatives are listed in Table 2.

The alternative methods accepted by Federal agencies include 17 non-animal methods and 10 methods that use fewer animals and/or provide for refinement by reducing or avoiding potential discomfort. Most of the *in vitro* methods do not require the use of animals when positive results are obtained. Recommendations for additional alternatives for allergic contact dermatitis (ICCVAM, 2009a, 2009b, 2009d) and ocular irritation are in development (ICCVAM, 2009c, 2009e).

3 The NICEATM-ICCVAM Five-Year Plan

In response to requests from the Appropriations Committees of the U.S. House of Representatives and U.S. Senate, ICCVAM and NICEATM prepared a five-year plan to describe goals and objectives for the years 2008 through 2012 (ICCVAM, 2008b). This plan was developed in conjunction with Federal agency program offices and describes how NICEATM and ICCVAM will promote the research, development, translation, validation, and regulatory acceptance of alternative test methods that reduce, refine, and replace the use of animals in testing while maintaining scientific quality and protecting human health, animal health, and the environment.

The plan identified four key challenges that NICEATM and ICCVAM will address in conjunction with agency programs. The first challenge is to identify priority toxicity testing areas for the next five years and to conduct and facilitate activities in those areas. Currently, the four highest priority areas are ocular toxicity, dermal toxicity, acute systemic toxicity, and biologics potency and safety testing. Their priority is based on the potential for the tests to cause significant animal pain and distress and the fact that these are among the most commonly conducted safety tests. Other priority areas include immunotoxicity, endocrine disruptors, pyrogenicity, reproductive and developmental toxicity, and chronic toxicity/carcinogenicity testing.

The second challenge is to identify and promote research initiatives that are expected to support future development of innovative alternative test methods. These new methods might incorporate techniques such as high throughput screening, computer modeling, informatics, and biomarkers.

The third challenge is for NICEATM and ICCVAM to foster the acceptance and appropriate use of alternative test methods through outreach and communication. This will be accomplished by sponsoring and participating in workshops, NICEATM-ICCVAM website communications, and the development and publication of standardized test method protocols.

Lastly, ICCVAM and NICEATM will develop partnerships and strengthen interactions with stakeholders to facilitate meaningful progress. These efforts are expected to facilitate research, development, translation, validation, and regulatory acceptance of alternative methods that will reduce, refine, and replace animal use while maintaining scientific quality and the protection of human health, animal health, and the environment.

In June 2009, ICCVAM and NICEATM released an Implementation Plan for the NICEATM-ICCVAM Five-Year Plan (ICCVAM, 2008b). This working document describes how ICCVAM and NICEATM are implementing the strategies outlined in the Five-Year Plan and includes goals, specific objectives, and planned activities for implementation. The Implementation Plan also describes NICEATM and ICCVAM accomplishments since February 2008 that relate to the goals and objectives outlined in the Implementation Plan.

NICEATM and ICCVAM recently created an alternative test methods milestones page on the NICEATM-ICCVAM website to provide a comprehensive overview of available alternative methods and a quick summary of current validation and evaluation activities (http://iccvam.niehs.nih.gov/methods/mile-

Tab. 2: U,S. and international acceptance of alternative methods 1998-2009

Approved/Endorsed Alternative Safety Testing Method	Test Guidelines, Regulations, and Guidances	Other References
Acute Systemic Toxicity • Up and Down Procedure (UDP) • Fixed Dose Procedure (FDP) • Acute Toxic Class Method (ATC) • Acute Toxicity In Vitro Starting Dose Procedure, 3T3 cells • Acute Toxicity In Vitro Starting Dose Procedure, NHK cells	OECD TG 420 (OECD, 2001a) OECD TG 423 (OECD, 2001b) OECD TG 425 (OECD, 2008a) EPA OPPTS 870.1100 (EPA, 2002)	(ICCVAM, 2001a) (ICCVAM, 2001b) (ICCVAM, 2001c) (ICCVAM, 2006a) (Stokes et al., 2008)
Dermal Absorption • *In vitro* dermal absorption methods	OECD TG 428 (OECD, 2004a)	
Dermal Corrosivity and Irritation • CORROSITEX® Skin Corrosivity Test • EpiSkin™ Skin Corrosivity Test • EpiDerm™ Skin Corrosivity Test • SkinEthic™ Skin Corrosivity Test • Rat TER Skin Corrosivity Test	OECD TG 430 (OECD, 2004b) OECD TG 431 (OECD, 2004c) OECD TG 435 (OECD, 2006) 67/548 EEC	(ICCVAM, 1999a) (ICCVAM, 2002) (ICCVAM, 2004)
Dermal Phototoxicity 3T3 NRU Phototoxicity Test 3T3 NRU Phototoxicity Test: Application to UV Filter Chemicals	OECD TG 432 (OECD, 2004d) 67/548 EEC	(Spielmann et al. 1998)
Allergic Contact Dermatitis • Local Lymph Node Assay	OECD TG 429 (OECD, 2002) EPA OPPTS 870.2600 (EPA, 2003)	(Dean et al., 2001) (Haneke et al., 2001) (ICCVAM, 1999b) (ISO, 2002) (Sailstad et al., 2001)
Ocular Corrosivity and Irritation • Bovine Corneal Opacity and Permeability (BCOP) Test Method • Isolated Chicken Eye (ICE) Test Method	OECD TG 437 (OECD, 2009d) OECD TG 438 (OECD, 2009e)	(ICCVAM, 2006b) (ICCVAM, 2006c) (ICCVAM, 2006d) (ICCVAM, 2006e)
Pyrogenicity • Human Whole Blood/Interleukin-1B *In Vitro* Pyrogen Test • Human Whole Blood (Cryopreserved) Interleukin-1B *In Vitro* Pyrogen Test • Human Whole Blood/Interleukin-6 *In Vitro* Pyrogen Test • Human Peripheral Blood Mononuclear Cell/Interleukin-6 *In Vitro* Pyrogen Test • Monocytoid Cell Line Mono Mac 6/Interleukin-6 *In Vitro* Pyrogen Test	U.S. FDA (FDA, 2009) European Pharmacopoeia (EDQM, 2009a)	(ICCVAM, 2008a)
Vaccine Potency and Safety Testing • Use of Humane Endpoints in Animal Testing of Biological Products • Rabies Vaccine, Humane Endpoints • ELISA Test for Batch Potency Testing of Erysipelas Vaccines (refinement) • Relevance of the Target Animal Safety Test for Batch Safety Testing of Vaccines for Veterinary Use • ELISA Test for Batch Potency Testing of Human Tetanus Vaccines refinement • ToBI Test for Batch Potency Testing of Human Tetanus Vaccines refinement	9 CFR 117.4e 27 CFR 610.10 9 CFR 113.4 CVB Notice Number 04-09 (Hill, 2004) European Pharmacopoeia (EDQM, 2009b)	(Hendriksen et al., 1994) (Hendriksen, 2009) (Johannes et al., 2003) (Rosskopf-Streicher et al., 2001)

stones.htm). The page provides a comprehensive summary of the status of ongoing and completed NICEATM-ICCVAM alternative test method evaluation projects and projects to which NICEATM, ICCVAM, and agency scientists are contributing (ICCVAM, 2009f). The page also provides timelines for each project with links to more detailed information and relevant documents. Additional information about NICEATM and ICCVAM test method evaluation activities can be found at http://iccvam.niehs.nih.gov.

4 The International Cooperation on Alternative Test Methods

On April 27, 2009, a Memorandum of Cooperation for International Cooperation on Alternative Test Methods (ICATM) was signed by representatives of the United States, European Union, Canada, and Japan (ICATM, 2009). Signing the agreement were:

– Dr. Linda Birnbaum, Director, NIEHS and NTP, USA
– Dr. Masahiro Nishijima, Director, National Institute of Health Sciences, Japan
– Dr. Elke Anklam, Director, Institute of Consumer Protection and Health, JRC, European Commission
– Dr. David Blakey, Director, Health and Safety Bureau, Health Canada.

The agreement provides a framework for enhanced international cooperation, collaboration, and communication in three critical areas:

– Test method validation studies
– Independent peer review of the validation status of test methods
– Development of harmonized formal test method recommendations for regulatory authorities

The initial participating validation organizations are the NICEATM and ICCVAM, the Japanese Center for the Validation of Alternative Methods (JaCVAM), the European Centre for the Validation of Alternative Methods (ECVAM), and the Environmental Health Science and Safety Bureau within Health Canada. The cooperation among these organizations facilitated by this agreement is expected to accelerate international adoption of scientifically valid alternative test methods (Blakey, 2009; Kojima, 2009; Kreysa, 2009; NIEHS, 2009; Stokes and Wind, 2009; Wind et al., 2009).

5 Alternative methods for allergic contact dermatitis: new versions and applications of the murine Local Lymph Node Assay

NICEATM and ICCVAM convened an international independent scientific peer review panel in 2008 and 2009 to review several new versions and applications of the murine Local Lymph Node Assay (LLNA), an alternative test method for assessing the allergic contact dermatitis potential of chemicals and products. ICCVAM originally evaluated the scientific validity of the LLNA in 1998 (ICCVAM, 1999b; Dean et al., 2001; Haneke et al., 2001; Sailstad et al., 2001). U.S. Federal agencies accepted the ICCVAM recommendations that the LLNA was a valid substitute for the current guinea pig test methods used to assess allergic contact dermatitis and that it could be used to evaluate most but not all types of substances. ICCVAM also concluded that the LLNA has many advantages over the traditional test methods, including using fewer animals, eliminating the potential discomfort that can occur from substances that are sensitizers, and providing dose-response information. The LLNA was subsequently incorporated into national and international test guidelines for the assessment of skin sensitization (ISO, 2002; OECD, 2002; EPA, 2003).

Based on a nomination in 2007 from the U.S. Consumer Product Safety Commission, NICEATM and ICCVAM developed LLNA performance standards and evaluated the validation status of the following new versions and applications of the LLNA:

– Three modified non-radioactive versions of the traditional LLNA:
  • LLNA: BrdU-ELISA (Takeyoshi et al. 2001; OECD 2009b)
  • LLNA: BrdU-Flow Cytometry (MB Research Laboratories)
  • LLNA: DA (Idehara et al. 2008; OECD 2009a)
– The reduced LLNA (rLLNA; also referred to as the LLNA limit dose procedure)
– Use of the traditional LLNA to test mixtures, metals, and aqueous solutions
– Use of the LLNA for potency categorization

NICEATM and ICCVAM evaluated the new versions and applications in conjunction with ECVAM and JaCVAM. This included the development of internationally harmonized performance standards for the LLNA that can be used to more rapidly and efficiently determine the validity of nonradioactive and other modified versions of the LLNA (ICCVAM, 2008a, 2008b). The evaluations included two meetings of an international independent scientific peer review panel (Panel) in 2008 and 2009 (ICCVAM, 2009c).

The Panel agreed with ICCVAM's recommendations that the available data and test method performance support the use of two of the non-radioactive modified versions of the LLNA to identify substances as potential skin sensitizers and nonsensitizers, with certain limitations. The Panel concluded that the LLNA could be used to test any chemical or product for allergic contact dermatitis potential, including pesticides and substances such as fragrances and dyes, unless there are unique physiochemical properties associated with these materials that might affect their ability to interact with immune processes. The Panel also endorsed ICCVAM's revised protocol for the LLNA and draft LLNA performance standards and a reduced LLNA procedure.

ICCVAM submitted a revised Test Guideline (TG) 429 to the Organization for Economic Cooperation and Development (OECD) in June 2009 that includes the ICCVAM updated LLNA protocol, the reduced LLNA procedure, and harmonized performance standards (OECD, 2009c). The updated protocol provides for a 20% reduction in animal use, and the reduced protocol can decrease animal use by an additional 40%. IC-

CVAM, in collaboration with JaCVAM, also submitted draft OECD Test Guidelines for the LLNA:DA and the LLNA:BrdU-ELISA, the two non-radioactive methods endorsed by the Panel (OECD, 2009a, 2009b).

NICEATM and ICCVAM are also evaluating the application of *in vitro* methods and integrated decision strategies to the regulatory safety assessment of allergic contact dermatitis hazards. This includes validation of two *in vitro* cell culture-based methods, the human cell line activation test (h-CLAT) (Sakaguchi et al., 2006), and the Myeloid U937 Skin Sensitization Test (MUSST) assay (Ovigne et al., 2008), and a chemistry-based assay, the direct peptide reactivity assay (DPRA) (Gerberick et al., 2007). An ECVAM/JACVAM Study Management Team that includes ICCVAM and NICEATM liaisons is coordinating the validation studies. NICEATM is also evaluating the integration of various types of physical-chemical properties and *in vitro* data to determine decision algorithms that can be used for regulatory safety decisions. NICEATM and ICCVAM will also recommend *in vitro* assays that may help increase the accuracy of *in vitro* predictions of allergic contact dermatitis for inclusion in high throughput screening (HTS). NICEATM will then evaluate the HTS results to identify useful assays.

6 Ocular toxicity test method evaluation activities

Ocular safety testing is one of ICCVAM's top four priorities because it is one of the four most common safety tests required and therefore involves a significant number of animals. Such testing also has the potential to cause significant unrelieved pain and distress when chemicals and products cause eye damage (ICCVAM, 2008b). In 2008, two ICCVAM-recommended *in vitro* safety testing methods, the bovine corneal opacity and permeability (BCOP) and isolated chicken eye (ICE) test methods, were accepted by U.S. Federal agencies for identifying substances with the potential to cause severe or permanent damage to eyes without using live animals. The OECD Council formally adopted TGs for these test methods in 2009, with the BCOP described in TG 437 and the ICE in TG 438 (OECD, 2009a, 2009b). These methods can now be used worldwide to identify substances that may cause severe or irreversible eye damage. Positive results can be used for hazard classification without the need to use live animals, thereby avoiding the pain and distress that may have resulted if animals had been required.

NICEATM and ICCVAM organized an international independent scientific peer review panel that met in May 2009 to evaluate nine alternative methods and strategies for ocular safety testing. These included:

- The routine use of topical anesthetics, systemic analgesics, and humane endpoints to avoid and minimize pain and distress during *in vivo* ocular irritation testing (ICCVAM, 2009g).
- The use of the BCOP, the Cytosensor® Microphysiometer (CM), the ICE, the isolated rabbit eye, and the hen's egg test – chorioallantoic membrane test methods for identifying moderate and mild ocular irritants and for identifying substances that do not cause sufficient eye injury to require ocular hazard labeling (non-labeled category) (IIVS, 2008; ICCVAM, 2009h, 2009i, 2009j).
- The *in vivo* low volume eye test (ICCVAM, 2009k).
- Non-animal testing strategies using the BCOP, CM, and/or EpiOcular™ test methods to assess the eye irritation potential of antimicrobial cleaning products and to determine their appropriate U.S. Environmental Protection Agency ocular hazard classification (ICCVAM, 2009l).

Highlights of the Panel's conclusions and recommendations include:

- Topical anesthetics and systemic analgesics should routinely be used prior to all *in vivo* ocular irritancy testing. The Panel recommended an enhanced protocol of specific pain-relieving drugs and schedule of administration to effectively avoid or minimize discomfort.
- The BCOP and CM test methods could be used as screening tests to identify some products and substances that would not require hazard labeling for eye irritation. These methods will be considered for potential use in a "bottom-up testing approach" to identify substances that are not expected to cause sufficient injury to require classification as an ocular irritation hazard.
- Proposed non-animal testing strategies using three *in vitro* test methods to assess the eye irritation potential of antimicrobial cleaning products for EPA ocular hazard classification and labeling purposes appear promising. The Panel recommended that studies should be conducted to further characterize the *in vitro* test methods and that testing strategies should be designed in coordination with ICCVAM.

ICCVAM also provided recommendations to an ECVAM Scientific Advisory Committee peer review of four cell function-based assays proposed for the identification of substances that can cause mild to moderate ocular irritation and to identify substances that do not require ocular hazard labeling (Cytosensor® Microphysiometer, Flourescein Leakage, Neutral Red Release, Red Blood Cell Hemolysis).

Information on the NICEATM and ICCVAM ocular toxicity test method evaluations can be found on the NICEATM-ICCVAM website at http://iccvam.niehs.nih.gov/methods/ocutox/ocutox.htm. The report from the peer panel meeting is currently available with ICCVAM test method evaluation reports, including final ICCVAM test method recommendations, to be available in 2010.

In conjunction with ECVAM, ICCVAM is currently evaluating two other *in vitro* test methods for ocular irritation testing that are proposed for identifying substances that do not cause classifiable ocular irritation. An ECVAM-led SMT with NICEATM and ICCVAM liaison members is currently planning validation studies for these two methods, Epiocular™ and SkinEthic™ .

7 Acute systemic toxicity test method evaluation activities

Acute systemic toxicity testing is one of the top four NICEATM-ICCVAM priorities because it is the most common safety test

performed worldwide and because of the potential for significant pain and distress to animals when substances produce toxic effects (ICCVAM, 2008b). In 2008, ICCVAM forwarded recommendations on the use of two *in vitro* test methods for estimating starting doses for acute oral systemic toxicity studies (ICCVAM, 2006a; Stokes et al., 2008). ICCVAM recommended that these test methods should always be considered before using animals for acute oral toxicity testing and that the methods should be used where determined appropriate. These recommendations were accepted by U.S. Federal agencies, and data from the test methods should now be used in a weight-of-evidence approach for determining starting doses for *in vivo* studies. Using these *in vitro* methods where appropriate is expected to reduce the number of animals required for each acute toxicity study.

A draft guidance document describing use of these two *in vitro* test methods for estimating starting doses for acute oral systemic toxicity tests has been prepared and forwarded to OECD for consideration (OECD, 2009f). In addition, the ICCVAM Acute Toxicity Working Group provided comments on five new and revised OECD Test Guidelines and Guidance Documents for assessment of acute toxicity hazards (OECD, 2008b, 2008c, 2008d, 2009g, 2009h). These additional methods are expected to further reduce animal use for acute systemic toxicity testing.

In February 2008, NICEATM and ICCVAM, in collaboration with ECVAM and JaCVAM, organized an International Workshop on Acute Chemical Safety Testing: Advancing *In Vitro* Approaches and Humane Endpoints for Systemic Toxicity Evaluation. The workshop was based on prior ICCVAM and expert panel recommendations that standardized procedures to collect information pertinent to an understanding of toxicity mechanisms should be included in future *in vivo* rat acute oral toxicity studies. The workshop participants recommended ways to collect data to identify key toxicity pathways for acute systemic toxicity so this mechanistic information can be used to target the development of predictive *in vitro* alternative test methods. The workshop also recommended that systematic collection of mechanistic data from required *in vivo* studies would help identify predictive biomarkers of systemic toxicity that could be used as earlier, more humane endpoints during in vivo tests to further reduce or avoid pain and distress (ICCVAM, 2009e).

More than one hundred people from seven countries attended the workshop, representing U.S. Federal agencies, academia, industry, international organizations, and the animal welfare community. Attendees participated in breakout group discussions to address questions for the following topics:

- Identifying key pathways leading to acute systemic toxicity
- Improving current acute systemic toxicity injury and toxicity assessments
- Identifying earlier humane endpoints for acute systemic toxicity testing
- Applying *in vivo* mode of action and mechanistic information to the development and validation of *in vitro* methods for assessing acute systemic toxicity
- Increasing industry involvement in test method development, validation, and use

NICEATM and ICCVAM are participating in the Validation Management Group of an ECVAM metabolism validation study. The goal of the study, the *ECVAM Validation Study in the Field of Toxicokinetics and Metabolism: Provision of a Standard for Human Hepatic Metabolism and Toxicity by Assessing as an Indicator Biotransformation Enzyme Induction using HepaRG Cells and Cryopreserved Human Hepatocytes*, is to develop a standard *in vitro* test system for human hepatic metabolism and metabolism-mediated toxicity and assess the potential for cytochrome P450 induction at clinically relevant doses. Issues being evaluated include 1) reliability of the test systems, 2) phenotypic stability of the cells, 3) within/between laboratory reproducibility, 4) transferability, and 5) predictivity for assessing *in vivo* human induction.

A high-throughput screening program at the NIH NCGC recently evaluated 13 other cytotoxicity test methods for predicting acute oral toxicity in humans (Xia et al., 2008). The generation of high-quality cytotoxicity data on a library of 1408 known compounds using HTS demonstrates the potential of this methodology to profile a much broader array of assays and compounds, which, in aggregate, may be valuable for prioritizing compounds for further toxicologic evaluation, identifying compounds with particular mechanisms of action, and potentially predicting *in vivo* biological response. NICEATM and ICCVAM will continue to identify candidate test methods for inclusion in the HTS that may help increase the accuracy of *in vitro* methods for predicting *in vivo* acute systemic toxicity.

8 Evaluations of endocrine disruptor screening methods

In 2002, an ICCVAM independent expert panel review concluded that there were no adequately validated *in vitro* estrogen receptor (ER) or androgen receptor based *in vitro* test methods available for screening of chemicals with potential endocrine disruptor activity. Based on the Panel's conclusions and recommendations, ICCVAM developed recommendations for minimum procedural standards and a list of 78 reference substances that should be used to standardize and validate such test methods. ICCVAM then invited the nomination of *in vitro* test methods that met the published recommendations for validation studies. Xenobiotic Detection Systems (XDS), Inc. subsequently nominated their LUMI-CELL® ER assay for the detection of ER agonists and antagonists to ICCVAM for validation. LUMI-CELL® is a stably transfected transcriptional activation assay (STTA) that uses a human cell line with human ERs. A joint NICEATM-ECVAM-JaCVAM international validation study of the LUMI-CELL® assay using the 78 recommended reference substances is currently being conducted in laboratories in Italy, the U.S., and Japan. An independent scientific peer review panel will evaluate the results of the study in 2011. CertiChem, Inc. also nominated its MCF-7 cell proliferation assay for validation studies. Like the LUMI-CELL® assay, it uses a human cell line that identifies human estrogen receptor agonist and antagonist activity. An interlaboratory validation study of the CertiChem assay is planned. Data from these validation studies will be used

to develop OECD test guidelines and test method performance standards for inclusion in the test guidelines.

9 Dermal safety assessment test method activities

In 2004, ICCVAM established performance standards for *in vitro* test methods for skin corrosion (ICCVAM, 2004). The performance standards were based on four *in vitro* test methods evaluated by ICCVAM for the identification of substances with the potential to cause skin corrosion. The standards can be used to evaluate the reliability and accuracy of other test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. NICEATM and ICCVAM have submitted revisions to OECD TG 430 (rat skin transcutaneous electrical resistance assay) and TG 431 (human skin model systems) that incorporate the ICCVAM-recommended performance standards (ICCVAM, 2004; OECD, 2009i, 2009j). ICCVAM scientists are also participating in OECD Expert Consultation meetings to evaluate several *in vitro* skin irritation assays for their inclusion in a draft TG.

NICEATM is currently conducting a study to determine how *in vitro* dermal irritation test methods (i.e., EpiDerm™ , EPISKIN™ , and SkinEthic™) will classify corrosive substances incorrectly identified as non-corrosives by *in vitro* corrosivity test methods. Current false negative rates for *in vitro* corrosivity assays range from 12% to 21% (ICCVAM, 1999a). This study will also confirm the extent to which a new procedure to identify substances that directly reduce MTT (3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyltetrazolium bromide) will reduce the false negative rate in corrosivity assays. A completely non-animal testing strategy for dermal irritation and corrosion must be able to accurately identify corrosive substances since these substances can cause permanent injuries and even death from severe chemical burns.

10 Genetic toxicity test method evaluation activities

The ICCVAM Genetic Toxicity Working Group (GTWG) recently provided comments on cytotoxicity evaluation procedures for a draft OECD Test Guideline 487 for the *in vitro* micronucleus test (OECD, 2007). The GTWG also provided comments on the proposed study plan, protocol, and reference substances for a JaCVAM-led international validation study of the *in vivo* rodent comet assay and nominated experts for an ESAC Peer Review Panel on a cell transformation assay. Similarly, the GTWG has provided technical comments on a planned JaCVAM study of a cell transformation assay.

11 Pyrogen test method evaluation activities

ICCVAM forwarded recommendations on five *in vitro* test methods proposed for assessing potential pyrogenicity of pharmaceuticals and other products to Federal agencies in November 2008 (ICCVAM, 2008a). All applicable Federal agencies, including the U.S. FDA, accepted or endorsed the ICCVAM recommendations (ICCVAM, 2009m). These test methods were also recently adopted by the European Pharmacopoeia Commission during its 133rd session in March 2009 for inclusion in the European Pharmacopeia in 2010 (EDQM, 2009a). These methods should now be considered prior to conducting *in vivo* pyrogenicity testing and should be used where determined appropriate for specific testing situations. The availability of these test methods may reduce the number of animals required for pyrogenicity testing.

12 Biologics test method evaluation activities

Biologics potency and safety testing is one of the four highest ICCVAM-NICEATM priorities because it is required by multiple agencies and can require large numbers of animals that may experience significant pain and distress during testing (ICCVAM, 2008b). NICEATM and ICCVAM, in conjunction with ECVAM and JaCVAM, are organizing an international workshop on alternative methods to reduce, refine, and replace the use of animals in vaccine potency and safety testing. The workshop is scheduled for September 14-16, 2010, at the William H. Natcher Conference Center on the main campus of the NIH in Bethesda, MD. More information about the workshop will be posted at http://iccvam.niehs.nih.gov/methods/biologics/biologics.htm as it is available. NICEATM and ICCVAM are also planning an evaluation of U.S. Department of Agriculture/Michigan State University study of *in vitro* potency tests for Leptospirosis vaccines.

13 Conclusions

Since its establishment, ICCVAM has led or contributed to the evaluation of 27 alternative test methods that have now been accepted or endorsed by national and international authorities. Numerous other test methods are expected to be adopted for regulatory use over the next few years as NICEATM and ICCVAM implement their Five-Year Plan. In addition, many new test methods are expected to emerge from NICEATM-ICCVAM collaborations with member agencies to conduct research, development, translation, and validation efforts relevant to alternative methods. These efforts will integrate scientific advances and new technologies into new test methods and strategies. Implementation of the International Cooperation on Alternative Test Methods is also expected to result in collaborations that will expedite the adoption of alternative methods that will support improved safety assessments while reducing, refining, and replacing animal use.

References

Blakey, D. H. (2009). Health Canada's role in ICATM. *ALTEX 26, Spec. Issue*, 173. (Accessed September 17, 2009).

Dean, J. H., Twerdok, L. E. et al. (2001). ICCVAM evaluation of

the murine local lymph node assay: II. Conclusions and recommendations of an independent scientific peer review panel. *Regulatory Toxicology and Pharmacology 34(3),* 258-273. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035697810&partnerID=40&rel=R5.6.0 (Accessed September 17, 2009).

EDQM (2009a). *Report of the 133rd Session of the European Pharmacopoeia Commission* 17-18th March 2009. Available at: http://www.edqm.eu/en/News-and-General-Information-43.html (Accessed September 17, 2009).

EDQM (2009b). *European Pharmacopeia*. Strasbourg: Council of Europe.

EPA, U. S. (2003). Health Effects Test Guideline OPPTS 870.2600 - Skin Sensitization. EPA 712-C-03-197. Washington, DC, U.S. Environmental Protection Agency. Available at: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Revised/870r-2600.pdf (Accessed: September 17, 2009).

EPA, U. S. (2002). Health Effects Test Guideline OPPTS 870.1100 - Acute Oral Toxicity. EPA 712-C-03-197. Washington, DC, U.S. Environmental Protection Agency. Available at: http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Revised/870r-1100.pdf (Accessed: September 17, 2009).

FDA (2009). Letter from Norris E. Alderson, Ph.D., FDA Associate Commissioner for Science, to RADM William S. Stokes, Director NICEATM. Available at: http://iccvam.niehs.nih.gov/methods/pyrogen/transmitNov08/FDA-Response.pdf (Accessed: September 17, 2009).

Gerberick, G. F., Vassallo, J. D. et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicological Sciences 97(2),* 417-427. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17400584 (Accessed September 17, 2009).

Haneke, K. E., Tice, R. R. et al. (2001). ICCVAM evaluation of the murine local lymph node assay: III. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. *Regulatory Toxicology and Pharmacology 34(3),* 274-286. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035702267&partnerID=40&rel=R5.6.0 (Accessed September 17, 2009).

Hendriksen, C. F. (2009). Replacement, reduction and refinement alternatives to animal use in vaccine potency measurement. *Expert Rev Vaccines 8(3),* 313-22. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19249973

Hendriksen, C. F., Woltjes, J. et al. (1994). Interlaboratory validation of in vitro serological assay systems to assess the potency of tetanus toxoid in vaccines for veterinary use. *Biologicals 22(3),* 257-268. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7811460

Hill, R. E. (2004). Notice clarifying policy on the use of humane endpoints for animal challenge potency tests on biologics. Center for Veterenary Biologics Notice No. 04-09. Ames, IA, Center for Veterenary Biologics, US USDA. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/USDA/CVB04-09.pdf (Accessed: September 17, 2009).

ICATM (2009). Memorandum of Cooperation Regarding International Cooperation on Alternative Test Methods (ICATM). Available at: http://iccvam.niehs.nih.gov/docs/about_docs/ICATM-MOC.pdf (Accessed September 17, 2009).

ICCVAM (2009a). ICCVAM Test Method Evaluation Report - The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. NIH Publication Number 09-6439. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-LD/TMER.htm (Accessed: September 17, 2009).

ICCVAM (2009b). Independent Scientific Peer Review Panel Report: Updated Validation Status of New Versions and Applications of the Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemical and Products. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf (Accessed: September 17, 2009).

ICCVAM (2009c). Independent Scientific Peer Review Panel Report: Evaluation of the Validation Status of Alternative Ocular Safety Testing Methods and Approaches. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/ocutox_docs/OcularPRPRept2009.pdf (Accessed: September 17, 2009).

ICCVAM (2009d). Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf (Accessed: September 17, 2009).

ICCVAM (2009e). Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/acutetox/toxwksp-rpt.htm (Accessed: September 17, 2009).

ICCVAM (2009f). Alternative Test Method Project Milestones. Available at: http://iccvam.niehs.nih.gov/methods/milestones.htm (Accessed September 17, 2009).

ICCVAM (2009g). Draft Background Review Document: Use of Topical Anesthetics, Systemic Analgesics, and Earlier Humane Endpoints to Minimize Pain and Distress in Ocular Toxicity Testing. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/pretreat/BRD.pdf (Accessed September 17, 2009).

ICCVAM (2009h). Draft Background Review Document: Current Status of In Vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: The Isolated Chicken Eye (ICE) Test Method. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/mildmod/ICE-BRD.pdf (Accessed September 17,

2009).
ICCVAM (2009i). Draft Background Review Document: Current Status of In Vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: The Bovine Corneal Opacity and Permeability (BCOP) Test Method. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/mildmod/BCOP-BRD.pdf (Accessed September 17, 2009).
ICCVAM (2009j). Draft Background Review Document: Current Status of In Vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: The Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) Test Method. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/mildmod/HET-CAM-BRD.pdf (Accessed September 17, 2009).
ICCVAM (2009k). Draft Summary Review Document: The In Vivo Low Volume Eye Test (LVET). Available at: http://iccvam.niehs.nih.gov/methods/ocutox/antimicro/LVET-BRD.pdf (Accessed September 17, 2009).
ICCVAM (2009l). Draft Summary Review Document: Strategy for U.S. Environmental Protection Agency Ocular Hazard Classification and Labeling of Antimicrobial Cleaning Products Using In Vitro Alternative Test Methods. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/antimicro/BRD.pdf (Accessed September 17, 2009).
ICCVAM (2009m, September 3, 2009). In Vitro Pyrogen Test Methods. Available at: http://iccvam.niehs.nih.gov/methods/pyrogen/pyrogen.htm (Accessed September 17, 2009).
ICCVAM (2008a). ICCVAM Test Method Evaluation Report: Validation Status of Five In Vitro Test Methods Proposed for Assessing Potential Pyrogenicity of Pharmaceuticals and Other Products. NIH Publication Number 08-6392. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/pyrogen/pyr_tmer.htm (Accessed: September 17, 2009).
ICCVAM (2008b). The NICEATM-ICCVAM Five-Year Plan (2008-2012). NIH Publication Number 08-6410. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/5yrPlan/NICEATM5YR-Final.pdf (Accessed: September 17, 2009).
ICCVAM (2006a). ICCVAM Test Method Evaluation Report (TMER): In Vitro Cytotoxicity Test Methods for Estimating Starting Doses For Acute Oral Systemic Toxicity Testing. NIH Publication Number 07-4519. Research Triangle Park, NC, National Insitute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_tmer.htm (Accessed: September 17, 2009).
ICCVAM (2006b). ICCVAM Test Method Evaluation Report - In Vitro Ocular Toxicity Test Methods for Identifying Severe Irritants and Corrosives. NIH Publication Number 07-4517. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/ocutox_docs/OTeval/OTevalrpt.pdf (Accessed: September 17, 2009).
ICCVAM (2006c). Background Review Document: Current Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability (BCOP) Test Method. NIH Publication Number 06-4512: Volume 1 of 2. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/ocutox_docs/ocubrd/bcop/bcopbrd.pdf (Accessed: September 17, 2009).
ICCVAM (2006d). Background Review Document: Current Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability (BCOP) Test Method. NIH Publication Number 06-4512: Volume 2 of 2. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/ocutox_docs/ocubrd/bcop/bcopapp.pdf (Accessed: September 17, 2009).
ICCVAM (2006e). Background Review Document - Current Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye (ICE) Test Method. NIH Publication Number 06-4513. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/ocutox_docs/ocubrd/ice/icebrd.pdf (Accessed: September 17, 2009).
ICCVAM (2004). Recommended Performance Standards for In Vitro Test Methods for Skin Corrosion: Prepared by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 04-4510. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf (Accessed: September 17, 2009).
ICCVAM (2002). Evaluation of EPISKIN™ , EpiDerm™ , and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: In Vitro Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication Number 02-4502. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/cwgfinal02/cwgfinal.pdf (Accessed: September 17, 2009).
ICCVAM (2001a). Report of the International Workshop on In Vivo Methods for Assessing Acute Systemic Toxicity: Results of an International Workshop Organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 01-4499. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/acutetox_docs/finalrpt/finalall0801.pdf (Accessed: September 17, 2009).
ICCVAM (2001b). ICCVAM Test Method Recommendations on Acute Systemic Toxicity (Appendix I of Workshop Report). NIH Publication Number 01-4499. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/acutetox_docs/finalrpt/finappi2.pdf (Accessed: September 17, 2009).
ICCVAM (2001c). The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals: Results of an Independent Peer Review Evalua-

tion Organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). NIH Publication Number 02-4501. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/acutetox_docs/udpProc/udpfin01/vol_1.pdf (Accessed: September 17, 2009).

ICCVAM (1999a). Corrositex®: An In Vitro Test Method for Assessing Dermal Corrosivity Potential of Chemcials - An Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4495. Research Trinagle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf (Accessed: September 17, 2009).

ICCVAM (1999b). The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication Number 99-4494. Research Triangle Park, NC, National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf (Accessed: September 17, 2009).

Idehara, K., Yamagishi, G. et al. (2008). Characterization and evaluation of a modified local lymph node assay using ATP content as a non-radio isotopic endpoint. *Journal of Pharmacological and Toxicological Methods 58(1),* 1-10. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-42649119427&partnerID=40&rel=R8.0.0 (Accessed September 17, 2009).

IIVS (2008). Background Review Document of Existing Methods for Eye Irritaiton Testing: Silicon Microphysiometer and Cytosensor Microphysiometer. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/CM/ECVAM-CMBRD-Aug08cleaned.pdf (Accessed September 17, 2009).

ISO (2002). Biological Evaluation of Medical Devices - Part 10: tests for irritation and delayed-type hypersensitivity reaction.

Johannes, S., Hartinger, J. et al. (2003). Humane endpoints in the efficacy testing of swine erysipelas vaccines. *ALTEX 20(1),* 11-15. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12579350

Kojima, H. (2009). JaCVAM's role in the 3Rs and ICATM. *ALTEX 26, Spec. Issue*, 172. (Accessed September 17, 2009).

Kreysa, J. (2009). ICATM's importance for ECVAM and the 3Rs in the EU. *ALTEX 26, Spec. Issue,* 171-172. (Accessed September 17, 2009).

MB Research Laboratories. Local Lymph Node Assay Using Flow Cytometric Endpoints: An Alternative to the Radioactive LLNA. TOXNOTE. Available at: http://www.mbresearch.com/TOXNOTE/TOXNOTE-LLNA.pdf (Accessed September 17, 2009).

NIEHS (2009). Press Release: Countries Unite to Reduce Animal Use in Product Toxicity Testing Worldwide - U.S.,Canada, Japan and European Union Sign International Agreement. Available at: http://www.nih.gov/news/health/apr2009/niehs-27.htm (Accessed September 17, 2009).

NIEHS (1997). Notice of Establishment. *Federal Register 62*, 48290. Available at: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/FR/FR_97_24273.pdf (Accessed September 17, 2009).

OECD (2009a). Draft Proposal for a New Test Guideline: Skin Sensitisation: Local Lymph Node Assay: DA Version. Proposed June 26, 2009. Available at: http://www.oecd.org/dataoecd/63/5/43302184.pdf (Accessed September 17, 2009).

OECD (2009b). Draft Proposal for a New Test Guideline: Skin Sensitisation: Local Lymph Node Assay: BrdU-ELISA. Proposed June 26, 2009. Available at: http://www.oecd.org/dataoecd/63/7/43302529.pdf (Accessed September 17, 2009).

OECD (2009c). Draft Proposal for an Update to Test Guideline 429: Skin Sensitisation: Local Lymph Node Assay. Proposed June 26, 2009. Available at: http://www.oecd.org/dataoecd/39/10/43345215.pdf (Accessed September 17, 2009).

OECD (2009d). OECD Guidelines for the Testing of Chemicals. Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants. Adopted September 7, 2009. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s54/p1 (Accessed: September 17, 2009).

OECD (2009e). OECD Guidelines for the Testing of Chemicals. Test No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants. Adopted September 7, 2009. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s55/p1 (Accessed: September 17, 2009).

OECD (2009f). Draft Proposal for a New Guidance Document: Using Cytotoxicity Tests to Estimate Starting Doses for Acute Oral Systemic Toxicity Tests. Proposed June 26, 209. Available at: http://www.oecd.org/dataoecd/17/0/43325517.pdf (Accessed September 17, 2009).

OECD (2009g). Draft Proposal for a New Guidance Document on Histopathology for Inhalation Toxicity Studies, Supporting TG 412 and 413. Proposed June 15, 2009. Available at: http://www.oecd.org/dataoecd/10/20/43062801.pdf (Accessed September 17, 2009).

OECD (2009h). Draft Proposal for a New Guidence Document on a Threshold for Fish Acute Toxicity. Available at: http://www.oecd.org/dataoecd/49/3/43226061.pdf (Accessed September 17, 2009).

OECD (2009i). Draft Proposal for an Update to Test Guideline 430: In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). Available at: http://www.oecd.org/dataoecd/14/2/43295353.pdf (Accessed September 17,

2009).
OECD (2009j). Draft Proposal for an Update to Test Guideline 431: In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method. Available at: http://www.oecd.org/dataoecd/63/6/43302385.pdf (Accessed September 17, 2009).
OECD (2008a). OECD Guidelines for the Testing of Chemicals. Test No. 425: Acute Oral Toxicity: Up-and-Down Procedure. Adopted October 3, 2008. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s25/p1 (Accessed: September 17, 2009).
OECD (2008b). Draft Proposal for an Update to Test Guideline 403: Acute Inhalation Toxicity. Proposed November 28, 2008. Available at: http://www.oecd.org/dataoecd/54/55/41761261.pdf (Accessed September 17, 2009).
OECD (2008c). Draft Guidance Document on Acute Inhalation Toxicity Testing No. 39. Proposed November 28, 2008. Available at: http://www.oecd.org/dataoecd/54/24/41761788.pdf (Accessed September 17, 2009).
OECD (2008d). Draft Proposal for an Update to Test Guideline 436: Acute Inhalation Toxicity - Acute Toxic Class (ATC) Method. Proposed November 28, 2008. Available at: http://www.oecd.org/dataoecd/54/11/41761436.pdf (Accessed September 17, 2009).
OECD (2007). Draft Proposal for an Update to Test Guideline 487: In Vitro Mammalian Cell Micronucleus Test (MNvit). Proposed December 13, 2007. Available at: http://www.oecd.org/dataoecd/38/58/39780112.doc (Accessed September 17, 2009).
OECD (2006). OECD Guidelines for the Testing of Chemicals. Test No. 435: In Vitro Membrane Barrier Test Method for Skin Corrosion. Adopted July 19, 2006. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s33/p1 (Accessed: September 17, 2009).
OECD (2004a). OECD Guidelines for the Testing of Chemicals. Test No. 428: Skin Absorption: In Vitro Method. Adopted April 13, 2004. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s28/p1 (Accessed: September 17, 2009).
OECD (2004b). OECD Guidelines for the Testing of Chemicals. Test No. 430: In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER). Adopted April 13, 2004. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s30/p1 (Accessed: September 17, 2009).
OECD (2004c). OECD Guidelines for the Testing of Chemicals. Test No. 431: In Vitro Skin Corrosion: Human Skin Model Test. Adopted April 13, 2004. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s31/p1 (Accessed: September 17, 2009).
OECD (2004d). OECD Guidelines for the Testing of Chemicals. Test No. 432: In Vitro 3T3 NRU Phototoxicity Test. Adopted April 13, 2004. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s32/p1 (Accessed: September 17, 2009).
OECD (2002). OECD Guidelines for the Testing of Chemicals. Test No. 429: Skin Sensitisation: Local Lymph Node Assay. Adopted April 24, 2002. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s29/p1 (Accessed: September 17, 2009).
OECD (2001a). OECD Guidelines for the Testing of Chemicals. Test No. 420: Acute Oral Toxicity - Fixed Dose Procedure. Adopted December 17, 2001. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s20/p1 (Accessed: September 17, 2009).
OECD (2001b). OECD Guidelines for the Testing of Chemicals. Test No. 423: Acute Oral toxicity - Acute Toxic Class Method. Adopted December 17, 2001. Available at: http://puck.sourceoecd.org/vl=6381888/cl=19/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s23/p1 (Accessed: September 17, 2009).
Ovigne, J. M., Martinozzi-Teissier, S. et al. (2008). The MUSST for in vitro skin sensitization prediction: Applicability domains and complementary protocols to adapt to the physico-chemical diversity of chemicals. *Toxicology Letters 180*, S216-S216. Available at: <Go to ISI>://000259252100703 (Accessed September 17, 2009).
Rosskopf-Streicher, U., Johannes, S. et al. (2001). Quality control of inactivated erysipelas vaccines: results of an international collaborative study to establish a new regulatory test. *Vaccine 19(11-12),* 1477-1483. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11163671
Sailstad, D. M., Hattan, D. et al. (2001). ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process. *Regulatory Toxicology and Pharmacology 34(3),* 249-257. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035697809-&partnerID=40&rel=R5.6.0 (Accessed September 17, 2009).
Sakaguchi, H., Ashikaga, T. et al. (2006). Development of an in vitro skin sensitization test using human cell lines; human Cell Line Activation Test (h-CLAT) II. An inter-laboratory study of the h-CLAT. *Toxicology in Vitro 20(5),* 774-784. (Accessed September 17, 2009).
Spielmann, H., Balls, M. et al. (1998). The international EU/COLIPA in vitro phototoxicity validation study: results of phase II (blind trial), part 1: the 3T3 NRU phototoxicity test. *Toxicology in vitro 12*, 305-327.
Stokes, W. and Wind, M. (2009). Current and future contributions of NICEATM and ICCVAM to the International Cooperation on Alternative Test Methods. *ALTEX 26, Spec. Issue*, 172. (Accessed September 17, 2009).
Stokes, W. S., Casati, S. et al. (2008). Unit 20.4. Neutral red uptake cytotoxicity tests for estimating starting doses for acute oral toxicity tests. *Current Protocols in Toxicology 36*, 20.4.1-20.4.20.
Stokes, W. S. and Schechtman, L. M. (2007). *Validation and regulatory acceptance of new, revised, and alternative toxicological methods. Principles and Methods of Toxicology*. A. W. Hayes. Philadelphia, Pennsylvania: Taylor and Francis: 1103-1128.
Takeyoshi, M., Yamasaki, K. et al. (2001). Development of

non-radio isotopic endpoint of murine local lymph node assay based on 5-bromo-2?-deoxyuridine (BrdU) incorporation. *Toxicology Letters 119(3),* 203-208. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035826075&partnerID=40&rel=R6.0.0 (Accessed September 17, 2009).

USC (2000). ICCVAM Authorization Act of 2000. Public Law 106-545. Available at: http://frwebgate.access.gpo.gov/cgi-bin/useftp.cgi?IPaddress=162.140.64.183&filename=publ545.pdf&directory=/diska/wais/data/106_cong_public_laws (Accessed: September 17, 2009).

Wind, M., Blakey, D. et al. (2009). What is the International Cooperation on Alternative Test Methods (ICATM) and what is its role? *ALTEX 26, Spec. Issue*, 171. (Accessed September 17, 2009).

Xia, M., Huang, R. et al. (2008). Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ. Health Perspect 116(3),* 284-291. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18335092 (Accessed September 17, 2009).

Acknowledgements

The authors acknowledge the expert assistance of Steven Morefield, Catherine Sprankle, and David Allen from ILS, Inc., the NICEATM Support Contractor, and Debbie McCarley, NICEATM/NIEHS, in the preparation of this manuscript. The content of this manuscript was supported by the Intramural Research Program of the National Institutes of Health.

Correspondence to

Dr. William S. Stokes
National Institute of Environmental Health Sciences
P.O. Box 12233, Mail Stop K2-16
Research Triangle Park, North Carolina
USA 27709
e-mail: stokes@niehs.nih.gov

Theme 3: Progress in life science domains

Breakout Sessions

BS21: Skin and eye toxicity I

Mechanistic Assays for the *In Vitro* Prediction of Skin Sensitization: Integrating Biological and Chemical Measurements in a Holistic Approach

Andreas Natsch

Givaudan Schweiz AG, Duebendorf, Switzerland

Summary

The skin sensitization reaction to small molecules is a complex immunological process, yet it is relatively well understood at the molecular level: In order for a molecule to be a sensitizer it must (i) gain access to the viable epidermis, (ii) react with skin proteins, (iii) the modified proteins need to be presented by MHC molecules on dendritic cells, (iv) in parallel the molecule must activate a "danger signal" in keratinocytes and/or dendritic cells to initiate migration of dendritic cells to lymph nodes, and finally (v) the dendritic cells must activate proliferation of hapten-specific T-cells in the lymph node. Given the complexity of this process, there is a paradigm widely held within the scientific community, that no single in vitro assay can reliably predict the sensitization potential of chemicals. Instead, a battery of tests will be needed. In this paper, progress made on the development of (a) bioavailability models, (b) dendritic cell activation models, (c) protein reactivity models, (d) other cellular models reflecting the "danger signal" induction and (e) T-cell activation are briefly reviewed. Results from our current simplified test battery are discussed: By combining (i) data from peptide reactivity as a purely chemical measurement and (ii) cell based data on the induction of the Nrf2-KeapI-ARE signalling pathway, a good prediction can already be made, and this test battery is already used in our routine research process.

Keywords: skin sensitizers, peptide reactivity, battery of assays, integrated testing, antioxidant response element, dendritic cell activation

1 Introduction

The cosmetic legislation in Europe has put a ban on animal testing for the prediction of the skin sensitization risk of cosmetic ingredients after 2013. Yet, the risk of contact sensitization is a critical issue in the development of novel ingredients for cosmetic products. To guarantee safe products after 2013, predictive *in vitro* tests are thus urgently needed. The risk for contact sensitization is currently estimated with the local lymph node assay in mice (LLNA), in which the cellular proliferation in the draining lymph nodes is measured after repeated topical application of the test compound onto the ears. Results are expressed as EC3 values, indicating the concentration which induces a threefold enhanced cellular proliferation (Basketter et al., 2002).

Animal tests by their very nature give a holistic outcome at the end of a reaction cascade, thus a compound is topically applied in the LLNA and finally the cellular proliferation in the lymph node induced by a chain of chemical and biological events is

measured. Animal tests give a holistic outcome, but they may be viewed as a "black box" at the same time with no information on the single steps in the process being recorded. Since the multi-step immunological process leading to skin sensitization cannot be brought into the laboratory in the form of a single *in vitro* assay, all *in vitro* testing must dissect the pathways into single elements, and the critical, mechanistically most relevant and/or experimentally most accessible elements need to be modelled with individual *in vitro* assays. A major challenge then is to integrate the information generated in these single assays.

2 The key steps in the skin sensitization reaction

In Figure 1, the different cell types and the key steps involved in the skin sensitisation reaction are summarized. The key cell types involved are (i) the specific T-cells (the effector cells), (ii) the dendritic, or more specifically the Langerhans cells (presenting the antigens to the T-cells) and (iii) the keratinocytes (the primary cells in contact with the sensitizers and thought to contribute important "danger signals" in the sensitization phase).

In a first step, the molecule applied topically must reach the viable epidermis (step 1), and skin penetration or more specifically "bioavailability in the skin" therefore is thought to be a prerequisite for the skin sensitization reaction. Certain molecules are considered pro-haptens, indicating that they themselves are non-sensitizers and need to be metabolically converted by skin enzymes to become the causative agents of a sensitization reaction (step 2; in brackets as only needed for certain molecules). A key step (step 3) in the process is the formation of a covalent adduct between the skin sensitizer and endogenous proteins and/or peptides in the skin. The modified peptides then must be presented by the Langerhans cells (i.e. the dendritic cells in the skin) on their MHC molecules (step 4). It is thought, that the mere binding of modified peptides on the dendritic cells is not sufficient to initiate the immune reaction, but that additional signals, often referred to as "danger signals", are needed to stimulate the dendritic cells (step 5). These danger signals appear to be part of the innate immune reaction and are formed in absence of specificity conveyed by specific T-cells. The complete chain of events in the formation of these danger signals is unknown, but these innate reactions may be the most valuable targets to look at when developing highly standardized *in vitro* tests as discussed further below.

Stimulated by the triggers/danger signals coming from the innate reactions, the dendritic cells mature and emigrate from the skin into the local lymph node (step 6), where they finally present the modified peptides to the T-cells and stimulate the proliferation of specific T-cell clones (step 7). This complete process summarized above and depicted in Figure 1 is referred to as the "sensitization phase" of the skin sensitization reaction. It is only this phase which is modelled by the current gold-standard, the LLNA assay, and virtually all *in vitro* tests try to model elements of this phase. Nevertheless it should not be left unmentioned that only the "elicitation phase", when the specific T-cell clones trigger skin inflammation upon secondary contact with the causative agents, is the real disease state (known as contact allergy) causing problems for the sensitized individuals. This elicitation phase had been part of the older animal tests (e.g. Guinea pig maximisation test and Buehler test), and it is

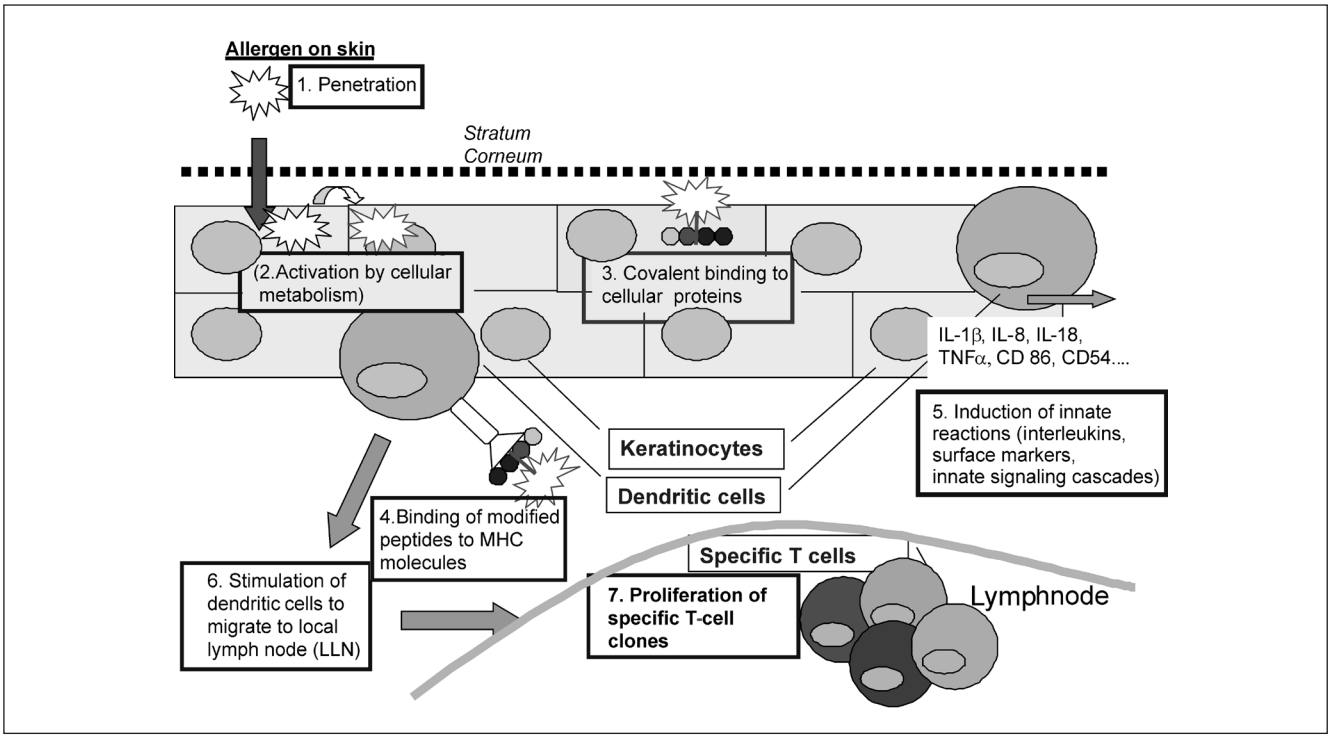


Fig. 1: A summary of seven key steps thought to be involved in the sensitization phase of skin sensitization.

also the endpoint looked at when performing human tests, be it the diagnostic patch tests conducted by dermatologists on sensitized individuals or the predictive human repeated insult patch test (HRIPT) occasionally used to assess the sensitization potential of individual chemicals. In Table 1, the different steps of the process are summarized, along with some key ongoing assay developments to model the different steps. These assays are described in more detail below.

2.1 Modelling skin penetration

In principle, when looking at the physicochemical parameters of skin sensitizers, there seems to be no clear limitation for either hydrophobic or hydrophilic chemicals to gain access to the viable epidermis and cause a sensitization reaction, and the importance of the skin penetration was therefore questioned by some authors (Roberts and Aptula, 2008). On the other hand, there are case reports of similarly reactive chemicals with very different bioavailability, for which a clear effect of skin penetration on sensitization has been reported (Aeby et al., 2004).

There is much literature on models to predict skin penetration, and there are several *in vitro* test systems available using pig skin, reconstituted human epidermis or primary human skin tissue samples to assess skin penetration. Predictive computer models have also been developed from the data generated with these tests. This literature is not reviewed here, and reviews can be found elsewhere (Basketter et al., 2007). There is a critical limitation in the majority of the classical skin penetration data and models: they assess partition from an aqueous solution into the more hydrophobic epidermis, whereas in the case of skin sensitisation testing, the chemicals are mostly applied in hydrophobic olive oil, a donor solution which yields a very different partition as compared to an aqueous vehicle. One detailed approach to model skin penetration from the actual vehicles used in the LLNA is thus being developed by G. Kasting (Basketter et al., 2007; Kasting et al., 2008) and may become very useful as part of a skin sensitization prediction model.

2.2 Modelling pro-hapten activation

Very little is known about the actual metabolic capacity of the skin towards xenobiotics. Several studies have shown that mRNA for a variety of P450 enzymes is expressed in the skin, and one research group (Bergström et al., 2007) has developed a skin-like P450 cocktail based on recombinant human P450 enzymes. They have shown that a specific skin sensitizing putative pro-hapten can be activated by this cocktail. This enzyme cocktail was reconstituted based on data from RT-PCR studies on mRNA and not based on enzyme activities really detected in the skin samples nor on proteomic data, and the evidence for these enzymes being active in the skin therefore is indirect. The only P450 enzyme activities really detected in keratinocytes are AhR-ligand-inducible activities such as 7-ethoxycoumarin-O-deethylase and ethoxyresorufin-de-ethylase activity (Cotovio et al., 1996). Liver microsomes may also be used to activate a variety of putative pro-haptens (Bergström et al., 2006, 2007), but this approach may be overpredictive for the real situation, as the metabolic capacity of the skin is certainly much lower as compared to the liver. Thus also non-sensitizing terpenes were activated to epoxides and conjugated to glutathione in incuba-

Tab. 1: The different steps in the sensitization phase of skin sensitization and selected approaches in development to model these steps *in vitro*, *in chemico* and *in silico*

Mechanistic step	Selected assay / model system
1) Penetration	Classical skin penetration assays (Basketter et al., 2007) Bioavailability model (Kasting et al., 2008)
2) Pro-hapten activation	Activation by liver microsomes (Bergström et al., 2006) Activation by skin-like P450 cocktail (Bergström et al., 2007)
3) Modelling the reaction with skin proteins	DPRA peptide depletion assay (Gerberick et al., 2004; Gerberick et al., 2007) LC-MS based assays (Aleksic et al., 2009; Natsch and Gfeller, 2008) Kinetic profiling assay (Roberts and Natsch, 2009)
4) Modelling the binding of modified peptides to MHC molecules	No separate assay, but part of T-cell assay
5) Modelling the activation of innate reactions / danger signal formation	h-CLAT assay (CD54 and CD86 expression on THP-1 cells) (Sakaguchi et al., 2006) MUSST assay (CD86 expression on U-937) IL-18 formation in keratinocytes (Corsini et al., 2009; Van Och et al., 2005) Induction of antioxidant dependent genes (Ade et al., 2009; Natsch and Emter, 2008)
6) Modelling dendritic cell emigration	Dendritic cell migration assay developed within the Sens-it-iv project http://www.sens-it-iv.eu/files/newsletter/Sens-it-iv_Newsletter_22.html
7) Modelling T-cell activation	T-cell proliferation assay developed within the Sens-it-iv project http://www.sens-it-iv.eu/files/newsletter/Sens-it-iv_Newsletter_26.pdf

tions with liver microsomes (Bergström et al., 2006). Finding a physiologically relevant model for enzymatic activation of pro-haptens thus remains a major challenge.

2.3 Modelling the reaction with skin proteins

The reaction of the skin sensitizer with skin proteins is a key step in the sensitization reaction, and it might turn out to be the rate-limiting step (Roberts and Aptula, 2008). In principle, one can therefore try to predict the skin sensitization reaction based on the reactivity of a test compound towards peptides and proteins in a purely *chemical assay* (reviewed in (Gerberick et al., 2008)). Currently, the nature of the major modified proteins in the skin is unknown. We do not know whether these are few specific proteins, extra- or intracellular proteins, or whether the sensitizers even directly modify endogenous MHC-bound peptide ligands. Thus all reactivity assays are based on reactions between a "protein model" and the test chemical. The most often used protein models are (i) native proteins such as BSA (Aleksic et al., 2007, 2008), (ii) model peptides, (iii) simple amino acid derivatives such as acetyl-cysteine and acetyl-lysine or even (iv) small nucleophiles such as aniline or butylamine. Most studies published in recent years focused on the reactivity of test chemicals with a model peptide. These usually contain one or several nucleophilic residues (mainly cysteine and lysine) and are 3-10 amino acid residues long. The peptides are incubated with the test chemical in a solvent/co-solvent system and analyzed either after a fixed time or in a kinetic time-course experiment. Analysis can be performed either on the disappearance of the peptide (peptide depletion) or it can focus on the formation of covalent adducts by using LC-MS analysis or NMR. Gerberick et al. (Gerberick et al., 2004) developed a peptide depletion assay, in which the peptide is incubated with an excess of a test chemical, and peptide depletion by the test chemical is recorded as endpoint. This assay is called DPRA (direct peptide reactivity assay). It underwent ring-studies and has been submitted to ECVAM for prevalidation. Gerberick et al. (2007) reported an overall accuracy of 89% on a set of 81 chemicals. We have further developed this approach by integrating LC-MS detection in order to simultaneously record peptide depletion and peptide-adduct formation (Natsch and Gfeller, 2008; Natsch et al., 2007). This approach allowed us to discriminate whether peptide depletion is (i) due to simple peptide oxidation or (ii) due to real adduct formation. A similar approach was also taken by Aleksic et al. (2009). In addition we had shown that reactivity can further be quantified by measuring kinetic rate constants in addition to the 24 h endpoint value proposed in the DPRA assay (Roberts and Natsch, 2009).

In general these assays are technically straightforward, accuracy for predicting standard sets of skin sensitizers is good, and this approach addresses a very key step in the sensitization reaction. The most critical issue now is the fact that these assays do not incorporate a metabolic component which might be crucial to detect pro-haptens, and they fail to detect reactivity of some chemicals with very low reactivity. The latter however may be overcome with very sensitive analysis of adduct formation by LC-MS.

2.4 Modelling binding to MHC molecules

Processing of modified proteins by the dendritic cells and binding of modified peptides to the MHC molecules is a process which in general is not considered rate-limiting or chemical-dependent and thus not important in the profiling of different chemicals. Specific assays mimicking/modelling this process are therefore currently not developed. However, the more holistic and more complex T-cell assays described below do include this step, as allergen-modified immature dendritic cells are used in the assays to stimulate T-cell proliferation (see below).

2.5 Modelling the activation of innate reactions / danger signal formation

The challenge in developing a predictive biological assay for skin sensitizers amenable to full validation is significant. The probably most interesting step to be modelled by such an assay is the innate reaction / danger signal formation, since these reactions occur in the dendritic cells and/or the keratinocytes, and stable cell lines related to these cell types are available. As a key requirement, a relevant marker/signalling pathway needs to be identified which is:

– induced in an immortal cell line
– induced by sensitizers but not irritants

What are the best strategies to find the optimal markers? Three approaches have been taken:

– In the "classical approach" researchers looked *at typical immunological markers* asking: "What is the expected reactions of cells to sensitizer challenge?"
– In the gene chip/omics approach several research groups have challenged cells with sensitizers and looked for *all possible reactions* (all genes/proteins).
– In a more mechanistic or functional approach we asked for the likely signalling pathways which could be selective for sensitizers. Or in other words: How can cells *sense the sensitizing potential*?

The classical approach mostly focused on known surface markers expressed by activated dendritic cells, namely CD86, CD54 and HLA-proteins. This approach has been developed into two standardized assays: The h-CLAT test developed by the cosmetic companies Shishiedo and Kao measures induction of CD86 and CD54 in THP-1 cells (Sakaguchi et al., 2006), whereas the MUSST assay developed by L'Oréal quantifies induction of CD86 in U-937 cells. Both the h-CLAT and MUSST assay have gone through ring trials to assure lab-to-lab transferability, and both assays were submitted for pre-validation to ECVAM. Other markers looked at in the "classical approach" are cytokines known to be expressed *in vivo* in allergen-treated skin. Some more recent studies focused specifically on the interleukin IL-18 (Corsini et al., 2009; Van Och et al., 2005), which is interesting as IL-18 has been established as a key and specific factor induced by sensitizers but not irritants in the skin (Antonopoulos et al., 2008; Cumberbatch et al., 2001).

With ***the omics approach*** a number of new sensitizer-specific markers were identified in primary dendritic cells derived

from blood samples. Thus studies by Ryan et al. (Ryan et al., 2004) found a series of markers which were further validated in detail by RT-PCR (Gildea et al., 2006). Similarly, the research group at the VITO institute found a range of markers in primary dendritic cells and developed these findings into an assay called "VitoSens™" (Hooyberghs et al., 2008). Unfortunately, this assay needs to be performed with primary cells, and it proved difficult to transfer the approach to a stable cell line (Lambrechts et al., 2009): The response in the stable cell line was found to be significantly different from the response in primary cells. This difficulty in transferring the results from primary cells to a standardized protocol with a cell line certainly remains a hurdle for validation. The strength of the most recent study (Python et al., 2009) was therefore to look for reactions in primary cells and in a stable cell line (MUTZ-3) in parallel. Interestingly, also this group found an extremely low overlap between the reactions in primary cells and in stable cell lines, but still four stable overlapping markers were identified (see below). Nevertheless, no new assay based on a stable cell line has resulted from the "omics approach" yet.

In a functional or mechanistic approach we asked the question from a different angle, as we found it difficult to understand how an innate reaction should a priori be able to discriminate sensitizers from irritants. So we had to ask what distinguishes allergens from irritants. The answer is, as pointed out above, that the key feature is the intrinsic reactivity of sensitizers. We then turned the question around, "Can cells sense reactivity of molecules?" And indeed this is the case: There is much literature on anti-oxidant response element (ARE) regulated genes. These genes are induced by the KeapI-Nrf2-ARE signalling pathway, which specifically reacts to electrophiles and forms a "cellular sensor" for reactive molecules (Dinkova-Kostova et al., 2005). The general scheme of this pathway and how it is presumed to be activated by skin sensitizers is summarized in Figure 2. We thus investigated to what extent skin sensitizers are able to induce this pathway and found, in a screening of >100 molecules, that the majority of skin sensitizers do induce ARE-dependent luciferase activity, whereas typical irritants such as SDS, which are false-positive in many assays, do not activate this pathway (Natsch and Emter, 2008; Natsch et al., 2009). This original work was performed in the cell line AREc32 derived from a breast cancer cell line (Xiu et al., 2006). We have recently further developed this approach by developing the stable cell line KeratinoSens, which contains a stable insertion of a luciferase reporter gene linked to the anti-oxidant response element from the gene AKR1C2, which had been found to be a sensitizer-induced gene in gene-chip and RT-PCR studies (Gil-

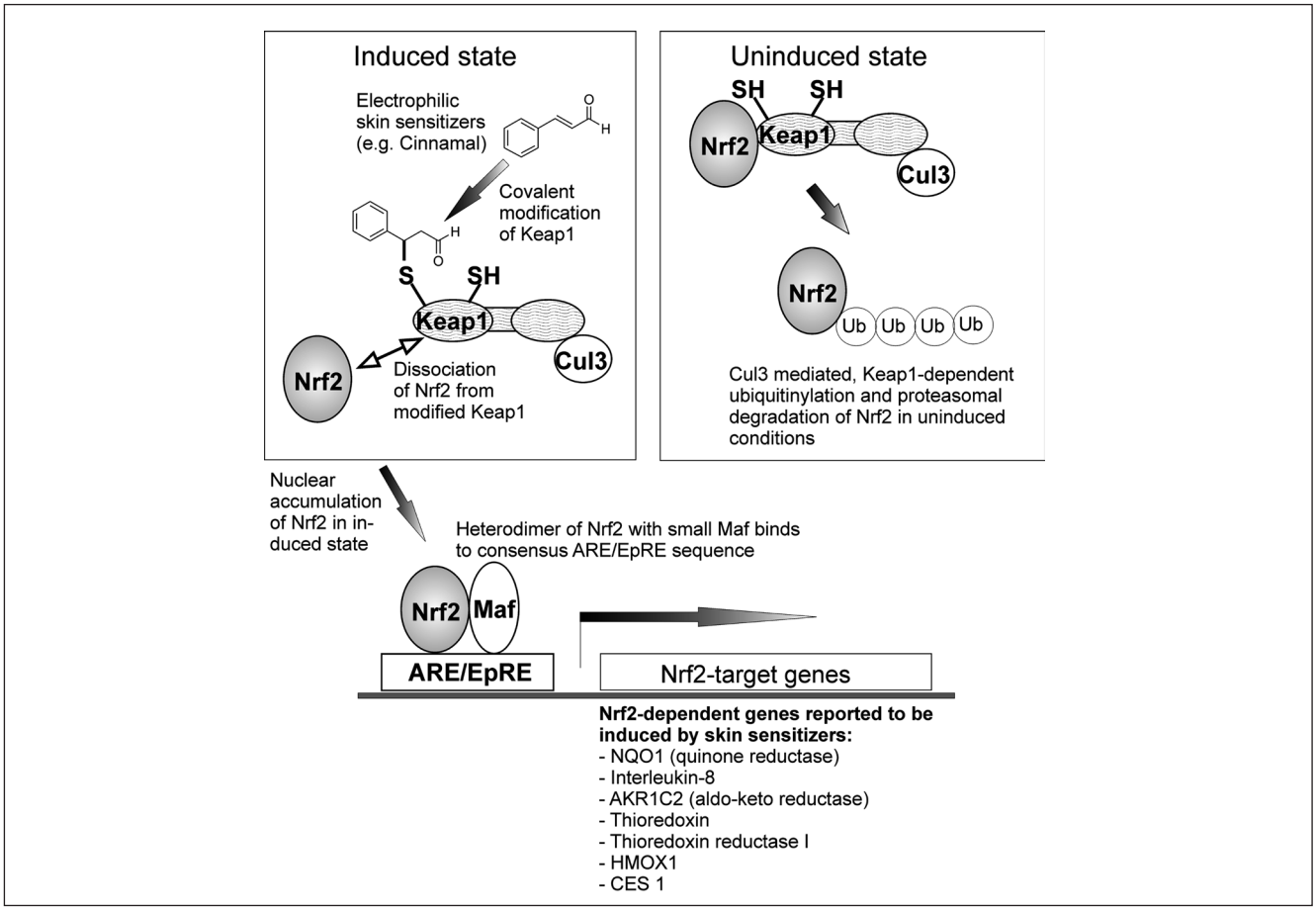


Fig. 2: A general view of the Keap1-ARE-Nrf2 pathway and how it is conceived to be induced by skin sensitizers along with the list of Nrf2-dependent genes specifically reported to be activated by skin sensitizers.

dea et al., 2006; Ryan et al., 2004). This assay yields very stable results and is amenable to high-throughput screening. A standard operating procedure was developed: It involves testing of a full dose-response curve at 12 concentrations for each chemical, simultaneously evaluating ARE-dependent luciferase induction and cell viability. Typical dose-response curves from this detailed and standardized analysis are shown in Figure 3 for the skin sensitizer benzyl-salicylate in comparison to the structurally related skin irritant salicylic acid. A clear dose-dependent gene induction at sub-cytotoxic concentrations of benzyl-salicylate can be seen. The gene induction for most compounds is not linked to cytotoxicity and starts at clearly subcytotoxic concentrations. This is in contrast to some assays which perform tests always at slightly cytotoxic concentrations. Most importantly, the system reacts in the presence of many weak and moderate sensitizers, which often remain undetected in a number of *in vitro* tests. This assay thus yielded an accuracy of 93.8% in the list of chemicals proposed by ECVAM for the development of alternative tests (Casati et al., 2009), and it gives an accuracy of 90.9% for the list of chemicals published by ICCVAM (http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/AppxE.pdf) to validate alternative endpoints in the LLNA (Emter et al.). Whereas our screening results at first hand only proved a good correlation between ARE-dependent gene induction and sensitization potential, a number of recent studies have proven the *in vivo* relevance of this signalling pathway in the skin sensitization reaction (Kim et al., 2008) and the induction of this pathway in dendritic cells (Ade et al., 2009; Python et al., 2009). A review of these studies and a detailed hypothesis on the biological relevance are presented elsewhere (Natsch, Toxicological Sciences, 2010).

2.6 Modelling dendritic cell emigration

As the emigration of dendritic cells from the skin may be a decisive step in the sensitization phase, several research groups have developed dendritic cell migration assays (see e.g. http://www.sens-it-iv.eu/files/newsletter/Sens-it-iv_Newsletter_22.html). These tests, very elegant in their design and relevant in the mechanistic step they do reflect, are technically demanding and still at early development stages, and they are thus not reviewed further here.

2.7 Modelling T-cell activation

The real hallmark of the induction phase of skin sensitization is the proliferation of specific T-cell clones. Thus the "ideal *in vitro* assay" would directly model this process. Indeed, several research groups were able (see e.g. http://www.sens-it-iv.eu/files/newsletter/Sens-it-iv_Newsletter_26.pdf) to bring a sensitizer specific T-cell proliferation reaction into the test tube. In principle, a donor (tests were run both with mice and men) must donate both dendritic cell progenitors and a mixture of naïve T-cells. The dendritic cells are derived from the progenitor cells by *in vitro* differentiation with a specific cytokine cocktail. They are then treated with the test compound and presented to the mixed T-cell population. Proliferation of T-cells is then monitored after 3-10 days by thymidine incorporation assays or by measuring IFN-γ positive cells. As long as mice are used, this is still an *ex vivo* animal test. However, with human volunteers donating the progenitors for the dendritic cells and the T-cells, it can be run entirely without animal tissue. From a conceptual point of view this certainly is

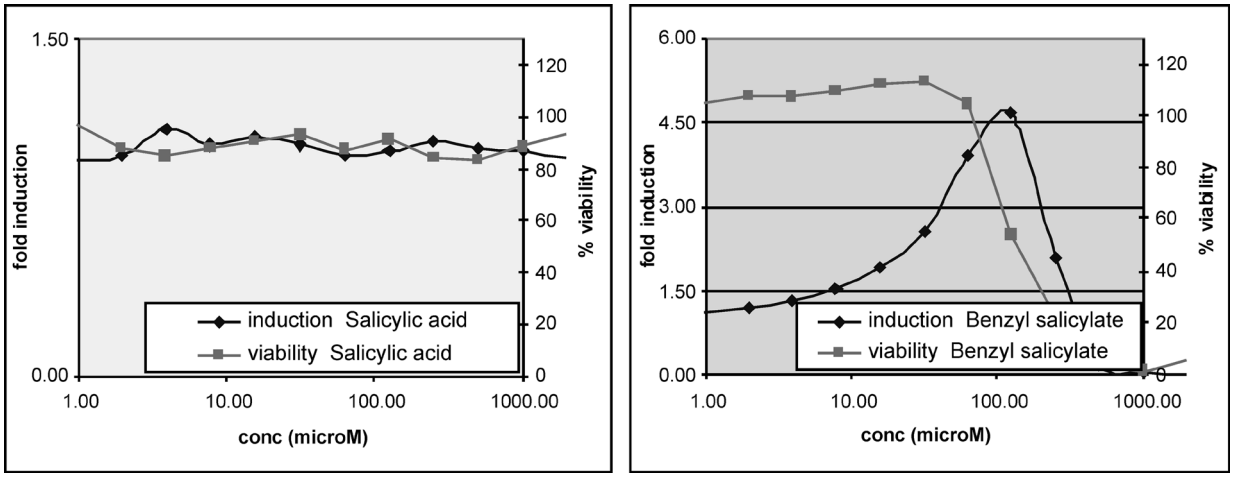


Fig. 3: Example of a full dose response in the KeratinoSens assay. The skin sensitizer benzyl-salicylate induces a significant activity of the ARE-regulated luciferase activity at subtoxic concentrations. The skin irritant salicylic acid does not induce any gene activity.

the "ideal assay", bringing some key steps of the sensitization phase into a single *in vitro* test. Yet most published data with this approach only showed proliferation induced by extreme sensitizers such as DNCB and DNFB. The technical complexity of this assay with the prolonged cultivation and *in vitro* differentiation of different primary cells types should also not be underestimated. In addition, since this assay requires a full array of T-cells, it relies on human donors and cannot be done with stable cell lines, which brings with it an intrinsic donor-to-donor variability and which will make validation of this "ideal assay" a major challenge.

3 Integrating the information

As outlined and reviewed above, there are many steps in the skin sensitisation reaction, all being modelled with different assays. Whereas some authors proposed, that only the key and rate limiting step needs to be modelled for the prediction of skin sensitization (Roberts and Aptula, 2008), the general paradigm still widely held states that a "battery" of tests will be needed to replace the holistic animal tests, and that data from these different tests need somehow to be integrated. This concept had been formalized for the first time by Jowsey et al. (Jowsey et al., 2006), and this paper was frequently cited. We had recently tried, based on the simplified array of tests available in our laboratory, to fill this concept with data. We integrated data from the peptide reactivity test, the AREc32 test, modelling based on TIMES SS and bioavailability (simply expressed as cLogP) (Natsch et al., 2009). Statistical analysis indicated that both peptide reactivity and ARE-dependent gene induction equally contributed to the prediction of the sensitization potential and that the combination of these two assays may offer a first simplified battery of experimentally straightforward assays. A mathematically more elaborate approach to integrate data from various tests is now being developed in the European OSIRIS project (http://www.osiris.ufz.de/), in which data integration to model skin sensitization forms a case study.

4 Conclusion

An array of assays is being developed to predict skin sensitization, and it is generally assumed that integrating these assays will yield best predictions. While the first assays are already in the pre-validation phase, there are key questions to be addressed:

1) Which assays are most specific, discriminating irritants and non-sensitizers from sensitizers?
2) Which assays yield additive information if used in a battery and which ones give redundant information, just addressing the same question with different methods?
3) How can data best be integrated from a weight-of-evidence point of view?
4) How can the specificity of a battery of assays be assured? If in a battery of assays any positive indication from a single assay is used to rate a chemical positive, and each single assay has a somewhat limited specificity generating some false-positives, too many chemicals may turn out as false-positives, as is currently probably the case by combining the evidence from all different genotoxicity tests for single chemicals.

References

Ade, N., Leon, F., Pallardy, M. et al. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: Role of the Keap1/Nrf2 pathway. *Toxicol. Sci. 107*, 451-460.

Aeby, P., Wyss, C., Beck, H. et al. (2004). Characterization of the sensitizing potential of chemicals by in vitro analysis of dendritic cell activation and skin penetration. *J. Invest. Dermatol. 122*, 1154-1164.

Aleksic, M., Thain, E., Roger, D. et al. (2009). Reactivity profiling: Covalent modification of single nucleophile peptides for skin sensitization risk assessment. *Toxicol. Sci. 108*, 401-411.

Aleksic, M., Pease, C. K., Basketter, D. A. et al. (2008). Mass spectrometric identification of covalent adducts of the skin allergen 2,4-dinitro-1-chlorobenzene and model skin proteins. *Toxicology in Vitro 22*, 1169-1176.

Aleksic, M., Pease, C. K., Basketter, D. A. et al. (2007). Investigating protein haptenation mechanisms of skin sensitisers using human serum albumin as a model protein. *Toxicology in Vitro 21*, 723-733.

Antonopoulos, C., Cumberbatch, M., Mee, J. B. et al. (2008). IL-18 is a key proximal mediator of contact hypersensitivity and allergen-induced Langerhans cell migration in murine epidermis. *Journal of Leukocyte Biology 83*, 361-367.

Basketter, D., Pease, C., Kasting, G. et al. (2007). Skin sensitisation and epidermal disposition: The relevance of epidermal disposition for sensitisation hazard identification and risk assessment: The report and recommendations of ECVAM workshop 59a. *ATLA 35*, 137-154.

Basketter, D. A., Evans, P., Fielder, R. J. et al. (2002). Local lymph node assay – Validation, conduct and use in practice. *Food and Chemical Toxicology 40*, 593-598.

Bergström, M. A., Ott, H., Carlsson, A. et al. (2007). A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites. *Journal of Investigative Dermatology 127*, 1145-1153.

Bergström, M. A., Luthman, K., Nilsson, J. L. G. and Karlberg, A. T. (2006). Conjugated dienes as prohaptens in contact allergy: In vivo and in vitro studies of structure-activity relationships, sensitizing capacity, and metabolic activation. *Chemical Research in Toxicology 19*, 760-769.

Casati, S., Aeby, P., Kimber, I. et al. (2009). Selection of chemicals for the development and evaluation of in vitro methods for skin sensitisation testing. *ATLA 37*, 305-312.

Corsini, E., Mitjans, M., Galbiati, V. et al. (2009). Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitizers from irritants and low molecular weight respiratory allergens. *Toxicol. In Vitro 23*, 789-796.

Cotovio, J., Roguet, R., Pion, F. X. et al. (1996). Effect of imidazole derivatives on cytochrome P-450 enzyme activities in a reconstructed human epidermis. *Skin Pharmacol. 9*, 242-249.

Cumberbatch, M., Dearman, R. J., Antonopoulos, C. et al. (2001). Interleukin (IL)-18 induces Langerhans cell migration by a tumour necrosis factor-α- and IL-1α-dependent mechanism. *Immunology 102*, 323-330.

Dinkova-Kostova, A. T., Holtzclaw, W. D. and Kensler, T. W. (2005). The role of Keap1 in cellular protective responses. *Chemical Research in Toxicology 18*, 1779-1791.

Emter, R., Ellis, G., Natsch, A.; Performance of a novel keratinocyte-based reporter cell line to sreen skin sensitizers in vitro. *Toxicol. Appl. Pharmacol.* (in press).

Gerberick, F., Aleksic, M., Basketter, D. et al. (2008). Chemical reactivity measurement and the predictive identification of skin sensitisers. *ATLA 36*, 215-242.

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M. et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicol. Sci. 97*, 417-427.

Gerberick, G. F., Vassallo, J. D., Bailey, R. E. et al. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci. 81*, 332-343.

Gildea, L. A., Ryan, C. A., Foertsch, L. M. et al. (2006). Identification of gene expression changes induced by chemical allergens in dendritic cells: Opportunities for skin sensitization testing. *Journal of Investigative Dermatology 126*, 1813-1822.

Hooyberghs, J., Schoeters, E., Lambrechts, N. et al. (2008). A cell-based in vitro alternative to identify skin sensitizers by gene expression. *Toxicology and Applied Pharmacology 231*, 103-111.

Jowsey, I. R., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). A future approach to measuring relative skin sensitising potency: A proposal. *Journal of Applied Toxicology 26*, 341-350.

Kasting, G. B., Miller, M. A. and Bhatt, V. D. (2008). A spreadsheet-based method for estimating the skin disposition of volatile compounds: application to N,N-diethyl-m-toluamide (DEET). *J. Occup. Environ. Hyg. 5*, 633-644.

Kim, H. J., Barajas, B., Wang, M. and Nel, A. E. (2008). Nrf2 activation by sulforaphane restores the age-related decrease of T(H)1 immunity: role of dendritic cells. *J. Allergy Clin. Immunol. 121*, 1255-1261 e1257.

Lambrechts, N., Verstraelen, S., Lodewyckx, H. et al. (2009). THP-1 monocytes but not macrophages as a potential alternative for CD34+ dendritic cells to identify chemical skin sensitizers. *Toxicology and Applied Pharmacology 236*, 221-230.

Natsch, A., Emter, R. and Ellis, G. (2009). Filling the concept with data: Integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicol. Sci. 107*, 106-121.

Natsch, A., 2010. The Nrf21-AREtoxicity pathway as a cellular sensor for skin sensitizers – functional relevance and a hypothesis on innate reactions to skin senstizers. *Toxicol. Sci. 113*, 284-292

Natsch, A. and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci. 102*, 110-119.

Natsch, A. and Gfeller, H. (2008). LC-MS-based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. *Toxicol. Sci. 106*, 464-478.

Natsch, A., Gfeller, H., Rothaupt, M. and Ellis, G. (2007). Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicology in Vitro 21*, 1220-1226.

Python, F., Goebel, C. and Aeby, P. (2009). Comparative DNA microarray analysis of human monocyte derived dendritic cells and MUTZ-3 cells exposed to the moderate skin sensitizer cinnamaldehyde. *Toxicol. Appl. Pharmacol.*

Roberts, D. W. and Natsch, A. (2009). High Throughput Kinetic Profiling Approach for Covalent Binding to Peptides: Application to Skin Sensitization Potency of Michael Acceptor Electrophiles. *Chem Res Toxicol.,* E-Pub ahead of print.

Roberts, D. W. and Aptula, A. O. (2008). Determinants of skin sensitisation potential. *Journal of Applied Toxicology 28*, 377-387.

Ryan, C. A., Gildea, L. A., Hulette, B. C. et al. (2004). Gene expression changes in peripheral blood-derived dendritic cells following exposure to a contact allergen. *Toxicology Letters 150*, 301-316.

Sakaguchi, H., Ashikaga, T., Miyazawa, M. et al. (2006). Development of an in vitro skin sensitization test using human cell lines; human Cell Line Activation Test (h-CLAT) II. An inter-laboratory study of the h-CLAT. *Toxicology in Vitro 20*, 774-784.

Van Och, F. M. M., Van Loveren, H., Van Wolfswinkel, J. C. et al. (2005). Assessment of potency of allergenic activity of low molecular weight compounds based on IL-1? and IL-18 production by a murine and human keratinocyte cell line. *Toxicology 210*, 95-109.

Xiu, J. W., Hayes, J. D. and Wolf, C. R. (2006). Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of Nrf2 by cancer chemotherapeutic agents. *Cancer Research 66*, 10983-10994.

Correspondence to
Andreas Natsch
Givaudan Schweiz AG
Ueberlandstrasse 138
8600 Duebendorf
Switzerland
e-mail: andreas.natsch@givaudan.com

In Vitro Dermal Toxicity Testing of Cosmetic Ingredients: An Iterative Approach to Complete Regulatory Acceptance

Marleen Pauwels and Vera Rogiers

Department of Toxicology, Vrije Universiteit Brussel, Belgium

Summary

The safety evaluation of cosmetic ingredients belonging to the Annexes of the EU Cosmetic Directive 76/768/EEC, namely colorants, preservatives, UV-filters and substances for which some concern exists with respect to human health (e.g. hair dyes), is carried out at the Commission level by the so-called Scientific Committee on Consumer Safety (SCCS, formerly SCCP and SCCNFP). The reports of these scientific exercises are publicly available "opinions" that can be freely downloaded from the European Commission web pages. They contain valuable information and data with respect to the implementation and use of animal experiments and/or validated alternative methods in the safety evaluation process of cosmetic ingredients in the EU.

This multitude of knowledge has been translated into a carefully tailored searchable database. This VUB-database is used here to gain post-validation information on dermal toxicity tests carried out on cosmetic ingredients, with particular emphasis on the implementation and use of validated 3R-alternative methods.

It became clear that post-validation studies are very useful (e.g. Local Lymph Node Assay, LLNA; 3T3 Neutral Red Uptake Phototoxicity Test, 3T3 NRU PT) to continuously double-check and update the applicability domain and to follow up the implementation and use of alternative methods in other sectors. Furthermore, pro-active cooperation with "regulators" speeds up the regulatory acceptance process in the long run (e.g. eye irritation tests).

When all parties involved work together in this iterative approach, this can help to establish a more efficient regulatory acceptance of 3R-alternatives.

Keywords: 3R-alternatives, cosmetics, dermal toxicity, regulatory acceptance, post-validation

Introduction: The European legislative context of 3R-alternatives for cosmetic safety evaluation

Cosmetic safety evaluation is a complex process, which in Europe requires knowledge of the extensive web of Directives, Regulations and Recommendations intended to ensure the free movement and safe use of cosmetic products (Pauwels and Rogiers, 2007).

Of most relevance is the Cosmetic Products Directive (EU, 1976), laying down the principal rules for marketing and labelling of cosmetic products in the EU. One of its major provisions is that every cosmetic product on the EU market must be safe for use. In order to substantiate this, a specific set of technical data must be assembled and made readily available to the competent authorities of the EU Member States. This so-called "dossier" needs to take into consideration the toxicological profile of the ingredients, their chemical structure and their level of exposure. A second and most debated feature of the EU cosmetic legislation is that the safety of the finished products and their ingredients must be guaranteed without the performance of animal experiments (EU, 2003). This is substantiated in a European animal testing ban on cosmetic ingredients from 11 March 2009 on, accompanied by a gradual marketing ban for cosmetic ingredients tested on animals, becoming final on 11 March 2013. As such, cosmetics are still considered to be inherently safe, but the tools that have enabled the scientific exercise of hazard and risk assessment of cosmetic ingredients to date will be significantly restricted in the near future and are already now limited to a certain extent.

Alternative methods are usually situated in one of the 3R-fields, being refinement, reduction and replacement as introduced by Russell and Burch (Russell et al., 1959), and can be applied to answer a variety of questions for different types of substances and products during their development. When used, however, in a regulatory context of safety assessment in Europe, alternative methods need to fulfil certain criteria. Within the current chemical EU Regulation N° 1907/2006, commonly referred to as "REACH" (Registration, Evaluation, Authorization, Restriction of Chemicals) (EU, 2006), "suitable" alternative methods are defined as "being sufficiently well developed methods according to internationally agreed test development criteria". For cosmetics, however, only "validated" alternative methods (EU, 2003) are allowed and these are further restricted to only replacement methods, shortening the list of currently available validated alternatives.

For long-term toxicity testing (repeated dose toxicity, chronic toxicity, carcinogenicity and reproductive toxicity) and toxicokinetics no validated replacement alternatives currently exist.

This becomes problematic for quantitative risk assessment and toxic potency determination, in particular after March 2013, if no real scientific breakthrough occurs.

With respect to acute and local toxicity, the situation is more favourable (see Table 1).

For the safety evaluation of cosmetic ingredients in the EU, two distinct channels are operative, namely one at the Commission level, being carried out by the SCCS (Scientific Committee on Consumer Safety, previously called SCCP[1] and SCCNFP[2]), and one at the cosmetic industry level, done by independent safety assessors (see Fig. 1). The latter group evaluates all ingredients present in a cosmetic product, whereas the SCCS (and former SCC(NF)P) focuses on the safety of the ingredients present on the Annexes of Directive 76/768/EEC (EU, 1976). These are colorants, preservatives, UV-filters and substances for which some health concerns exist, such as hair dyes.

Over time, the former SCC(NF)P and SCCS today, have studied an important number of cosmetic ingredients' human health safety dossiers and more than 270 opinions have been issued. In particular during the last years, some results ob-

Tab. 1: Validated 3R-alternative methods available today.

Validated replacement alternatives available	No validated replacement alternatives available	
→ endpoints not affected by EU testing or marketing ban	→ endpoints affected by EU testing ban (2009) & EU marketing ban (2009)	→ endpoints affected by EU testing ban (2009) & EU marketing ban (2013)
- skin corrosivity / irritation - dermal absorption - mutagenicity / genotoxicity - phototoxicity	- acute toxicity - eye irritation	- skin sensitization - repeated dose toxicity - carcinogenicity - reproductive toxicity - toxicokinetics

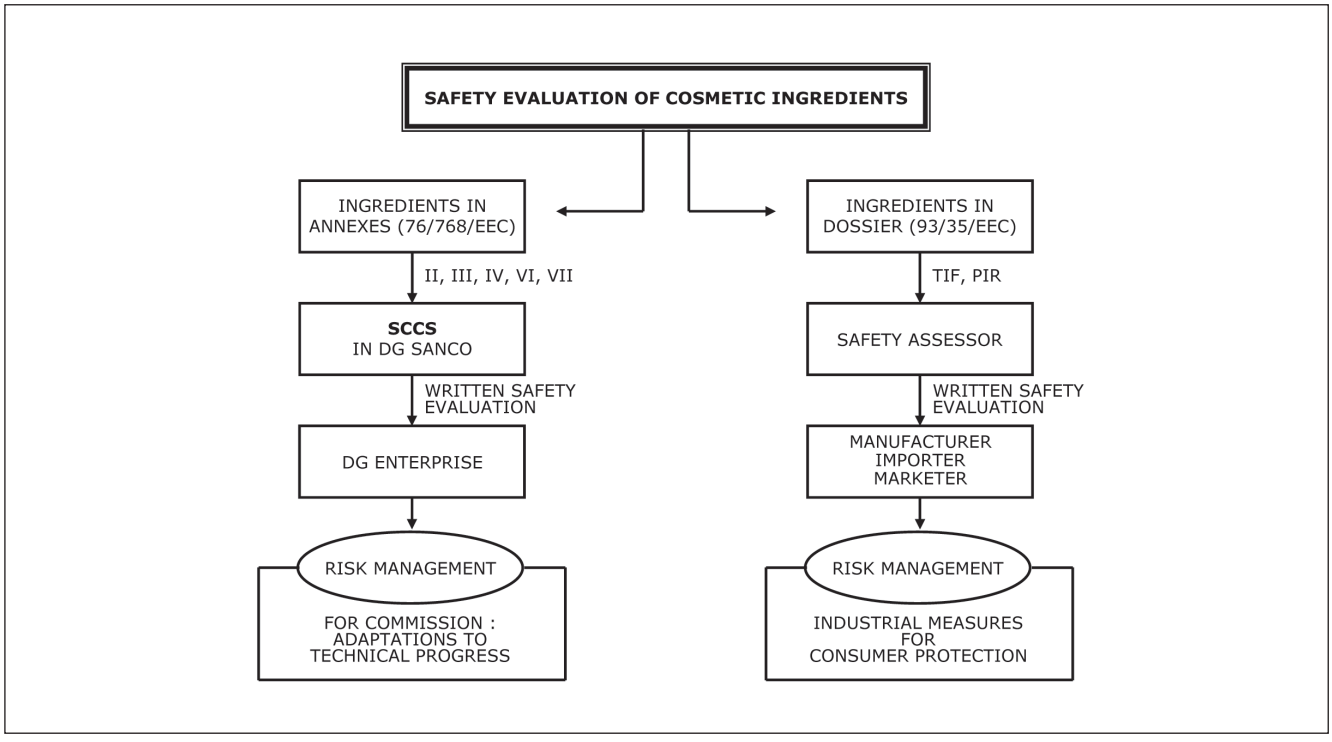


Fig. 1: Two pathways for the safety evaluation of cosmetic ingredients in the EU (SSCP, 2006).
SCCS = Scientific Committee on Consumer Safety
DG = Directorate General
TIF = Technical Information File
PIR = Product Information Requirement

[1] Scientific Committee on Consumer Products
[2] Scientific Committee on Cosmetic and Non-Food Products intended for consumers

tained through alternative methods have been introduced into these dossiers.

As the detailed opinions issued by the SCCS and former SCC(NF)P can be freely downloaded from the European Commission's web pages[3], the idea emerged to translate this multitude of knowledge into a carefully tailored searchable database (Rogiers and Pauwels, 2008).

With the help of this tool relevant information can be gathered, e.g. the typical content of an EU cosmetic ingredient dossier; most commonly encountered problems in the safety assessment process of cosmetics and their ingredients; in-depth information on the implementation of 3R-alternatives in the risk assessment process, etc. (Pauwels and Rogiers, 2009).

Here the updated database has been used to gain in particular post-validation information and to check the implementation and use of validated 3R-alternative methods in the safety evaluation of cosmetic ingredients present on the Annexes. In this work, the focus lies on local toxicity testing.

Materials and Methods

Use was made of the Microsoft Access Database present at the Vrije Universiteit Brussel (VUB), Department of Toxicology, as described previously (Rogiers and Pauwels, 2008).

It contains information extracted from 109 SCCNFP and 145 SCCP documents, dated from February 2000 to February 2009. They reflect more than 270 individual SCC(NF)P opinions on 221 substances in total.

As only one SCCS opinion has become publicly available at the time of this publication, the results presented here cover the work of the SCC(NF)P only.

Results

(i) Skin corrosion

Although the number of validated alternative tests in the field of skin corrosion is high and the tests are so-called replacement tests, they were not encountered in the SCCS and SCC(NF)P dossiers between 2000 and 2009.

(ii) Skin irritation

Very recently, validated skin irritation replacement tests became available, namely the Episkin™ method (ESAC, 2007a; ESAC, 2009), the modified EpiDerm™ SIT[4] (ESAC, 2008; ESAC, 2009) and the Skin Ethic™ RHE[5] assay (ESAC, 2008; ESAC, 2009). For obvious reasons, none of these was yet encountered in the SCCS and SCC(NF)P submissions made between 2000 and 2009.

Of the 129 cosmetic compounds for which *in vivo* skin irritation data were available in the dossiers submitted, 86 were non-irritating, 23 were slightly/mildly irritating and 17 were clear skin irritants. One was even severely irritating and one was corrosive. One compound was giving indecisive results. 31 of the 129 compounds provoked *in vivo* discoloration of the skin. The indecisive result mentioned above, was due to discoloration of the skin.

These *in vivo* data were shared with ECVAM (European Centre for the Validation of Alternative Methods) and the cosmetic industry, since it is believed that these cosmetic ingredients could be of further interest as reliable *in vivo* data are available for these compounds. They could therefore represent interesting reference molecules for further validation studies.

(iii) Eye irritation

Validated eye irritation tests that can be used in quantitative risk assessment are not yet available. Eye irritation screening tests do exist. These are the BCOP (Bovine Cornea Opacity Permeability) test, HET-CAM (Hen's Egg Test-Chorio Allantoic Membrane) test, NRU (Neutral Red Uptake) test and RBC (Red Blood Cell) test (ESAC, 2007b).

In the SCC(NF)P dossiers (2000-2009), *in vivo* tests were present for 135 substances, and *in vitro* screenings for 14 substances. Only in three cases were *in vitro* methods used as a stand-alone. One compound turned out to be a severe skin irritant. Of the latter a previous rabbit skin test showed severe irritancy, wherefore instillation into the eye was excluded for ethical reasons. For the others, combinations of *in vitro* tests were accepted based on expert judgment. Out of the *in vivo* eye irritation data it appeared that 30% were non-irritating, 20% slightly to mildly irritating, 28% (moderately) irritating and 17% severely irritating.

(iv) Skin sensitization

The murine LLNA (Local Lymph Node Assay) is a validated test (EU, 2008d; OECD, 2002b), considered to be a stand-alone assay for testing of skin sensitization. The data presented in Table 2 were retrieved from the SCC(NF)P dossiers. They show that after 2002, representing the official incorporation date of the test into OECD Guideline 429, the LLNA was well incor-

Table 2: Occurrence of LLNA data for SCC(NF)P dossiers (2000-2009)

Test(s) performed	< 2002	≥ 2002
Non-LLNA	71	1
Both non-LLNA and LLNA	26	
Only LLNA		41

[3] http://ec.europa.eu/health/ph_risk/committees/sccp/sccp_opinions_en.htm (consulted 08/2009)
http://ec.europa.eu/health/ph_risk/committees/04_sccp/sccp_opinions_en.htm (consulted 08/2009)
http://ec.europa.eu/health/ph_risk/committees/04_sccs/sccs_opinions_en.htm (consulted 08/2009)

[4] Skin Irritation Test

[5] Reconstructed Human Epidermis

porated in the cosmetic dossiers studied. The results of "old" guinea pig studies, carried out before 2002, were still present in the dossiers together with the "new" LLNA results. This was found in nearly 20% of all cases, but in 30% LLNA results were shown as stand-alone tests. There was also one illegal case for which non-LLNA data were produced after the 2002 deadline.

Where both LLNA data and guinea pig assay data are available, it is possible to check how well the results correlate. The 26 cases present in our database (see Fig. 2) showed that there is a relatively good correlation. When coloration was a potential problem in the guinea pig test, the LLNA data could eliminate this in 4/5 of the cases available to us.

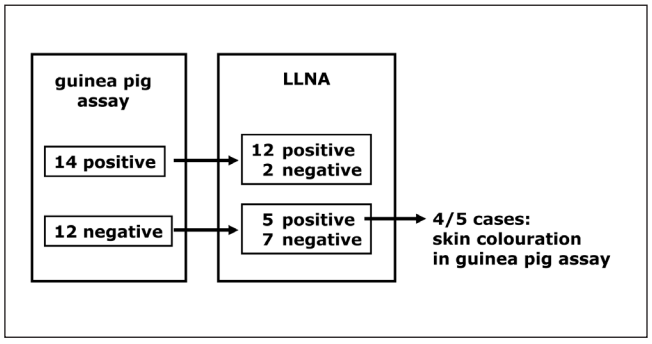


Fig. 2: Availability of both LLNA and guinea pig data in testing of skin sensitization by cosmetic substances in SCC(NF)P dossiers (2000-2009).

(v) Phototoxicity testing

The 3T3 NRU PT (3T3 Neutral Red Uptake Phototoxicity Test) is a validated replacement test that became available in Europe in 2000 as EC method B.41 and as an OECD Guideline 432 in 2004 (EU, 2000; OECD, 2004c).

The test is of particular interest for UV-filters used in sunscreens and other cosmetic products such as day creams and anti-aging preparations. The test is also used for chemicals and pharmaceuticals absorbing UV.

The database contains 27 substances with phototoxicity assays. For 14 of these a 3T3 NRU PT assay is provided but only in 4 instances the test was used as a stand-alone test. *In vivo* phototoxicity data were found for 23 substances in total and were mainly performed on guinea pigs, mice and human volunteers.

A surprising finding was that four animal phototoxicity assays were reported with testing dates after 2000, i.e. when these tests were no longer legal.

Discussion

(i) Skin corrosion

Although in Europe three validated replacement tests, namely the TER (Transcutaneous electrical resistance) assay (EU, 2008b; OECD, 2004a) and two human skin models Episkin™ and EpiDerm™ (EU, 2008c; OECD, 2004b), are available to measure the skin corrosive properties of chemical substances, it is not surprising that no replacement alternatives were encountered in the cosmetic ingredient dossiers examined between 2000 and 2009 by the SCCS and former SCC(NF)P.

Indeed, cosmetic ingredients are usually selected for their mild character and when strong acids or bases are present in cosmetics, it is often for their neutralising properties and to adjust the pH of a cosmetic formula. Only in the case of production mistakes or serious misuse of cosmetics may corrosivity occur.

(ii) Skin irritation

That no alternative tests for skin irritation were found in the submissions examined between 2000 and 2009 is probably not only due to their quite recent introduction but also to the problems encountered with colorants and reductive substances.

Although Episkin™ and other reconstructed human skin equivalents are promising replacement alternatives, some concern was expressed by the SCCP with regard to the applicability domain (SCCP, 2007b). More specifically, the endpoint of the test consists of a so-called MTT[6] reduction, in which colour formation is essential. Objective observations may thus be affected in the cases where hair dyes, colorants and reductive substances are present. The MTT colorimetric method for dyes was already described as problematic before in another setting, namely the use of the Episkin™ model for the assessment of *in vitro* phototoxicity (Lelièvre et al., 2007). This limit in applicability domain has been taken up in the recent OECD draft on *in vitro* testing of skin irritation (OECD, 2009).

As the deadline of 11 March 2009 has passed, all cosmetic ingredients that need testing for their irritative and non-irritative properties must be analysed, both in and outside Europe, using a validated replacement method. Otherwise they cannot be marketed within the EU. The need for legislative approval is thus very urgent, since quite a number of cosmetic ingredients seem to be potential skin irritants, as seen from the data analysed via our database.

(iii) Eye irritation

Eye irritation is an important safety test for cosmetic ingredients and finished products, in particular when they are intended to be applied to the face and around the eyes. As the deadline for *in vivo* eye irritation testing has passed, both in and outside the Member States (11 March 2009), the validation of the *in vitro* methods that are presently in the pipeline becomes very urgent. It is interesting to mention that high quality *in vivo* data could be retrieved from the SCC(NF)P and SCCS dossiers and could potentially be used as a basis for ingredient selection for coming validation studies that are necessary in the immediate future for cosmetic ingredients.

(iv) Skin sensitization

For more than 35 years two guinea pig assays have been routinely performed for the evaluation of the sensitizing potential

[6] 3-(4,5)-dimethyl-2-thiazolyl-2,5-dimethyl-2H-tetrazolium bromide

of test substances, the Magnusson-Kligman GPMT and the Buehler test (EU, 2008a; OECD, 2002a).

An alternative method, the murine LLNA, was endorsed by ESAC (ECVAM Scientific Advisory Committee) in 2000 (ESAC 2000) and translated into OECD Guideline 429 in 2002 (OECD, 2002b). It is a reduction and refinement animal test using mice instead of guinea pigs and it requires that the animal is only subjected to the induction phase and not the elicitation phase. As more than one substance application is inherent to the test to demonstrate contact allergic reactions, the test is seen as a kind of "repeated dose" test and therefore may be used in a legal setting until 11 March 2013 for testing outside the Member States for cosmetic ingredients destined for the European market (SCCP, 2007a).

From the results described (Table 2 and Fig. 2) it can be learned that the LLNA is used as a stand-alone assay more and more frequently. It has some clear advantages over the "old" guinea pig assay, in particular when coloration occurs after exposure to the compound under investigation.

Indeed we could analyse 26 cases. In 14 of these, a positive reaction was obtained using the guinea pig assay whereas 12 were positive in the LLNA.

Of the 12 negative ones in the guinea pig assay, 5 were positive and 7 negative in the LLNA. 4 of these 5 cases showed skin coloration in the guinea pig assay, pointing towards the usefulness of the LLNA to avoid some false negative results due to coloration problems.

(v) Phototoxicity testing

Some problems are still encountered with the validated 3T3 NRU PT. It has been reported that false positives occur (De Smedt, 2007), leading to extra testing. The latter can be done using the 3-dimensional human skin equivalents that are commercially available (Liebsch et al., 1995), but that strategy has not yet been validated as a whole concept.

Problems are often related to the solubility and concentration of the compounds, their pH, the type of UV source used, the occurrence of complex mixtures, the use of different non-standardised protocols, the bioavailability of the compound, etc. False positives may lead to loss of valuable compounds and should therefore be avoided as much as possible.

Some general considerations are important:

- The applicability domain of alternative methods needs to include those substances that are subject to the most stringent legislative requirements. As in Europe this happens to be the case for cosmetic ingredients, it is important that in the validation process of alternative methods enough cosmetic substances, belonging to the Annexes of Directive 76/768/EEC, are included. In our data analysis, the example of Episkin™ became quite prominent.
- From the results present in the database with respect to eye irritation, it also became clear that identification of and proactive cooperation with regulators (e.g. SCCS) can speed up the regulatory acceptance process in the long run.
- From the data present in the VUB-database, with respect to the LLNA and the 3T3 NRU PT alternatives, it can be learned that post-validation studies are indispensable to continuously double-check and update the applicability domain, to optimise alternative methods over time and to follow up the implementation and use of alternatives in other sectors.

In conclusion, if all parties involved work together in this iterative approach, a more efficient regulatory acceptance of 3R-alternatives can probably be achieved.

Acknowledgements

The authors wish to thank the scientific and administrative colleagues at the SCCS and former SCC(NF)P, at Colipa and the VUB for their constructive collaboration and useful advice.

References

De Smedt, A. (2007). In vitro 3T3 NRU Phototoxicity test. Oral communication during the Annual conference 2007 of the European Partnership for Alternative Approaches to Animal Testing: "Regulatory Acceptance". 5 November 2007 Brussels, Belgium.

ESAC (2007a). Statement on the scientific validity of In vitro Tests for Skin Irritation. Adopted by the ECVAM Scientific Advisory Committee (ESAC) at its 26th Meeting of 26-27 April 2007 at ECVAM, Ispra, Italy.

ESAC (2007b). Statement on the conclusions of the ICCVAM retrospective study on Organotypic in vitro assays as screening tests (BCOP and ICE) to identify potential ocular corrosives and severe irritants as determined by US EPA, EU (R41) and UN GHS classifications in a tiered testing strategy, as part of a weight of evidence approach. *Adopted by the ECVAM Scientific Advisory Committee (ESAC) at its 26th Meeting of 26-27 April 2007 at ECVAM, Ispra, Italy*.

ESAC (2008). Statement on the scientific validity of In vitro Tests for Skin Irritation testing. Adopted by the ECVAM Scientific Advisory Committee (ESAC) at its 29th Meeting of 4-5 November 2008 at ECVAM, Ispra, Italy.

ESAC (2009). Statement on the performance under UN GHS of three in vitro assays for skin irritation testing and the adaptation of the reference chemicals and defined accuracy values of the ECVAM skin irritation performance standards. Adopted by the ECVAM Scientific Advisory Committee (ESAC) at its 30th Meeting of 9-10 March 2009 at ECVAM, Ispra, Italy.

EU (1976). Council Directive 76/768/EEC of 27 July 1976 on the approximation of the laws of the Member States relating to cosmetic products. *Official Journal L262*, 169-200, 27 September 1976.

EU (2000). B.41. Phototoxicity - in vitro 3T3 NRU phototoxicity test. Commission Directive 2000/33/EC of 25 April 2000 adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal L136*, 98-107, 08 June 2000.

EU (2003). Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the

Member States relating to cosmetic products. *Official Journal L066*, 26-35, 11 March 2003.

EU (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Official Journal L396*, 1-849, 30 December 2006. Corrigendum in *Official Journal L136*, 3-280, 29 May 2007.

EU (2008a). B.6. Skin sensitisation. Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal L 142*, 31/05/2008 p.202.

EU (2008b). B.40. In vitro skin corrosion: transcutaneous electrical resistance test (TER). Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal L 142*, 31/05/2008 p.384.

EU (2008c). B.40 bis. In vitro skin corrosion: human skin model test. Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal L 142*, 31/05/2008 p.394.

EU (2008d). B.42. Skin sensitisation : Local Lymph Node Assay. Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *Official Journal L 142*, 31/05/2008 p.414.

Liebsch M., Döring B., Donelly TA., Logemann P., Rheins L.A. and Spielmann H., 1995. Application of the human dermal model Skin2 ZK 1350 to phototoxicity and skin corrosivity testing. *Toxicology In vitro 9*, 557–562.

OECD (2002a). OECD Guideline for testing of chemicals - Guideline 406: Skin Sensitisation. Organization for Economic Cooperation and Development, Paris, adopted 12 May 1981, last updated 17 July 2002.

OECD (2002b). OECD Guideline for testing of chemicals - Guideline 429: Skin Sensitisation: Local Lymph Node Assay. Organization for Economic Cooperation and Development, Paris, adopted 24 April 2002.

OECD (2004a). OECD 430 - OECD Guideline for testing of chemicals - Guideline 430: In vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER) Organization for Economic Cooperation and Development, Paris, adopted 13 April 2004.

OECD (2004b). OECD 431 - OECD Guideline for testing of chemicals - Guideline 431: In vitro Skin Corrosion: Human Skin Model Test Organization for Economic Cooperation and Development, Paris, adopted 13 April 2004.

OECD (2004c). OECD Guideline for testing of chemicals - Guideline 432: In vitro 3T3 NRU phototoxicity test. Organization for Economic Cooperation and Development, Paris, adopted 13 April 2004.

OECD (2009). Draft proposal for a new Guideline: In Vitro Skin Irritation: Reconstructed Human Epidermis (RhE) Test Method (Version 6). Organization for Economic Cooperation and Development, Paris, 20 March 2009.

SCCP (2006). SCCP/1005/06: The SCCP's Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation, adopted by the SCCP during the 10th plenary meeting of 19 December 2006.

SCCP (2007a). SCCP/1111/07: Memorandum on actual status of alternative methods on the use of experimental animals in the safety assessment of cosmetic ingredients in the European Union, adopted by the SCCP during the 12th plenary meeting of 19 June 2007.

SCCP (2007b). SCCP/1145/07: Memorandum on the in vitro test EPISKIN™ for skin irritation testing, adopted by the SCCP during the 14h plenary meeting of 18 December 2007.

Pauwels, M. and Rogiers, V. (2007). EU legislations affecting safety data availability of cosmetic ingredients. *Regulatory Toxicology and Pharmacology 49*, 308-315.

Pauwels, M. and Rogiers, V. (2009). Critical analysis of the SCCNFP/SCCP safety assessment of cosmetic ingredients (2000-2006). *Food and Chemical Toxicology 47*, 898-905.

Rogiers, V. and Pauwels, M. (Eds.) (2008). *Current Problems in Dermatology Vol.36: Safety Assessment of Cosmetics in Europe*. S. Karger, Basel, Switzerland, ISBN 978-3-8055-8655-9.

Russell, W. M. S. and Burch R. L. (1959). *The principles of Humane Experimental Technique*. Methuen and Co Ltd, London (reprinted by the Universities Federation for Animal Welfare UFAW, 1992, Potters Bar, Herts), UK, 1959.

Correspondence to:
Prof. Dr. Vera Rogiers
Department of Toxicology, Vrije Universiteit Brussel (VUB),
Laarbeeklaan 103,
B-1090 Brussels, Belgium

Session BS22: Skin and eye toxicity 2

ECVAM Bottom-Up/Top-Down Testing Approach: Testing Strategy to Reduce/Replace the Draize Eye Test and Validation/Regulatory Acceptance of *In Vitro* Assays: Current Status

Valérie Zuang, João Barroso, Thomas Cole, Mara Ceridono and Chantra Eskes

ECVAM, In Vitro Methods Unit, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Italy

Summary

To reduce and/or replace the Draize test, testing schemes combining strengths of particular in vitro assays were proposed during a 2005 ECVAM expert meeting. The testing scheme proposes, based on expected irritancy of the test substance, a Bottom-Up approach, beginning with test methods that accurately identify non-irritants, or a Top-Down approach, beginning with test methods that accurately identify severe irritants before progression of further in vitro testing. Furthermore, as its core activity, ECVAM participated in the retrospective validation of and has peer reviewed scientific validity of four organotypic assays and undertook retrospective validation of four cell function/cytotoxicity assays. The BCOP and ICE organotypic assays were ICCVAM and ESAC endorsed as scientifically valid for identifying severe irritants, and OECD Test Guidelines are under adoption. NRR, FL and CM cell function/cytotoxicity assays were recommended by an ECVAM Validation Management Group for identification of non-irritants or severe eye irritants in the Bottom-Up/Top-Down approaches. These assays were peer reviewed by ESAC during 2009. Finally, a joint ECVAM-COLIPA prospective validation study was initiated in 2008 to evaluate two Reconstructed human Tissue assays to discriminate non-classified materials from eye irritants, based on the proposed test strategies. The ultimate goal is to combine validated in vitro assays, based on their performances and applicability domains, to define the most suitable testing strategy to classify substances for eye irritation potential and ultimately replace the Draize test. This manuscript presents the proposed testing scheme and provides details on the validation/regulatory status of in vitro assays for use in this scheme.

Keywords: in vitro test, testing strategy, Bottom-Up testing approach, Top-Down testing approach, eye irritation

1 Introduction

Success in developing and validating alternative tests to fully replace the Draize rabbit eye irritation test has remained elusive in the past, despite major efforts by the European Centre for the Validation of Alternatives (ECVAM), governmental institutions, industry trade associations, individual companies and academia. This was due to many reasons; one of these being that no *in vitro* test taken individually could fully replace the Draize eye test.

In 2004, a thorough review was carried out on the status of the most promising alternatives for eye irritation testing (Eskes et al., 2005). Several recommendations to progress validation efforts in view of replacing the Draize rabbit eye test were made. Among those an important one was to make use of testing strategies that utilise the strengths of individual *in vitro* test methods to address required ranges of irritation potential and/or chemical classes.

To follow up on this, ECVAM organised an expert meeting, where the conceptual framework of promising testing strategies

for eye irritation was defined in collaboration with more than 30 representatives from industries, contract research organisations, regulators, academia and animal welfare organisations in 2005. Two testing schemes were identified proposing the use of a Bottom-Up (begin with using test methods that can accurately identify non-irritants) or Top-Down (begin with using test methods that can accurately identify severe irritants) progression of *in vitro* tests, based on expected irritancy of substances (Scott et al., 2009).

The performances and applicability domains of individual alternative methods, which are considered sufficiently promising to populate the proposed testing strategies, were or are currently being, determined through validation studies, as described below.

2 Conceptual framework of testing strategies

ECVAM convened an expert meeting to allow developers/users to nominate methods to be considered as a basis for an overall testing strategy. Assays were evaluated and categorised on the basis of their applicability domains (e.g. categories of severity of irritation, modes of action, chemical class and physico-chemical compatibility). The analyses were based on the data developed from current practice and published studies, expert opinion, the ability to predict depth of injury (within the applicable range of severity), modes of action that could be addressed and compatibility with different physico-chemical forms. The difficulty in predicting the middle category of irritancy (e.g. R36, GHS Categories 2A and 2B) was recognized. The testing scheme proposes using a Bottom-Up (begin with using test methods that can accurately identify non-irritants) or Top-Down (begin with using test methods that can accurately identify severe irritants) progression of *in vitro* tests, based on expected irritancy of substances as primarily identified by their physico-chemical properties. Irrespective of the starting point, the approach would identify non-irritants and severe irritants, leaving all others to the (mild/moderate) irritant GHS 2/R36 categories.

Fifteen assays were nominated, which can be divided in four major groups: cytotoxicity- and cell function-based assays, reconstructed human tissue models, organotypic assays and other assays. The progress in the validation and regulatory acceptance of the most promising assays for use in these strategies is described hereafter.

3 Cytotoxicity and cell-function based assays

The retrospective validation of four cytotoxicy- and cell function-based assays, i.e. the Neutral Red Release (NRR), the Red Blood Cell test (RBC), the Fluorescein Leakage assay (FL) and the Cytosensor Microphysiometer (CM) took place between May 2006 and October 2008. The study was based on the retrospective collection of existing data compiled according to the ECVAM Modular Approach to Validation and weight-of-evidence principles (Hartung et al., 2004; Balls et al., 2006). Based on the final results, recommendations were made by the Validation Management Group on the validity of the NRR and FL to be used in a Bottom-Up approach, discriminating non-irritants (GHS and EU non-classified) from all other classes, and the CM and FL in a Top-Down approach, discriminating severe irritants (GHS Cat 1, EU R41) from all irritant classes, for defined applicability domains.

In July 2009, ESAC endorsed the CM (Invittox protocol 102 modified) and the FL (Invittox protocol 71) as scientifically valid for being considered for regulatory purposes as an initial step within a Top-Down approach to identify ocular corrosives and severe irritants (GHS Cat 1, EU R41, EPA Cat I) from all other classes for water-soluble chemicals (substances and mixtures). Furthermore, the CM was considered to have been scientifically validated and to be ready for consideration for regulatory use as an initial step within a Bottom-Up approach to identify non-irritants (GHS NC, EU NC, EPA Cat IV) from all other classes, only for water-soluble surfactants and water-soluble surfactant-containing mixtures.

For the FL, additional testing and further refinement, in particular with respect to variability and definition of the applicability domain, by expanding the dataset of tested chemicals and direct comparison with *in vivo* data was recommended. With regard to the remaining tests, ESAC considered that the available evidence was insufficient to support a recommendation that they are ready for consideration for regulatory use.

Different VMG and ESAC recommendations were endorsed for the assays considered useful to initiate a Bottom-Up approach. Possible reasons for such differences could have been:

1) The acceptance criteria applied for the identification of substances not classified as irritants. While the VMG accepted up to 5-10% false negatives in the non severe irritancy range according to the GHS (Cat. 2) and EU (R36) C&L systems and a not too high percent of false positives, the ESAC accepted only 0% false negatives according to the GHS, EU and EPA classification systems, being more flexible towards false positives.
   Regarding the rate of false negatives accepted by the VMG, it should be noted that an ECVAM internal analysis on a total of 2039 new and existing chemicals showed that, by only considering the within-test variability of in vivo responses, the Draize test may underpredict up to 10% Cat. 2 (GHS) or 22% R36 (EU) substances as non-classified. If the in vivo between laboratory variability (BLV) was to be added to the within-test variability, similar to what was considered in the validation study of cell-based assays, the overall in vivo variability might even increase.
2) Differences in the subjective weighing of the evaluated evidence.

In any case, despite their limited applicability domains, these validated assays might contribute to further decreasing animal testing for eye irritation. Indeed, 79% of newly registered substances (out of 2497) were shown to be non-irritants, and 16% severe irritants (Scott et al., 2009). Furthermore, the fact that

for the first time an assay is validated to identify substances not classified as irritants to the eye based on a retrospective weight-of-evidence approach is an important step forward.

4 Reconstructed human Tissue model assays

Following corporate prevalidation and multi-industrial optimisation studies, the optimised protocol for two Reconstructed human Tissue (RhT) models, the SkinEthic™ Human Corneal Epithelium (HCE) and the EpiOcular™ OCL-200 models were submitted by COLIPA to ECVAM in 2008. Further to the request and review of additional data from COLIPA ring trials, ECVAM agreed to launch a validation study on the two RhT model assays.

The primary goal of the validation study is an evaluation of the ability of the *in vitro* tests to reliably discriminate ocular non-irritant (NI) chemicals from irritant ones (i.e. as first step in a Bottom-Up approach), as defined according to the OECD and United Nations proposal for a Globally Harmonised System (GHS) for the classification and labelling of ocular irritation (category 1/category 2, no category; UN, 2003) and as implemented in the European Commission Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006.

The validation study foresees the testing of 104 coded chemicals in 3 runs and in 3 laboratories. Three VMG meetings took already place since December 2008, where strategic decisions were taken including the definition of study design and the type and number of chemicals to be tested, in view of starting the experimental phase of the study at the earliest possible date (Freeman et al., present proceedings).

5 Organotypic assays

Further to an ECVAM survey with regulatory authorities, in July 2004, the European Commission stated in its Manual of Decisions for Implementation of the 6th and 7th Amendments to Directive 67/548/EEC on Dangerous Substances, that positive outcomes of four organotypic assays (Bovine Corneal Opacity & Permeability test (BCOP), Isolated Chicken Eye test (ICE), Isolated Rabbit Eye test (IRE) and Hen's Egg Test on the Chorio-Allantoic Membrane (HET-CAM)) are accepted for the classification and labelling of severe eye irritants, but that a negative result requires confirmation by an in vivo test (EC, 2004). Subsequently, between 2003 and 2006, ICCVAM-NICEATM conducted, with ECVAM collaboration, a retrospective validation study on these four assays, in which the ability of these assays to detect severe eye irritants and ocular corrosives was evaluated. After peer review, two assays, the BCOP and ICE tests, were endorsed as scientifically valid to identify ocular corrosives and severe irritants in the US and in the EU (ICCVAM, 2006; ESAC, 2007). OECD Test Guidelines on the two test methods were prepared by ICCVAM with ECVAM collaboration and were adopted in 2009 as TG 437 for the BCOP and TG 438 for the ICE. For the two other organotypic assays evaluated, the Hen's Egg Test on the Chorio-Allantoic Membrane (HET-CAM) assay and the Isolated Rabbit Eye (IRE) test, ESAC recommended that further work was performed before a statement on their scientific validity to identify ocular corrosives and severe irritants could be made.

With regard to the evaluation of the four organotypic assays for identifying mild or non irritants, a retrospective analysis of the collected data was carried out by ICCVAM. In May 2009, the ICCVAM Ocular Peer Review Panel recommended BCOP for the identification of non-classified materials, which is, as for the CM, an important step forward. None of the test methods was recommended for full replacement, because none of the methods are able to identify the mild/moderate ranges of ocular irritancy.

At ECVAM, follow-up work on further improvements of the prediction models and analyses for prediction of all ranges of irritancy, using data mining techniques, is currently ongoing.

6 Other assays

Finally, two other assays, i.e. the Low Volume Eye Test (LVET), a refinement method, and the Ocular Irritection® assay, underwent both external validation and were submitted to ECVAM for evaluation. ESAC recommended in July 2009, the use of existing LVET data for the use domain of household detergent and cleaning products and their main ingredients (e.g. surfactants) for making eye irritation classification and labelling decisions in a weight of evidence approach, and as reference data for the validation of *in vitro* test methods.

The Ocular Irritection® assay, an *in vitro* method that mimics the biochemical phenomena of corneal protein denaturation and disruption caused by irritant substances acting on the cornea, was submitted to ECVAM in January 2009. Further to the evaluation of the submission, ECVAM requested in May 2009 to provide additional information in a revised submission.

7 Conclusion and future prospects

The evaluation of the individual tests by retrospective and prospective validation studies will allow the determination of the most suitable strategies that utilise the strengths of specific *in vitro* assays for the classification of test substances according to their irritation potential.

While the main role of ECVAM is to coordinate and streamline the validation process of the individual *in vitro* assays available to populate the testing strategies, it is furthermore also closely working, in collaboration with COLIPA, on the construction of testing strategies combining in the most optimal way the different validated methods by using data mining techniques and statistical tools. The purpose of this work is to identify the most promising testing strategies based on the Top-Down and Bottom-Up concept which have amongst others the higher impact

in reducing animal testing, and the higher benefits and lower potential costs linked to testing and (in)correct predictions. Once available, it is foreseen that such strategies will be challenged with a new set of test chemicals prior to being forwarded for regulatory acceptance.

References

Balls, M., Amcoff, P., Bremer, S. et al. (2006). The principles of weight of evidence validation of test methods and testing strategies. *ATLA 34*, 1-19.

EC (2004). Manual of Decisions for Implementation of the 6th and 7th Amendments to Directive 67/548/EEC on Dangerous Substances. Updated version of July 2004 (EUR 20519). 189pp. Ispra, Italy: European Chemicals Bureau, European Commission JRC. Website: http://ecb.jrc.it/DOCUMENTS/New-Chemicals/Manual_of_decisions.pdf (accessed on 28.08.2009)

ESAC (2007). ESAC Statement on the conclusions of the ICCVAM retrospective study on Organotypic in vitro assays as screening tests to identify potential ocular corrosives and severe irritants as determined by US EPA, EU (R41) AND UN GHS classifications in a tiered testing strategy, as part of a weight of evidence approach. Available at: http://ecvam.jrc.ec.europa.eu/ under "Publications", "ESAC statements", accessed on 29.05.2009.

Eskes, C., Bessou, S., Bruner, L. et al. (2005). Subchapter 3.3. Eye Irritation. In C. Eskes and V. Zuang (eds.), Alternative (non-animal) Methods for Cosmetics Testing: Current Status and Future Prospects. *ATLA 33, Suppl. 1*, 47-81.

Freeman, S. J., Alépée, N., Barroso, J. et al. (2010). prospective validation study of reconstructed human tissue models for eye irritation testing. *ALTEX*, present proceedings.

Hartung, T., Bremer, S., Casati, S. et al. (2004). A Modular Approach to the ECVAM Principles on Test Validity. *ATLA 32*, 467-472.

ICCVAM (2006). ICCVAM Test Method Evaluation Report: In Vitro Ocular Toxicity Test Methods for Identifying Severe Irritants and Corrosives. NIH Publication No: 07-4517. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmer.htm. Accessed on 26.06.2009.

Scott, L., Eskes, C., Hoffman, S. et al. (2010). Bottom-up and top-down approach: Eye irritation testing strategy to reduce and replace in vivo studies. *Toxicology In Vitro 24*, 1-10.

UN (2003). *Globally Harmonised System of Classification and Labelling of Chemicals (GHS). Part 3: Health and Environmental Hazards* (107-228). New York, NY, USA, and Geneva, Switzerland: United Nations Organisation.

Correspondence to

Dr. Valérie Zuang
In Vitro Method Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre, European Commission
21027 Ispra (VA), Italy
e-mail: valerie.zuang@jrc.ec.europa.eu

BS22: Skin and eye toxicity 2

ECVAM Drives Expanded EpiOcular Applicability Domain for EU Legislation: Successful International Pre-Validation

Yulia Kaluzhny, Helena Kandarova, Patrick Hayden, Joseph Kubilus, Laurence d'Argembeau-Thornton and Mitchell Klausner

MatTek Corporation, Ashland, MA

Summary

The recently implemented EU Cosmetics directive and REACH legislation has heightened the need for in vitro ocular test methods. In response, the European Centre for Validation of Alternative Methods (ECVAM) eye irritation task force requested that submitters, model developers, and companies involved in EpiOcular™ pre-validation studies work toward expanding the test chemical applicability domain (AD) of the model. The EpiOcular model producer (MatTek Corp.) therefore undertook development of an expanded EpiOcular AD protocol. Based on results for 59 test materials, a prediction model (PM) was developed that uses a single exposure period and a single cut-off in tissue viability (determined by the MTT assay) for classification: ≤60% = irritant (I) (R36 and R41); >60% = non-classified (NC). Currently, we report results for an additional 35 materials (94 materials tested in total), including alcohols, hydrocarbons, amines, esters, and ketones, that were evaluated. For the 94 test materials, the PM for discriminating between ocular irritants and non-irritants resulted in 100.0% sensitivity and 68.0% specificity. This PM was subsequently evaluated in 2007/2008 by the European Cosmetics Association (COLIPA) in a multi-laboratory study. Twenty coded chemicals were tested in 7 laboratories (4 EU and 3 US). Overall, 298 independent trials were performed, demonstrating 99.7% agreement in prediction (NC/I) across the laboratories. Coefficients of variation for the % survival of tissues across laboratories were generally modest (<16%) except where tissue survival values were low. Using these data, a formal submission was sent to ECVAM in 2008 in support of the protocol's entry into a formal validation study. The expanded AD together with its long history of reproducibility and proven utility of EpiOcular for ultra-mildness testing make EpiOcular an extremely useful model for addressing current legislation related to animal use in the testing of potential ocular irritants.

Keywords: in vitro ocular tissue model, ocular irritation, validation, REACH

1 Introduction

Prediction of eye irritation potential is a requirement for the Registration, Labelling and Authorization of Chemicals (REACH) legislation, preparation of Material and Safety Data Sheets, labelling of cosmetic ingredients under the Cosmetics Directive, and labelling of mixtures and formulations where required by specific legislation (Eskes et al., 2005).

Damage to the corneal epithelium resulting from chemical or other environmental challenges may compromise tissue function, resulting in effects ranging from mild irritation to loss of ocular transparency and/or blindness. For the past twenty years, extensive research has been conducted to develop approaches for evaluating the eye irritation potential of cosmetic, household, and personal care products. This research involved developing a detailed understanding of the mechanism by which these products interact with eye tissue. Topical application of chemicals can cause cytotoxicity in several ways, including lysis of membranes (e.g. by surfactants, organic solvents), denaturation of proteins (e.g. by surfactants, organic solvents, alkalis and acids), saponifi-

cation of lipids (e.g. by alkalis), and alkylation or other covalent interactions with macromolecules (e.g. by bleaches, peroxides, alkylators) (Watt et al., 2004; Jester et al., 2001).

The Draize test performed on rabbits (Draize, 1944) has been the predominant test used for the assessment of ocular irritation/toxicity and for registering chemicals and cosmetics ingredients. The classification system takes into consideration the ocular effects produced, as well as the reversibility and the severity of the effects. A Draize modified maximum assessment score (MMAS) can range from 0 (non-irritants) to 110 (extreme irritants). In recent years, the Draize rabbit eye test has been increasingly criticised due to its lack of reproducibility, overestimation of human responses, cruelty to animals/animal rights issues (Balls et al., 1995; Curren and Harbell, 2002). The anatomy and biochemistry of the rabbit eye are not equivalent to that of the human eye (existence of the third eyelid, or nictitating membrane) and numerous physiological reasons (low tear production, blink frequency, ocular surface area, etc.) are also reasons why the rabbit eye often does not adequately predict human effects (Curren et al., 1997). These concerns have prompted many requests for refinement, reduction, and replacement of Draize ocular testing (Curren et al., 1997; Curren and Harbell, 1998).

The 7th amendment of the Cosmetics Directive (7th amendment, 2003) and REACH legislation (Regulation EC 1907/2006) prohibits or severely restricts the use of animals for safety assessment of chemical substances used in cosmetic products and in general use. For ocular irritation caused by surfactant and surfactant based materials, a prediction model has been developed (Blazka et al., 2003), and validation data are currently under review by ECVAM. The EpiOcular tissue model is also widely used to evaluate ultra-mild formulations that are designed for use in, or in the vicinity of, the eyes (McCain et al. 2002; Yin et al., 2009). The purpose of the present study was to further test a Prediction Model (PM) that expands the applicability domain of EpiOcular to a wider range of chemicals. The PM will allow industry to comply with the Cosmetics Directive and REACH legislation and address multiple concerns including: (a) consumer safety, (b) societal concern for animal welfare, and (c) testing cost.

Regulatory agencies and other users need to be reassured that the models will provide consistent, good quality data over time, not just during the validation process (Gupta et al., 2005). Recommended guidelines include "full characterization of cells or tissues, sampling of each lot … for performance, and regular use of controls and benchmark chemicals to provide assurance of consistency of assay performance" (Rispin et al., 2006). However, Draize testing has been criticised due to the lack of reproducibility in testing results both within and between laboratories. In order to ensure reproducible results, each batch of EpiOcular is tested against benchmark chemical treatment. Quality control of weekly batches of EpiOcular was performed using the MTT assay. The exposure time needed to reduce the viability to 50% (ET-50) for 0.3% Triton X-100 is determined. In order for the batch to be accepted, the ET-50 and negative control values must fall within ranges originally set in 1996.

This paper presents results: 1) Quality control testing results for EpiOcular showing long term stability and reproducibility of the tissue model, and 2) additional results expanded to 94 test articles for the new ocular irritation PM that broadens the applicability domain of the EpiOcular irritant/non-irritant test method.

2 Materials and methods

EpiOcular (OCL-200) reconstructed human ocular model

EpiOcular cultures (OCL-200) (Fig. 1A) are produced according to standardised procedures in the MatTek Corporation GMP tissue production facility. OCL-200 tissues are cultured using normal human epidermal keratinocytes (NHEK) obtained from

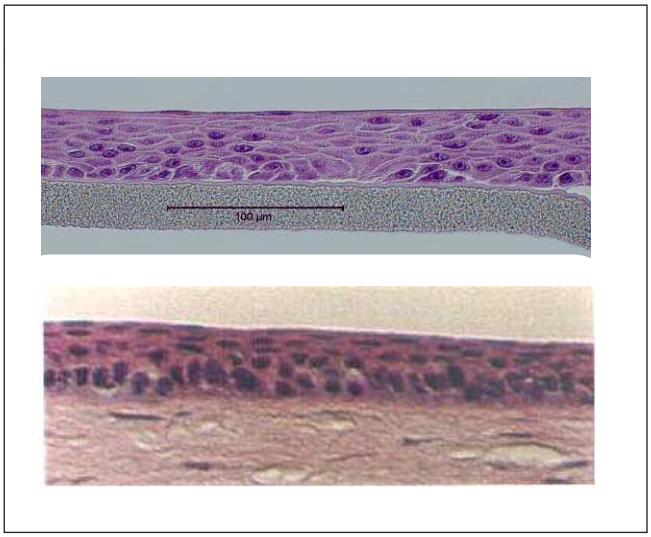


Fig. 1: H&E stained histological cross-sections of: A) EpiOcular™ (OCL-200) tissue model and B) rabbit cornea
Both tissues show highly organized basal cells, which progressively flatten out at the apical surface of the tissue.

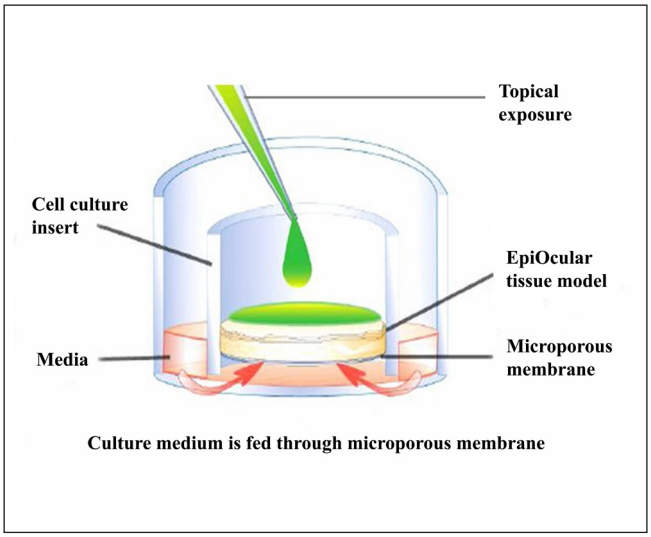


Fig. 2: A schematic representation of growth of the EpiOcular tissue at the air-liquid interface
The tissue is fed from the baso-lateral side, only allowing topical dosing onto the apical EpiOcular™ tissue surface, which otherwise is exposed to the atmosphere.

individual donors, and the cells from different donors are not pooled. All cell strains used for production are screened for HIV, Hepatitis-B, Hepatitis-C, mycoplasma, bacteria, yeast and fungi. The NHEK are seeded onto porous membrane inserts and cultured under submerged conditions until a confluent monolayer is formed. The inserts are then raised to the air/liquid interface (Fig. 2) and cultured in proprietary serum-free culture media to induce corneal differentiation and form the organotypic EpiOcular tissue model.

The EpiOcular tissue model exhibits *in vivo*-like morphological and growth characteristics, which are uniform and highly reproducible. EpiOcular consists of highly organised basal cells, which progressively flatten out as the apical surface of the tissue is approached, analogous to the normal *in vivo* corneal epithelium (Fig. 1) (Kubilus et al., 1997).

MTT tissue viability assay

The MTT assay is an inexpensive, robust, easy to use viability assay based upon the reduction of (3-(4,5-dimethylthiazole-2-yl)-2,5-diphenyl tetrazolium bromide) (MTT) by viable cells (Mosmann, 1983). The MTT assay measures metabolic function of mitochondria and its application to cytotoxicity assays has been well-established (Barile, 1998; Evans et al., 2001). The assay is routinely utilised with a variety of MatTek tissue models for both QC purposes as well as determination of tissue viability following experimental treatment. The MTT assay has been successfully utilised as the primary endpoint in validation and/or pre-validation programmes for ocular irritation (Blazka et al., 2003; McCain et al., 2002) as well as skin corrosivity, phototoxicity, and irritation (Fentem et al., 2001; Zuang et al., 2002; Kándárova et al., 2005).

Tissue viability is determined with the MatTek MTT viability assay kit (part# MTT-100) as follows. Approximately 1 hour prior to use, the MTT concentrate (5 mg/ml) is thawed and diluted (1:4), and 300 µl per well of the MTT solution is pipetted into a 24-well plate. EpiOcular tissue inserts are gently rinsed in PBS and placed in the MTT plate, making sure that no air bubbles are trapped underneath the cell culture insert. The tissues are then incubated at 37°C, 5% CO_2 for 3 hours. After incubation, each EpiOcular insert is removed from the plate and transferred into a 24-well plate containing 2.0 ml of the extractant solution. The plate is covered and sealed (to reduce evaporation) and extraction proceeds overnight at room temperature (RT) in the dark. After the extraction period is complete, the inserts are discarded, the extractant solution is mixed by pipetting up and down. 200 µl of the solution are pipetted into a 96-well microtiter plate in duplicate and the optical density of the samples is measured at 550 nm on a plate reader (Molecular Devices, Palo Alto, CA). % tissue viability is determined for each tissue using the following formula:

% Viability = [OD(treated tissues)/OD(negative control)] x 100%

Assessment of Direct Reduction of MTT by Test Article

Since the MTT assay is based on the reduction of the MTT formazan dye, it is known that materials which are strong reducers can interfere with the assay if they bind to the tissue (Fentem et al., 1998). This can result in false negative results (i.e. the tissue is actually non-viable but appears viable because MTT has been directly reduced by the residual test chemical that is not removed during rinsing).

To avoid false negative results, it is necessary to assess the ability of each test article to reduce MTT directly. 50 µl (liquid test articles) or one levelled spoonful (approximately 50 mg of solid test articles) are added to 0.5 ml of the MTT solution (1.0 mg/ml) and the mixture is incubated at standard culture conditions (37ºC, 5% CO_2) for two hours. A negative control (50 µl of deionised water) is run concurrently. If the MTT solution colour turns blue/purple, the test article has reduced the MTT.

In cases where the test article is shown to reduce MTT, a functional check using freeze-killed tissue controls is performed. Freeze-killed tissues are dosed, rinsed, and exposed to MTT according to the standard testing procedure. If the test material binds to the tissue, does not rinse off, and MTT reduction occurs, corrective measures must be used. In the current study, no materials caused MTT reduction and hence the concern related to false negatives was removed.

OCL-200 Quality Control

EpiOcular is produced under Good Manufacturing Procedures (GMP) to ensure tissue quality and reproducibility, both during the validation process and afterwards. Each OCL-200 tissue lot is evaluated by MatTek Corporation using established and certified QC procedures. Only tissue lots that meet the QC criteria are utilised. Random tissues from each weekly production batch are exposed to 100 µl of 0.3% Triton X-100 for exposure times of 5, 20, and 60 minutes (n=2 tissues/exposure time). In addition, tissues are exposed to 100 µl of ultrapure water for

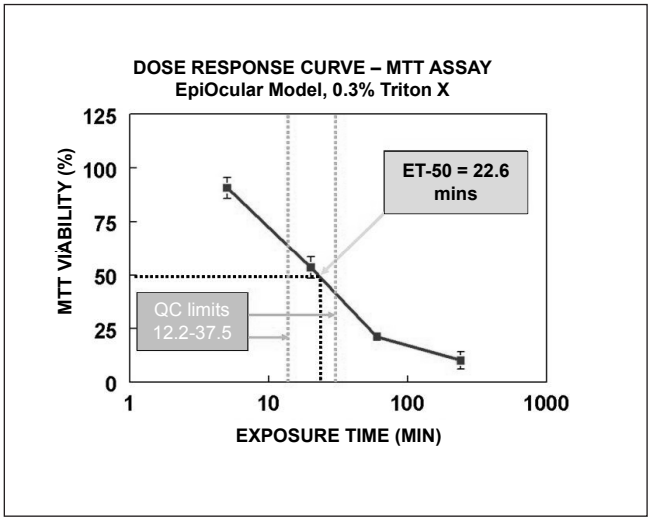


Fig. 3: ET-50 determination – graphical representation
N=2 tissues are exposed to 100 µl of 0.3% Triton X-100 for 5, 20, and 60 minutes to determine the ET-50 (negative control = ultrapure H_2O). Tissue viability is determined using the MTT assay. The exposure time at which the tissue viability is reduced to 50% (ET-50) is mathematically interpolated. Tissue to tissue variability within a lot can also be assessed by calculating the average for the coefficients of variation (AVG CV) for all points on the dose response curve.

60 minutes to serve as the negative control. Tissue viability is determined from the equation given above, and a dose response curve is constructed (Fig. 3). The exposure time that reduces the tissue viability to 50% (ET-50) was determined mathematically from: V = a + b * log(t), where V = % viability, t = time in h, and "a" and "b" are constants. The constants are determined using the viability data from 2 exposure times which bracket 50% viability, and then the equation is solved for t by setting V = 50%. A graphical representation of ET-50 determination is shown (Fig. 3). The ET-50 is an indirect measure of the tissue barrier properties, since the MTT monitors the viability of the basal cell layers of the tissue (Hines et al., 2003). Thus, a reproducible ET-50 indicates that the tissue barrier is constant, which is important since materials applied to the apical surface must penetrate to affect/interact with the viable cells within the tissue. Acceptance criteria for the EpiOcular tissue are given in Table 1.

Light microscopy evaluation: histological screening
Similar to *in vivo* changes observed clinically due to eye irritation, ocular irritants induce changes in tissue structure. In addition, materials which are strong reducers of MTT may lead to false negative results, and therefore another confirmatory assay is desirable. Histology can be used as a confirmatory assay, since it has proven to be an effective means of avoiding false negative results (Blazka et al., 2005; Kándárova et al., 2007).

For histological evaluation of control tissues, cultures were fixed with 10% formalin and embedded in paraffin; 5 μm cross-sections were cut, H&E stained, observed, and photographed using a Nikon Diaphot microscope. Histological sections are an accurate, facile, inexpensive screen for the general viability and architecture of the EpiOcular cultures.

Expanded Applicability Domain Protocol – Liquid Test Articles
Each test article and control was tested using duplicate tissues. After an overnight incubation at standard culture conditions (SCC, 37°C, 5% CO_2), the tissues were pre-treated with 20 μl of Calcium (Ca^{++}) and Magnesium (Mg^{++}) Free-DPBS (PBS). If the PBS did not spread across the tissues, the plate was tapped to insure that the entire tissue surface was wetted. Next, 50 μl of test article was applied topically onto each tissue (Fig. 2), and the tissues were incubated for 30 minutes under SCC. To prepare for rinsing the tissues, three 100 ml beakers were prepared with 100 ml each of PBS for each treatment group. After the incubation, each pair of tissues was successively rinsed by dipping, swirling, and decanting through its set of three beakers. After the final rinse and decanting, the tissues were immersed in 5 ml of EpiOcular Assay Medium (EAM) in a 12-well plate for 12 minutes (post-soak) at RT. The EAM was then decanted, and the tissues were transferred to a 6-well plate containing 1 ml of warm EAM. Tissues were incubated for 2 h at SCC (post-treatment incubation), and tissue viability was assessed using the MTT assay.

Expanded Applicability Domain Protocol – Solid Test Articles
Each test article and control was tested in duplicate tissues. After the overnight incubation, the tissues were pre-treated with 20 μl PBS, as above. Next, 50 mg of powder were applied topically onto the EpiOcular tissues using a calibrated scoop (~50 mg). The positive control was applied as above. The treated tissues were incubated for 2 hours at SCC. The rinsing and post-soak conditions were the same as described for the liquid articles; however, the solids-treated tissues were incubated for 18 hours in the post-treatment incubation. After the 18 hour post incubation period, tissue viability was determined using the MTT assay.

Protocol Transfer Training
Transferability of the expanded AD protocol was assessed in a multi-laboratory, COLIPA-sponsored study involving three US and four European laboratories (Harbell et al., 2009). Liquid test materials were chosen for which *in vivo* Draize scores were available in the literature (van Goethem et al., 2006); in addition, *in vivo* data for solid test articles were obtained from ECETOC and TSCA databases, as well as GHS BCOP Substances Chart Classes. Three laboratories tested a set of twenty liquid test articles; the four other laboratories tested 15 liquids and 5 solid test articles. All chemicals were coded by an independent laboratory and distributed by MatTek Corporation.

The laboratories received formal hands-on training from MatTek staff using the common protocol and laboratory documentation. Both experienced and naïve laboratories were included.

Tab. 1: Quality control (QC) acceptance criteria for EpiOcular (OCL-200) tissue.

ET-50 (minutes)	12.2 > ET-50 < 37.5
Negative control (OD)	> 1.00
Histological evaluation	Stratified non-keratinising squamous epithelium, 4-6 cell layers thick
Sterility	No signs of contamination following storage for 14 days
Notes:	
All tests are performed on randomly selected tissues from weekly production batches.	
ET-50: Exposure time in which 0.3% Triton X-100 decreases tissue viability to 50% as determined by MTT assay.	
Negative control (NC): NC tissues are exposed to ultrapure H2O. Optical density (OD) of the extract from the MTT assay is determined at 550 nm.	
Histological evaluation: H&E stained tissue cross-sections are evaluated by EpiOcular product manager (Ph.D. scientist) if ET-50 and/or NC data are suspect.	
Sterility: Tissues are cultured for 2 weeks after production for signs of bacteria or fungal contamination.	

Each laboratory tested five chemicals, and the data were submitted to MatTek for review. Upon successful completion of this review step, testing on the coded chemicals was begun. MatTek was not involved in the study after this initial training was completed.

3 Results

Quality Control Results

The yearly average ET-50s for batches of EpiOcular produced during calendar years 1996-2008 are presented in Table 2. As shown, the yearly average ET-50s have ranged from 22.0 (1999) to 27.3 minutes (2006) but have remained clustered around the QC acceptance value of 24.9 minutes established in 1996. Thus, tissue properties have remained constant over the past 13 years. In addition, the variability between tissues has been low as evidenced by the yearly average coefficients of variation (AVG CV) which have not exceeded 7%.

Expanded applicability domain protocol

A total of 94 test materials (74 liquids/20 solids) including alcohols, hydrocarbons, amines, esters, and ketones were evaluated using the EpiOcular assay (Tab. 3). The selection of chemicals

Tab. 2A: Summary of yearly average for quality control (QC) testing of OCL-200 tissue.
N=2 tissues were exposed to 100 µl of 0.3% Triton X-100 for 5, 20, and 60 minutes to determine the ET-50 (negative control = ultrapure H_2O). Average ET-50 for calendar year and the average coefficient of variation for the dose response curves is shown.

calendar year	ET-50 (min)	Avg CV (%)
1997	22.9	5.5
1998	25.2	5.5
1999	22.0	5.6
2000	22.9	5.5
2001	23.3	4.7
2002	22.4	5.0
2003	24.6	5.9
2004	22.2	6.7
2005	24.8	5.8
2006	27.3	6.4
2007	24.4	5.7
2008	25.0	4.9

Tab. 2B: Original Quality Control (QC).
QC parameters were established in 1996 based on 184 lots of EpiOcular tissue.

1996 Avg. ET-50 (min)	24.9 (N=184)
Std. Dev. (min)	6.3
QC upper limit (min)	37.5
QC lower limit (min)	12.2

Conclusion: Since 1996, the tissue properties (ET-50) have remained constant and the tissue to tissue reproducibility has been high (low AVG CV).

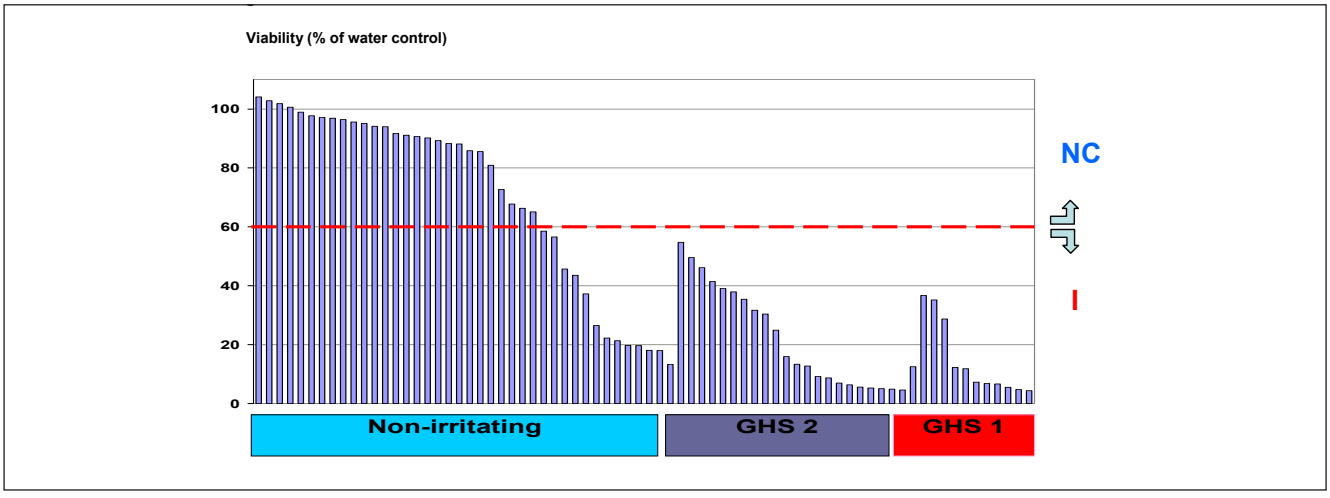


Fig. 4: Graphical representation of the *in vitro* and *in vivo* results for liquid test articles using the EpiOcular irritation protocol (data from Tab. 3A)
In vitro EpiOcular tissue viability is plotted versus the *in vivo* GHS ocular irritation category. Each liquid compound is represented by a single bar. Irritant/non-irritant prediction model cut-off value of 60% of water viability is represented by the dashed red line.
Conclusion: All GHS 2 and GHS 1 irritants were corrected predicted (tissue viability < 60%) as irritants.

Tab. 3: List of coded chemicals tested in expanded applicability EpiOcular irritation assay

A. Liquids

Chemicals	# of exp.	MMAS	GHS	60% cut off	Viability (% of water)	StDev
Alcohols						
1	2	80	1	I	12.4	9.7
2	2	65	2	I	6.9	1.0
3	2	61	2a	I	4.5	1.1
4	5	51	2a	I	37.9	9.8
5	3	41	2	I	30.4	15.6
6	2	34	1	I	4.3	0.3
7	3	34	1	I	28.6	23.1
8	2	31	1/2	I	9.1	2.9
9	2	25	2b	I	13.3	9.9
10	2	24	2	I	12.7	5.1
11	2	22	2b	I	4.8	0.3
12	2	13	2b	I	8.7	3.8
13	2	8	NI	I*	17.9	1.3
14	1	3	NI	NI	95.1	1.0
15	3	2	NI	NI	100.6	2.9
16	1	1	NI	NI	88.1	10.6
17	1	1	NI	NI	96.4	5.6
18	2	0	NI	NI	96.9	0.7
19	3	0	NI	NI	93.9	3.2
Alkali						
20	2	26	2	I	24.9	26.6
Amine						
21	3	89	1	I	5.4	2.1
Esters						
22	3	45	1	I	12.2	5.4
23	2	45	1	I	11.8	4.0
24	3	40	2a	I	35.4	1.3
25	3	28	2a	I	39.0	1.7
26	3	18	2b	I	46.1	2.8
27	3	15	NI	I*	13.2	6.5
28	3	8	NI	I*	26.5	15.0
29	3	5	NI	NI	98.9	2.0
30	3	4	NI	I*	37.2	12.5
31	3	3	NI	I*	19.6	8.8
Ether						
32	3	31	1	I	6.6	0.9
Hydrocarbons						
33	2	10	NI	NI	66.2	1.6
34	3	5	NI	NI	91.7	2.9
35	3	4	NI	NI	91.1	7.0
36	3	3	NI	I*	56.5	10.7
37	4	3	NI	NI	97.7	3.2
38	3	2	NI	NI	101.8	3.4
39	3	1	NI	NI	90.6	4.3
40	2	0	NI	NI	95.5	4.2
41	2	0	NI	NI	89.2	1.6

Chemicals	# of exp.	MMAS	GHS	60% cut off	Viability (% of water)	StDev
Hydrocarbons (cyclic)						
42	2	70	1/2	I	6.3	1.2
43	1	50	1/2	I	54.6	3.9
44	3	49	1	I	36.7	18.3
45	3	24	2a	I	41.4	3.5
46	3	9	NI	I*	45.6	15.3
47	3	7	NI	I*	58.5	3.3
48	4	5	NI	NI	67.7	4.9
49	2	4	NI	NI	88.2	3.2
50	2	2	NI	NI	65.0	0.0
51	2	2	NI	NI	85.8	4.6
Ketones						
52	2	66	2a	I	15.9	5.4
53	2	50	2a	I	5.0	1.1
54	3	13	NI	I*	22.2	10.7
55	3	7	NI	NI	72.6	7.7
56	2	5	NI	I*	19.6	5.6
Nitrile (cyanide)						
57	2	14	2b	I	31.6	0.7
Organic acids						
58	2	38	1	I	4.7	1.1
59	2	38	2a	I	5.5	1.4
PABA derivative						
60	2	0	NI	NI	97.1	2.9
Surfactants						
61	2	108	1	I	7.2	0.4
62	2	96	1	I	6.7	1.1
63	2	86	1	I	35.1	30.0
64	1	36	2a	I	49.6	47.3
65	2	32	2a/2b	I	5.2	0.5
66	1	5	NI	NI	85.5	5.5
67	2	4	NI	NI	102.8	4.2
68	2	3	NI	NI	80.9	11.8
69	2	2	NI	I*	43.5	0.2
70	2	2	NI	I*	18.0	6.8
71	2	1	NI	I*	21.3	0.9
72	3	0	NI	NI	94.1	2.1
73	1	0	NI	NI	90.1	5.1
74	1	0	NI	NI	104.1	5.2

Notes: I, overpredicted non-irritating chemicals*

B. Solids

Chemicals	# of exp.	MMAS	GHS	60% cut off	Viability (% of water)	StDev
Alkali						
1	2		NI	NI	92.9	14.1
Amine						
2	2	82	1	I	4.7	1.7
3	3	14.3	2b	I	37.4	2.5
4	3	0	NI	NI	105.7	9.0
5	2	-	NI	NI	94.3	6.9
6	3	-	NI	I *	31.0	23.8
7	2	-	NI	NI	97.1	6.2
Esters						
8	2	-	2a	I	5.8	0.2
9	3	-	NI	I *	5.0	0.2
Hydrocarbons(cyclic)						
10	2	82.3	1	I	16.9	2.4
11	3	15	2b	I	42.5	20.0
12	2	-	NI	NI	102.6	9.4

Chemicals	# of exp.	MMAS	GHS	60% cut off	Viability (% of water)	StDev
Organic acids						
13	2	33.7	1	I	27.8	7.0
14	2	-	NI	NI	71.8	4.3
Organic salt						
15	2	71.7	1	I	10.6	3.6
16	2	15.7	2b	I	16.7	11.1
17	2	-	1	I	14.9	6.1
Inorganic salt						
18	3	0	NI	NI	81.8	11.8
Amidine						
19	3	6.7	NI	I *	29.6	15.6
Nitro compound						
20	3	15.7	2b	I	22.4	11.2

was based on the availability of publicly available *in vivo* (rabbit) eye irritation data and the desire to have a broad range of chemical classes tested.

Negative control tissues were dosed with 50 µl of ultrapure water, and tissue viability was evaluated using the MTT assay as previously described. When the viability was ³60%, the test chemical was considered non-classified (NC); if tissue viability was <60%, the test chemical was predicted to be and irritant (I).

All *in vivo* and *in vitro* data are presented in Table 3. Data for liquid test materials are summarised in Figure 4; data for solid test materials are given in Figure 5. For liquid test materials, only 1 GHS 1 category irritant and 1 GHS 2 category irritant were under-predicted by the EpiOcular assay. For solid test materials, there were no under-predictions. As summarised in Table 4, the overall sensitivity of the prediction model was 95.7%; specificity for the prediction model was 68.0%

Protocol Transferability

The reproducibility within each of the 7 laboratories involved in the COLIPA-sponsored pre-validation study was assessed by calculating the tissue viability and coefficient of variation. Coefficients of variation for the % survival of tissues across laboratories was generally modest (<16%) except where tissue survival values were low (Harbell et al., 2009). In addition, there was good agreement between experiments within all laboratories.

For evaluation of reproducibility between laboratories, the average tissue viability for the 25 test chemicals for all labs was calculated. Fifteen of the liquid test materials were tested in all seven laboratories; an additional 5 liquids were tested in three laboratories and five solids were tested in four laboratories. As shown in Figure 6, the variability between laboratories, as shown by the error bars, is minimal. Remarkably, a total of 298 independent trials across seven laboratories were performed during the study. Of the 298 trials, 297 of the trials (99.7%) were in agreement for prediction of irritant/non-irritant status (Harbell et al., 2009).

4 Discussion

An important characteristic of any toxicology system is its long term stability and reproducibility. Clearly, if EpiOcular tissue properties changed over time, the utility of the model would be dramatically reduced and, at some point, the prediction model would become invalid. The long term reproducibility of the tissue model properties, as evidenced by a stable ET-50 (Tab. 2), means that the prediction model will perform in a constant manner both during the validation process and in subsequent years following validation. The high level of reproducibility of the EpiOcular model is undoubtedly due to the strict good manufacturing practice (GMP) techniques utilised for the production of EpiOcular tissues.

The purpose of this study was to evaluate a new EpiOcular prediction model to expand the applicability domain for ocular irritation testing. The new prediction model is meant to address legislation now in place due to the 7th amendment of the Cos-

Tab. 4: Results for the expanded applicability domain EpiOcular irritation prediction model.

	Liquids	Solids	Total
Sensitivity	100.0	100.0	100.0
Specificity	67.5	70.0	68.0
Accuracy	82.4	85.0	83.0
Positive predictivity	72.3	76.9	73.3
Negative predictivity	100.0	100.0	100.0
False positives	32.5	30.0	32.0
False negatives	0.0	0.0	0.0

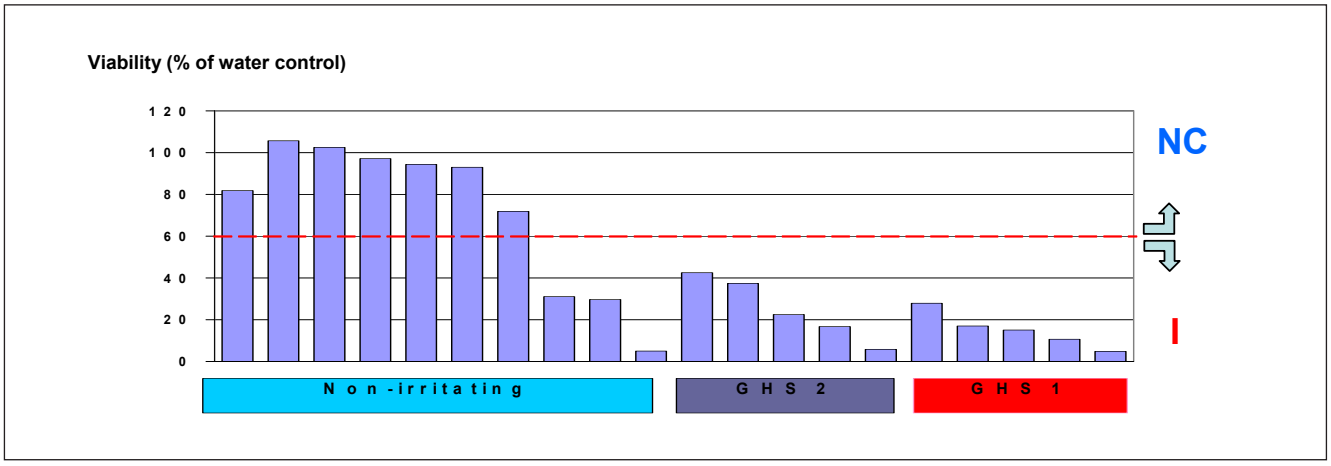


Fig. 5: Graphical representation of *in vitro* and *in vivo* results for solid test articles using the EpiOcular irritation protocol (data from Table 3B)

In vitro EpiOcular tissue viability is plotted versus the *in vivo* GHS ocular irritation category. Each solid compound is represented by a single bar. Irritant/non-irritant prediction model cut-off value of 60% of water viability is represented by the dashed red line. *Conclusion:* All GHS 2 and GHS 1 irritants were corrected predicted (tissue viability < 60%) as irritants.

metic Directive, which banned the marketing in the EU (as of March 11, 2009) of any cosmetic product that had been tested using animals. Likewise, the new prediction model will address REACH legislation now in place. Of note is that the new prediction utilises the MTT assay, which has been used previously in validated skin corrosion and skin irritation assays. Also, the new prediction model relies on a single assay time point, which simplifies testing and reduces cost. Due to the high level of tissue to tissue reproducibility, with a single OCL-200 kit of 24 tissues, the irritancy of 10 test materials (along with the positive and negative assay controls) can be evaluated.

The cut-off of 60% tissue viability was chosen in order to maximise the sensitivity of the assay without significantly compromising the specificity of the assay (sensitivity = 100.0%, specificity = 68.0%, see Tab. 4). In this way, the prediction model favours predictions that may err on the side of safety. However, if the viability cut-off were set at 50%, the specificity would rise to 72% while the sensitivity would still be 95.5%; or if the cut-off were set at 40%, the sensitivity would be 90.7% and specificity 76.0%. Therefore, it is possible to alter the prediction model to obtain a more balanced prediction while still maintaining high sensitivity and specificity.

Results for this assay previously reported included two false negatives. However, upon examination of the data, these chemicals were identified as decanol and dibenzoyl-l-tartaric acid (20%). Decanol is known to be an *in vivo* skin irritant (rabbits) and an *in vitro* (EpiDerm) skin irritant. When this material was originally tested, it was assumed that this chemical would be an ocular irritant as well. However, in fact not all skin irritants are ocular irritants (Kennedy and Banerjee, 2006). Regarding dibenzoyl-l-tartaric acid (20%), *in vivo* MMAS data are only available for the neat material. In fact, it is possible that at 20% the chemical is non-irritating. Therefore, in both cases, data for these two test materials were eliminated from the current manuscript.

References

7th amendment of the Cosmetics directive, Directive 2003/15/EC.

Balls, M. et al. (1995). The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicol. In Vitro 9*, 871-929.

Barile, F. A. (1998). Cardona M. Acute cytotoxicity testing with cultured human lung and dermal cell. *In Vitro Cell Dev. Biol. Anim. 34(8)*, 631-635.

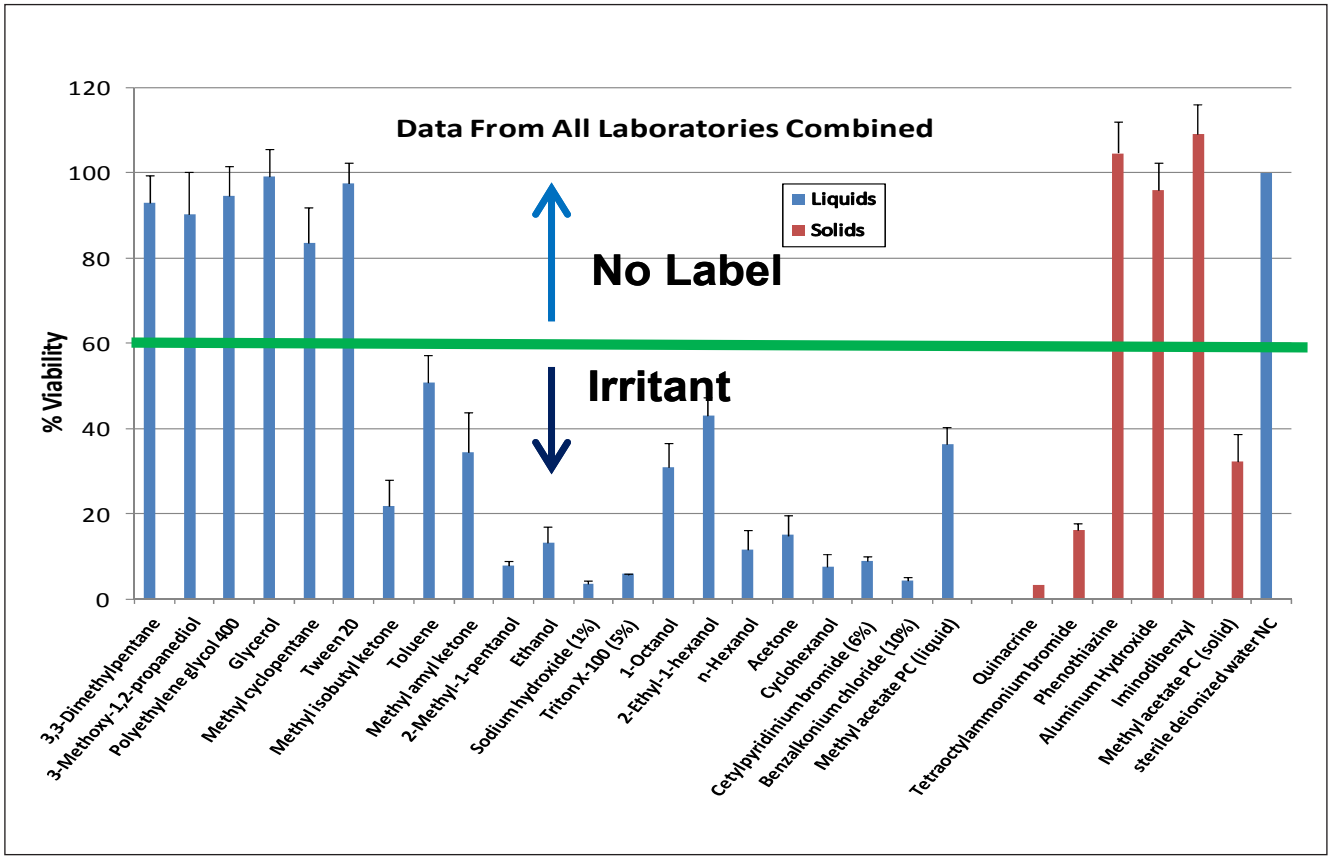


Fig. 6: Average viability for the 25 test articles and the positive control tested in the 7 laboratories, COLIPA-sponsored EpiOcular irritation pre-validation study (Harbell et al., 2009)

Blazka, M., Diaco, M., Harbell, J. et al. (2005). EpiOcularTM human cell construct: tissue viability and histological changes following exposure to surfactants. *The Toxicologist 84 (1),* 409.

Blazka, M., Harbell, J., Klausner, M. et al. (2003). Evaluating the ocular irritation potential of 54 test articles using the EpiOcular human tissue construct model (OCL-200). *The Toxicologist 72, S1*, 221.

Curren, R. D. and Harbell, J. W. (2002). Ocular safety: a silent (in vitro) success story. *ATLA 30, Suppl. 2*, 69-74.

Curren, R. D. and Harbell, J. W. (1998). In vitro alternatives for ocular irritation. *Environ Health Perspect. 106, Suppl. 2*, 485-492.

Curren, R. D. et al. (1997). Current approaches to the in vitro prediction of ocular irritation Comments in *Toxicol. 6*, 71.

Draize et al. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Pharmacol. Exp Ther. 82*, 377-390.

Eskes, C. et al. (2005). Eye irritation. *ATLA 33, Suppl. 1*, 47-81.

Evans, S. M., Casartelli A. et al. (2001). Development of a high throughput in vitro toxicity screen predictive of high acute in vivo toxic potential. *Toxicol In Vitro 15(4-5),* 579-584.

Fentem, J. H., Briggs, D. et al. (2001). A pre-validation study on in vitro tests for acute skin irritation: results and evaluation by the Management Team. *Toxicology in Vitro 15*, 57-93.

Fentem, J. H. et al. (1998). The ECVAM international validation study on in Vitro tests for skin corrosivity – 2. Results and evaluation by the management team, *Toxicol. In Vitro 12*, 483-524.

Gupta, K., Rispin, A., Stitzel, K. et al. (2005). Ensuring quality of in vitro alternative test methods: issues and answers. *Regul. Toxicol. Pharmacol. 43(3),* 219-224.

Harbell, J. W. et al. (2009). COLIPA's industry pre-validation program using reconstructed human tissue-based methods for predicting eye irritation for chemicals: MatTek EpiOcular. *ALTEX 26, Spec. Issue*, 289.

Hines, M. D., Gettings, S. D. and Jones, B. C. (2003). Evaluation of MTT metabolism as a meaningful indicator of viability in human corneal epithelial tissue models. *The Toxicologist 72*, S1, 156.

Jester, J.V., Li, L., Molai, A., Maurer, J.K., (2001). Extent of initial corneal injury as a basis for alternative eye irritation tests. *Toxicology In Vitro 15*, 115-130.

Kandárová H. et al. (2007). In Vitro Skin Corrosion Test: Long Term Reproducibility and Reliability for a Regulatory Accepted Method. *Toxicology Letters 172, Suppl. 1*, S82.

Kandárová, H., Liebsch, M., Gerner, I. et al. (2005). EpiDerm skin irritation test protocol – Assessment of the performance of the optimized test. *ATLA 33 (4),* 351-367.

Kennedy, G. L. Jr. and Banerjee, A. K. (2006). Relationship between eye and skin irritation in rabbits using a series of textile fiber finishes. *Journal of Applied Toxicology 12(4),* 281-284.

Kubilus, J. et al. (1997). EpiOcular prediction model. a reproducible in vitro tissue culture means of predicting Draize scores, M. MatTek Corp., Ashland, MA. *The Toxicologist 36 (1), 44*, Abstract #222.

McCain, N. E., Binetti, R. R., Gettings, S. D. and Jones, B. C. (2002). Assessment of ocular irritation ranges of market-leading cosmetic and personal-care products using an in vitro tissue equivalent. *The Toxicologist 66 (1-S),* 243.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays, *J. Immunol. Meth. 65*, 55.

Regulation EC 1907/2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals.

Rispin, A., Stitzel, K., Harbell, J. W. and Klausner, M. (2006). Ensuring quality of in vitro alternative test methods: Current practice. *Regul Toxicol Pharmacol. 45(2),* 97-103.

van Goethem, F. et al. (2006). Pre-validation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. *Toxicol In Vitro 20*, 1-17.

Watt, B. E. et al. (2004). Hydrogen peroxide poisoning. *Toxicol. Rev. 23(1),* 51-57.

Yin, X. J., Dang, A. H., Gill, A. et al. (2009). Prediction of ocular irritation potential of surfactants-based formulations at different concentrations using the EpiOcularTM model. *The Toxicologist 108 (1),* 378.

Zuang, V., Balls, M. et al. (2002). Follow-up to the ECVAM revalidation study on in vitro tests for acute skin irritation. ECVAM Skin Irritation Task Force Report 2. *ATLA 30*, 109-129.

Acknowledgements

The authors gratefully acknowledge the outstanding efforts of the testing labs and individuals listed below who contributed to the generation of the COLIPA pre-validation study results reported herein:

Pauline McNamee (The Procter & Gamble Company), John Harbell (Mary Kay Cosmetics), Claudius Griesinger (ECVAM), Teena Jacob (Avon), Uwe Pfannenbecker (Beiersdorf), Claudine Faller (Cosmital P&G), Jaqueline Rolle (Cosmital P&G), Greg Mun (IIVS), Hans Raabe (IIVS), Magalie Tailhardat (LVMH), Tiffany Florence (Mary Kay Cosmetics), Brigida McAnulty (Mary Kay Cosmetics), Sandrine Bessou-Touya (Pierre Fabre), and Marie Templier (Pierre Fabre).

Correspondence to

Mitchell Klausner, M.S.
Vice President Scientific Affairs
MatTek Corporation
200 Homer Ave
Ashland, MA 01721
USA
e-mail: mklausner@mattek.com

Prospective Validation Study of Reconstructed Human Tissue Models for Eye Irritation Testing

Stuart J. Freeman[1], Nathalie Alépée[2], João Barroso[3], Thomas Cole[3], Anna Compagnoni[3], Carina Rubingh[4], Chantra Eskes[5], Jan Lammers[4], Pauline McNamee[6], Uwe Pfannenbecker[7] and Valérie Zuang[3]

[1]Farino Consulting, UK, chair of the Validation Management Group (VMG); [2]L'Oréal, France; [3]ECVAM, Institute for Health & Consumer Protection, European Commission Joint Research Centre, Italy; [4]TNO, The Netherlands; [5]Independent consultant, Italy; [6]COLIPA, Belgium; [7]Beiersdorf, Germany

Summary

A prospective validation study of two in vitro test methods using Reconstructed human Tissue (RhT) models (MatTek EpiOcular™ and SkinEthic™ Human Reconstructed Corneal Epithelium (HCE)) for the detection of eye irritation effects by chemicals is currently being conducted by Colipa and ECVAM. Pre-validation studies with both test methods have served to optimise protocols and refine prediction models and have been shown to predict eye irritant properties of test substances with a high degree of accuracy, approximately 80% overall. The current validation study, managed by ECVAM and co-sponsored by Colipa and ECVAM, will evaluate the eye irritant properties of 104 test substances, identifying them as either not classified for eye irritancy (NC) or irritant (Cat. 1 and Cat. 2) within the United Nations Globally Harmonised System (UN GHS). The test methods will not differentiate between moderate (GHS Cat. 2) and severe (GHS Cat. 1) irritants but will seek to establish the relevance and reliability of the two test methods in an integrated testing strategy that will allow appropriate labelling of novel chemicals without the need for animal testing.

Keywords: in vitro alternatives, eye irritation, validation, EpiOcular, SkinEthic HCE

1 Introduction

In the field of eye irritation testing, numerous non-animal test methods have been developed over the years. Many of these have been subject to validation and/or evaluation studies (reviewed in Eskes et al., 2005), with the result that two such tests, the Bovine Corneal Opacity and Permeability (BCOP) test and the Isolated Chicken Eye (ICE) test, have been recommended for OECD approval as alternatives to the Draize rabbit eye test for the detection of chemicals that are corrosive or severely irritant to the eye (OECD, 2008a,b). Recently, two additional assays, Cytosensor Microphysiometer and Fluorescein Leakage have also been validated to identify corrosives and severe irritants to the eye. Furthermore, for the first time, two *in vitro* test methods, the Cytosensor Microphysiometer and the BCOP, have been validated to distinguish non-irritant (NC) from irritant (GHS Cat. 1 or 2) test substances (ESAC, 2009; ICCVAM, 2009a,b). However, due to their limited applicability domain, the assays do not cover the full range of physico-chemical properties for which regulatory testing is required. There are still no validated *in vitro* tests adopted in law to distinguish non-irritant (NC) from irritant (GHS Cat. 1 and 2) substances.

The replacement of animal tests to determine the toxicity of chemicals has long been a goal of toxicological research. Within the European Union, the Seventh Amendment to the Cosmetics Directive expressly forbids the use of animals in the safety evaluation of cosmetic products and ingredients. There is, therefore, an urgent need to continue to develop and to validate non-animal alternatives to established animal tests, including those for eye irritancy. Additional impetus has been given to the search for validated non-animal alternatives with the introduction of the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) legislation, which requires the undertaking of significant safety testing to support the use of new and existing substances.

The mechanisms of eye irritation studied *in vitro* can be modelled by the depth of injury theory described by Maurer and Jester (Jester et al., 2001; Maurer et al., 2002; Jester, 2006). Using confocal microscopy and traditional histology, these scientists observed that initial depth of injury to the cornea was highly predictive of the overall degree and duration of eye irritation. Depth of injury could be broken down by the tissue layers involved. Slight to mild irritation involved largely the corneal epithelium. Moderate irritants produced damage through the corneal epithelium and into the upper stroma. Damage into the deeper stroma (including damage to the endothelial cells) was predictive of severe irritation.

This paper describes the progress made to date in the validation of two test methods based on RhT models, MatTek EpiOcular™ and SkinEthic™ HCE to distinguish chemicals that

are irritant (GHS Cat. 1 and Cat. 2) to the eye from those that are non-irritant (NC). Irritants and non-irritants are defined according to the OECD and United Nations proposal for a Globally Harmonised System (GHS) for the classification and labelling of ocular irritants (category 1/category 2/no category; Anon, 2003) and as implemented in the European Commission Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No. 1907/2006 (EC, 2008). Neither test method is designed to distinguish between chemicals that are moderate (GHS Cat. 2) and severe (GHS Cat. 1) irritants. It is anticipated that successfully validated methods will be incorporated into a testing strategy (Scott et al., 2009) that will facilitate appropriate labelling of novel chemicals, ideally without the need for animal testing.

The European Centre for the Validation of Alternative Methods (ECVAM) has published the principles of a so-called modular approach to the validation and acceptance of test methods (Hartung et al., 2004). The current study is being conducted in accordance with these principles and with the guidelines published by OECD (2005). Thus, in a pre-validation study test definition, within laboratory variability (WLV) and transferability of the test method have been demonstrated for these methods and preliminary information obtained for between laboratory variability (BLV) and predictive capacity. This has led to some refinement of the protocol and prediction model for the SkinEthic™ HCE test method. Further information on the relevance and reliability of the optimised protocols, including BLV, Predictive Capacity and Applicability Domain will be obtained in the current full-scale validation study with the defined set of 104 chemicals.

2 Test Methods

2.1 EpiOcular™

The EpiOcular™ model uses normal human epidermal keratinocytes cultured to form a stratified squamous epithelium (Sheasgreen et al., 1996).

Use of the EpiOcular™ RhT model in the study of eye irritancy has been established for some years. The utility of the model in determining the degree of eye irritant potential of surfactants has been demonstrated (Blazka et al., 2000, 2003, 2005). These studies used a time-to-toxicity protocol (Ghassemi et al., 1997) that measured the time at which 50% of cultured cells (ET_{50}) remained viable, relative to negative controls. In this way, the degree of irritant effect, in the mild to moderate range, was determined. The protocol adopted for pre-validation and the present validation study differs in that it uses a single exposure time for each chemical tested. An irritant effect is inferred if cell viability falls below a pre-determined level (≤60%), relative to negative controls, as measured by succinate dehydrogenase reduction of MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide). The test protocol and prediction model were developed by MatTek Corporation using a total of 76 chemicals from across a range of chemical classes (Kaluzhny et al., 2007).

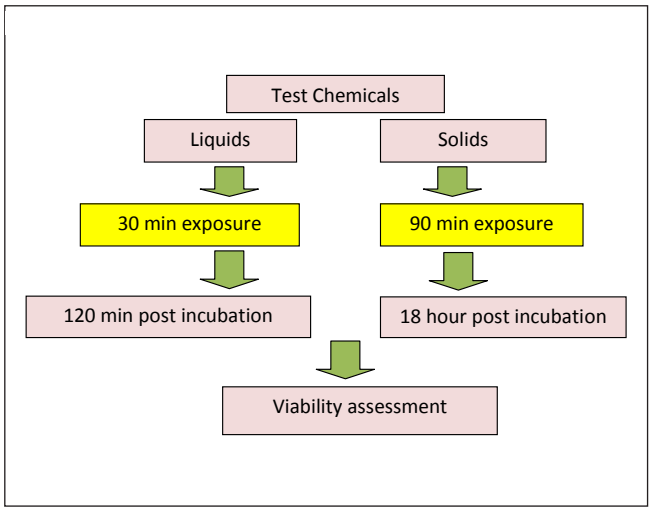


Fig. 1: Testing strategy for MatTek EpiOcular™

Separate protocols are employed for liquids and solids. Tissues are exposed to liquids for 30 minutes followed by a 120-minute post-treatment incubation and to solids for 90 minutes followed by an 18-hour post treatment incubation (Fig. 1).

The EpiOcular test method showed very good reproducibility (100%) in the prediction of irritant and non-irritant substances within and between laboratories. Furthermore, the predictive capacity of the test method for liquids and solids combined has been shown to give an overall accuracy of 81.3% under the EU classification system (NC versus R36/R41; Kaluzhny et al., 2007). Under GHS the overall accuracy of the test method improved to 83%, with a sensitivity of 100% and a specificity of 68%.

2.2 SkinEthic™ HCE

The SkinEthic™ HCE model uses immortalised human corneal cells which, when cultured in defined conditions, develop into a multi-layered tissue which resembles the human corneal epithelium morphologically and physiologically (Nguyen et al., 2003). The protocol for testing chemicals for their eye irritant potential was originally developed with 20 test substances from that published by van Goethem et al. (2006) and, like the EpiOcular™ method, uses cell viability measurement based on MTT reduction as its endpoint. Cell viability ≤50% would result in a chemical being classified as irritant; cell viability >50% would result in a classification of non-irritant. Evaluation of the protocol using a second set of about 100 test substances demonstrated that an increase in specificity was achieved by severe loss in sensitivity prediction (unpublished data), leading to development of an optimized SkinEthicTM HCE protocol including two exposure times. The short exposure time consists of a 10-minute exposure of tissue to the test substance with no post-treatment incubation, while the long exposure treatment exposes the tissue to the test substance for 1 hour with a further post-treatment incubation of 16 hours. In a pre-validation study involving at least 3 different laboratories, the SkinEthic™ HCE

test method consistently predicted (>95%) irritant and non-irritant test substances within (WLV) and between (BLV) laboratories. By combining the two exposure times in a paradigm that uses suitable selection criteria to allocate test chemicals to one or other treatment time, the overall accuracy was shown to increase to nearly 80%, with a sensitivity of 86.7% and a specificity of 68.9% (under GHS, reviewed by ECVAM). The criterion for allocation of test substances to either short or long exposure times is based on their intrinsic chemical reactivity, as defined by their electrophilic potential to react with cysteine- or lysine-containing peptides and measured through a direct peptide reactivity assay (DPRA; Gerberick et al., 2007). Chemical reactivity with the test substance is reported as percent depletion of the nucleophile, which is determined as the reduction of the peptide concentration in the samples relative to the average concentration of the controls. If the percent peptide depletion relative to the control is >5.95%, the test substance is categorized as reactive. If the percent peptide depletion is ≤5.95%, the substance is categorized as non-reactive. Thus substances which demonstrate an ability to bind in significant amounts to a cysteine- or lysine-containing peptide are deemed to be reactive (Gerberick et al., 2007), and are allocated to the short exposure (10 min) time treatment, while those substances which do not show significant binding to cysteine and lysine peptides and are considered non-reactive are allocated to the long exposure (1 h + 16 h post-treatment incubation) time treatment (Fig. 2). The validity of the dual protocol testing strategy will be determined in the post-study analysis of data.

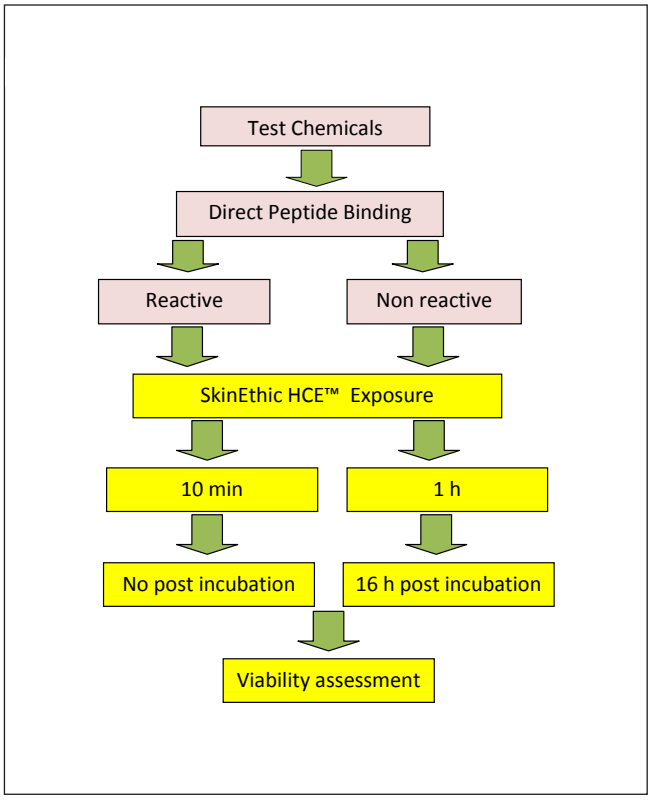


Fig. 2: Testing strategy for SkinEthic™ HCE

3 Study Design

3.1 Bio-analytical Assays

Chemical reactivity (peptide binding) is defined in this validation study as the electrophilic potential to react with cysteine- or lysine-containing peptides. In a first stage, test substances will have their chemical reactivity determined by the cysteine and lysine DPRA test method in a blind study in a single laboratory. Chemical reactivity for a subset of test substances will be conducted in two further laboratories following viability assessments to confirm reproducibility of the assay.

3.2 Biological Assays

Viability assessments for both the EpiOcular™ (liquids and solids exposure times) and SkinEthic™ HCE (short and long exposure times) protocols will be carried out in 3 different laboratories for each protocol. For the purposes of this validation study, all test substances will be evaluated in both exposure times of the SkinEthic™ HCE test method by all participating laboratories. Test substances will be coded, packaged and distributed by an independent co-ordinating organisation to each of the participating laboratories. Coding will be provided by an independent statistician at the co-ordinating organisation. It is anticipated that all test substances will be assessed in two (EpiOcular™) or three (SkinEthic HCE™) replicates in three separate runs in all laboratories.

Statistical analysis on data obtained during prevalidation of the EpiOcularTM test method showed that the use of only two tissue replicates will be sufficient for the validation study, since >90% of concurrently treated tissue replicates during prevalidation had a difference of viability below 5%.

4 Chemical Selection

Chemical selection for the study is the responsibility of the Chemical Selection Group (CSG) (see Fig. 3). The sample size of 104 substances has been statistically determined for the study.

Core VMG	Chair	Stuart John Freeman, Farino Consulting
	Co-chair	Valérie Zuang, ECVAM
	ECVAM sponsor	João Barroso, ECVAM
	COLIPA sponsor	Pauline McNamee, P&G
	Coordinating organisation	Jan Lammers, TNO
	Biostatisticians	Carina de Jong-Rubingh, TNO
		Anna Compagnoni, ECVAM
	External scientist	Chantra Eskes, ECVAM
	Chair of CSG	Thomas Cole, ECVAM
	SkinEthic™ HCE lead lab	Nathalie Alépée, L'Oréal
	EpiOcular™ lead lab	Uwe Pfannenbecker, Beiersdorf
	NICEATM liaison	Bill Stokes
	ICCVAM liaison	Jill Merrill
	JaCVAM liaison	Hajime Kojima

Fig. 3: VMG Membership

The substances will be derived from existing and new chemicals databases but will exclude substances used to define the prediction model of each test method and used for pre-validation. Criteria for chemical selection will be based on 1) availability of high quality *in vivo* data; 2) irritant (GHS Cat. 1 or Cat. 2) versus non-irritant (NC); 3) physical form (solids versus liquids); 4) chemically reactive versus non-reactive. Accordingly, a target of 45/55 split for irritant versus non-irritant substances, a 40/60 split for solids versus liquids and 50 ±15% for chemical reactivity has been set to achieve an appropriate balance among the test substances. Within the group of irritant substances, a 50/50 split of GHS Cat. 1 and Cat. 2 categories will also be attempted. Finally, if possible, a good representation of Cat. 2B substances will be sought.

5 Study Management

The management structure of the study is shown in Figure 4. The Validation Management Group (VMG), which is responsible for overseeing the conduct of all aspects of the study, comprises a chair, co-chair, sponsor representatives (Colipa, ECVAM), coordinating organisation's (TNO) representative, an independent biostatistician, an ECVAM biostatistician, an external scientist, the chair of the CSG and representatives of the lead laboratories for each test method (L'Oréal, Beiersdorf). In addition, in the framework of the International Cooperation on Alternative Test Methods, liaisons from the USA, Japan and Canada are represented on the VMG. Members of the VMG are shown in Figure 3. Discussions regarding chemical selection do not involve the lead laboratories' representatives.

6 Participating Laboratories

Three laboratories for each test method will participate in the study. For the EpiOcular™ test method, the three laboratories are Beiersdorf (Germany; lead laboratory), Harlan Laboratories (UK) and the Institute for In Vitro Sciences (IIVS; USA). The three participating laboratories for the SkinEthic™ HCE test method are L'Oréal (France; lead laboratory), CARDAM (Belgium) and CeeTox Inc. (USA). These laboratories have been selected on the basis of proven expertise in the conduct of *in vitro* toxicity testing.

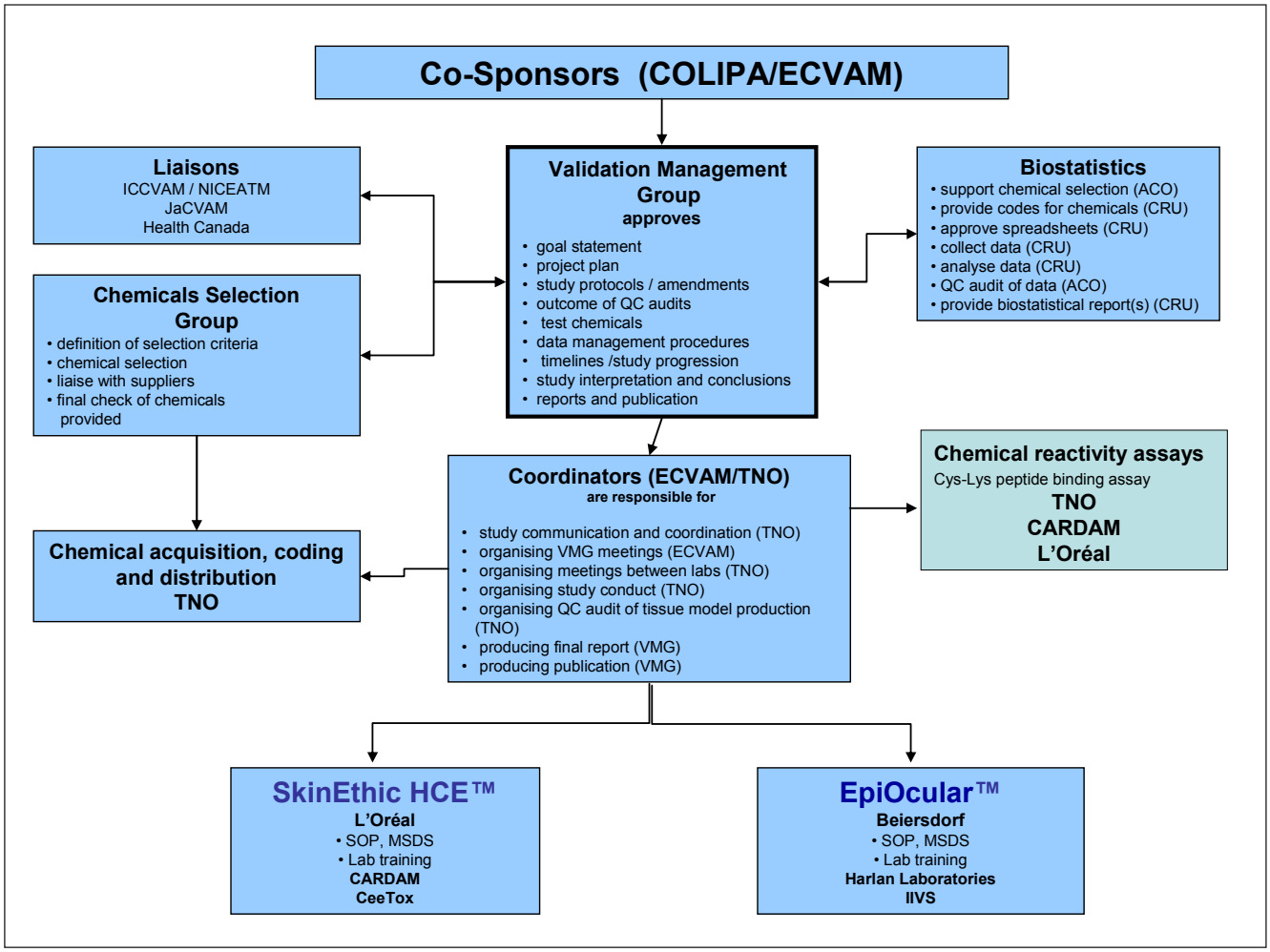


Fig. 4: Management Structure

DPRA tests on all test substances will be conducted by TNO (Netherlands). When the viability assays have been completed, a subset of test substances will undergo DPRA testing in two additional laboratories, L'Oréal and CARDAM. The final number of test substances in this subset will be determined by the study statisticians. When selecting the chemical subset for evaluation of DPRA reproducibility, weight will be given to chemicals that classify differently in the two time exposure regimens of the SkinEthicTM HCE protocol (10 min and 1 h + 16 h).

7 Data Collection, Handling and Analysis

Data collection, handling and analysis are the responsibility of the independent biostatistician. Results of the analysis will be provided in a final report which, together with statistical methodologies used, will be subject to quality control by the ECVAM biostatistician.

8 Quality Assurance

All participating laboratories and tissue suppliers will conduct work in the spirit of OECD GLP and will be subject to independent audit to ensure quality of work.

9 Timelines

Training of participating laboratories and transfer of methods is scheduled to take place by April 2010. Chemical selection and testing of chemicals in the DPRA shall be completed by the end of April 2010. Chemical coding and distribution will also have been completed by the end of May 2010 with the aim of enabling a start to the experimental phase as soon as possible thereafter.

References

Anon (2003). *Globally Harmonised System of Classification and Labelling of Chemicals (GHS). Part 3: Health and Environmental Hazards* (107-228). New York, NY, USA, and Geneva, Switzerland: United Nations Organisation.

Blazka, M. E. et al. (2005). EpiOcular™ human cell construct: tissue viability and histological changes following exposure to surfactants. *The Toxicologist 85*, 409.

Blazka, M. E., Harbell, J. W., Klausner, M. et al. (2003). Evaluating the ocular irritation potential of 54 test articles using the EpiOcular™ human tissue construct model (OCL-200). *The Toxicologist 72*, 221.

Blazka, M. E., Harbell, J. W., Klausner, M. et al. (2000). Colgate-Palmolive's program to validate the EpiOcular™ human tissue construct model. *The Toxicologist 54*, 188.

EC (2008). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. *Official Journal of the European Union L353*, 1-1355.

ESAC (2009). Statement on the Scientific Validity of Cytotoxicity/Cell-Function Based In Vitro Assays For Eye Irritation Testing. http://ecvam.jrc.ec.europa.eu/

Eskes, C., Bessou, S., Bruner, L. et al. (2005). Eye Irritation. In C. Eskes and V. Zuang (eds), Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Prospects. *ATLA 33, Suppl. 1*, 47-81.

Gerberick, F., Vassallo, J. D., Foertsch, L. M. et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicological Sciences 97 (2),* 417-427.

Ghassemi, A., Osborne, R., Kohrman, K. A. et al. (1997). Demonstrating the ocular safety of an eye cosmetic product using alternatives to animal eye irritation tests. *The Toxicologist 36*, 43.

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *ATLA 32*, 467-472.

Jester, J. V. (2006). Extent of corneal injury as a biomarker for hazard assessment and the development of alternative models to the Draize rabbit eye test. *Cutaneous and Ocular Toxicology 25*, 41-54.

Jester, J. V., Li, L., Molai, A. and Maurer, J. K. (2001). Extent of initial corneal injury as a basis for alternative eye irritation tests. *Toxicol. In Vitro 15*, 115-130.

Kaluzhny, Y., Thornton, L., Kandarova, H. et al. (2007). *Expanded Utilization of the EpiOcular™ Human Corneal Tissue Model for Ocular Irritation Testing*. Presented at 6th World Congress on Alternatives and Animal Use, Tokyo, August (2007).

ICCVAM (2009a). Draft Background Review Document – Current Status of In Vitro Test Methods for Identifying Mild/Moderate Ocular Irritants: Bovine Corneal Opacity and Permeability Test Method, National Institute of Environmental Health Sciences, Resesarch Triangle Park, North Carolina, USA (available: http://iccvam.niehs.nih.gov/methods/ocutox/PeerPanel09.htm).

ICCVAM (2009b). Independent Scientific Peer Review Panel Report: Evaluation of the Validation Status of Alternative Ocular Safety Testing Methods and Approaches, National Institute of Environmental Health Sciences, Resesarch Triangle Park, North Carolina, USA (available: http://iccvam.niehs.nih.gov/methods/ocutox/PeerPanel09.htm).

Maurer, J. K., Parker, R. D. and Jester, J. V. (2002). Extent of initial corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. *Regulatory Toxicology and Pharmacology 36*, 106-117.

Nguyen, D. H., Beuerman, R. W., de Wever, B. and Rosdy, M. (2003). Three-dimensional construct of the human corneal epithelium for in vitro toxicology. In H. Salem and S. A. Katz (eds), *Alternatives Toxicological Methods (147-159)*. CRC Press.

OECD (2008a). Draft Test Guideline 437: The Bovine Corneal Opacity and Permeability (BCOP) Test Method for Identifying Ocular Corrosives and Severe Irritants. http://www.oecd.org/dataoecd/37/8/41896718.pdf

OECD (2008b). Draft Test Guideline 438: The Isolated Chicken Eye (ICE) Test Method for Identifying Ocular Corrosives and Severe Irritants. http://www.oecd.org/dataoecd/37/10/41896968.pdf

OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (No. 34, OECD, 2005). http://www.oecd.org/document/30/0,3343,en_2649_34377_1916638_1_1_1_1,00.html

Scott et al. (2009). Bottom-up and Top-down Approach: Eye irritation testing strategy to reduce and replace in vivo studies (submitted).

Sheasgreen, J., Kubilus, J., Sennot, H. et al. (1996). Reproducibility and correlation of EpiOcular™, a three-dimensional tissue culture model of human corneal epithelium. *ATLA 24*, 284.

van Goethem, F., Adriaens, E., Alepee, N. et al. (2006). Prevalidation of a new in vitro reconstituted human cornea model to assess the eye irritating potential of chemicals. *Toxicol. In Vitro 20*, 1-17.

Correspondence to:
Dr. Stuart J. Freeman
Farino Consulting
Mercread
New Park Road
Cranleigh, Surrey UK
GU6 7HJ
e-mail: stuart.j.freeman@talktalk.net

BS24: Genotoxicity and carcinogenicity

short communication

ECVAM Prevalidation of Three Cell Transformation Assays

Philippe Vanparys[1,7], Raffaella Corvi[2,7], Marilyn Aardema[3,7], Laura Gribaldo[2,7], Makoto Hayashi[4,7], Sebastian Hoffmann[5,7] and Leonard Schechtman[6,7]

[1]Altoxicon BVBA, Vosselaar, Belgium; [2]In vitro Methods Unit/European Centre for the Validation of Alternatives Methods (ECVAM) and Molecular Biology and Genomics Unit, Institute for Health and Consumer Protection (IHCP), JRC of the European Commission, Ispra (Va), Italy; [3]The Procter & Gamble Co, Cincinnati Ohio, now Marilyn Aardema Consulting LLC, Fairfield, Ohio, USA; [4]Biosafety Research Center, Foods, Drugs and Pesticides, Shizuoka, Japan; [5]seh consulting + services, Cologne, Germany; [6]Innovative Toxicology Consulting, LLC, Lake Worth, Florida, USA; [7]Member of ECVAM Validation Management Team for Cell Transformation Assays

Summary

A prevalidation study on the cell transformation assays in SHE cells at pH 6.7, SHE cells at pH 7.0 and Balb/c 3T3 cell line was coordinated by ECVAM focussing on issues of standardisation of protocols, within-laboratory reproducibility, test method transferability and between-laboratory reproducibility. The Validation Management Team concluded that standardised protocols are now available that should be the basis for future use. The SHE pH 6.7, and the SHE pH 7.0 protocols and the assays system themselves are transferable between laboratories, and are reproducible within- and between-laboratories. For the Balb/c 3T3 method, some clarifications and modifications to the protocol were needed to obtain reproducible results. Overall, three methods have shown to be valuable to detect rodent carcinogens.

Keywords: validation, carcinogenicity, cell transformation assay, bioassay, regulatory toxicology

Overview of the prevalidation study

Development and ultimate utilisation of new chemical compounds requires, among other prerequisites, the assessment of human safety. One of the main endpoints in this assessment process is the determination of potential carcinogenicity. Historically, such an evaluation has necessitated the conduct of lifetime carcinogenicity bioassays in rats and/or mice. These studies take around 4 years (experimental phase and analysis of the results), cost around 1 million Euro per chemical and use significant numbers of animals. The European 7th Amendment to the Cosmetic Directive (EU, 2003), the new European chemical legislation REACH (Regulation EC, 2006) and the revised regulation on pesticides and biocides (Regulation EC, in press), all limit the use of animal tests, triggering the need for alternative methods. In fact, the 7th Amendment to the Cosmetic Directive completely bans animal testing for cosmetic ingredients and finished products since March 2009. REACH, on the other hand, requires data on carcinogenicity for chemicals manufactured in volumes greater than 1000 tons per year which are classified as somatic mutagens and are widespread in the environment or for which there is evidence for long term human exposure. As a result, it is expected that in the coming years, a high number of carcinogenicity studies will have to be carried out to fulfil the REACH requirements. Therefore there is a fundamental and critical need for the availability and implementation of validated alternative test models for carcinogenicity testing of chemicals, which can be used to reduce animal usage, refine current

in vivo test systems and replace animals that would otherwise be employed for such assays. Among the *in vitro* alternatives developed, the cell transformation assays (CTAs) are the most widely used. Despite their broad usage over decades of time, neither the identification of the ideal test method nor method standardisation of any of the available test methods have been fully resolved (Maurici et al., 2005). Nevertheless, the appeal offered by these assays is that they have been shown to involve a multistage process that closely models some stages of *in vivo* carcinogenesis (LeBoeuf et al., 1999), and therefore they are presumed to be worthy potential surrogates for rodent carcinogenicity systems.

Regulatory agencies have been reluctant to adopt these assays in their safety testing schemes, one of the reasons being the lack of formal validation.

On the basis of a conclusion made in a detailed review paper (DRP) of the OECD on cell transformation assays for the detection of chemical carcinogens (OECD, 2007) which concluded that the performance of the Syrian hamster embryo (SHE) and Balb/c 3T3 CTAs were sufficiently adequate and should be developed into formal OECD test guidelines. Further, the same OECD DRP recommended that although considerable and sufficient data on the performance of the assays were available, there was a need to develop standardised protocols and to assess the reproducibility of CTA results. On the basis of these conclusions and on recommendations of two expert meetings on cell transformation held at the European Centre for the Validation of Alternative Methods (ECVAM) (Combes et al., 1999), a formal prevalidation study on the Syrian hamster embryo (SHE) and Balb/c 3T3 CTAs was set up to address issues of standardisation of protocols, within-laboratory reproducibility, test method transferability, and between-laboratory reproducibility. Three variants of the CTAs were assessed: the CTA using SHE cells at pH 6.7, SHE cells at pH 7.0, and the Balb/c 3T3. In order to evaluate whether the tests would meet the criteria requested by the ECVAM principles on test validity (Balls et al., 1995), the modular approach to validation (Hartung et al., 2004) was followed. The reported study focused on the following four modules: test definition (e.g. definition of the test's scientific purpose, definite test protocol compliant with Good Laboratory Practice, prediction model, etc.), within laboratory reproducibility, transferability, between laboratory reproducibility. In addition, the fifth module, i.e. predictive capacity of the assay to predict the reference standard (i.e. *in vivo* test results), was preliminarily addressed in a limited way since only six chemicals were tested to serve that purpose.

In order to ensure that all study participants were adequately trained and that the respective test procedures were appropriately optimised, a preliminary study was conducted to specifically address those issues. Furthermore, the initial phase I of the study looked at the test definition and assessed both the within-laboratory reproducibility and transferability of the assay protocols by testing a non-coded and a coded compound. Subsequently, the between-laboratory reproducibility was determined by testing five additional coded compounds.

Each *in vitro* assay was conducted following the same agreed-upon protocol in four different laboratories for the SHE assay at pH 7.0 and in three different laboratories for the SHE assay at pH 6.7 and the Balb/c 3T3 assay. The laboratories involved encompassed industry, academia, contract research laboratories (CROs) and government establishments located in the USA, Japan and Europe. The chemicals were selected (Tab. 1) using data from the OECD DRP31 document (version August 2004) and a publication of Kirkland et al. (2005). The same chemicals were used for the SHE pH 6.7 assay and SHE pH 7.0 assay. Where possible the same chemicals were selected for the Balb/c 3T3 assay.

The following criteria were used to select the chemicals: (1) positive both in Balb/c 3T3 and in SHE, (2) negative both in Balb/c 3T3 and in SHE, (3) at least two references for each test chemical (for both Balb/c 3T3 and SHE), (4) if possible, data available using the SHE pH 7.0 and pH 6.7 protocol, (5) clear classification as *in vivo* carcinogen or non-carcinogen. All of these criteria could not be met in all cases.

As part of this validation exercise, photo catalogues for each variant of the respective CTAs were produced by the partici-

Tab. 1: Chemicals selected

Compound	*In vivo* carcinogenesis	SHE pH 6.7	SHE pH 7	Balb/c 3T3
Benzo[a]pyrene [a]	positive	X	X	X
2,4-Diaminotoluene	positive	X	X	
3-Methylcholanthrene [b]	positive	X	X	X
o-Toluidine HCl	positive	X	X	X
Anthracene	negative	X	X	X
Phthalic anhydride	negative	X	X	
2-Acetylaminofluorene	positive			X
Phenanthrene	negative			X

a: Positive control for the SHE pH 6.7 and pH 7.0 assay
b: Positive control for the Balb/c 3T3 assay

pating laboratories with the aim of establishing consistency in assessing colony/focus morphology and for the scoring experiments (see Fig. 1 and 2). The morphological criteria used to identify transformed colonies and transformed foci were adopted from Berwald and Sachs (1963, 1965), Kakunaga (1973), Reznikoff et al. (1973) and Schechtman (1985a,b).

Details on the test procedures followed, assay acceptance criteria, assessment criteria and photo catalogues and the results of the study will be published in a special issue on CTAs in Mutation Research (in preparation).

The three CTA assays are now in the reporting phase. In the present validation study three optimised and standardised protocols for the CTA in SHE cells at pH 6.7, in SHE cells at pH 7.0 and in Balb/c 3T3 cells have been established and assessed for their reproducibility and reliability. Each variant of the assay showed good within-laboratory reproducibility in all laboratories. The transferability of the assays was shown to be successful. Furthermore, some of the laboratories had no previous experience in working with such assays. This suggests that CTA test methods can be easily transferred to any laboratory that has experience in cell culture techniques. However, since scoring of transformed colonies is at the moment still done manually under the microscope, training is necessary to ensure a scoring which is as objective and consistent as possible. It also should be noted that the dose-level selection is a crucial step for the success of CTAs, since the right doses need to be hit in order to detect a significant number of transformed colonies when the cells are treated with transforming agents. The between-laboratory reproducibility was shown to be satisfactory for the three assays. The concordance between the CTAs and the carcinogenicity classification of the chemicals assessed was satisfactory. Unexpected results were produced with phthalic anhydride in SHE cells at pH 6.7 and with phenanthrene in Balb/c 3T3 cells.

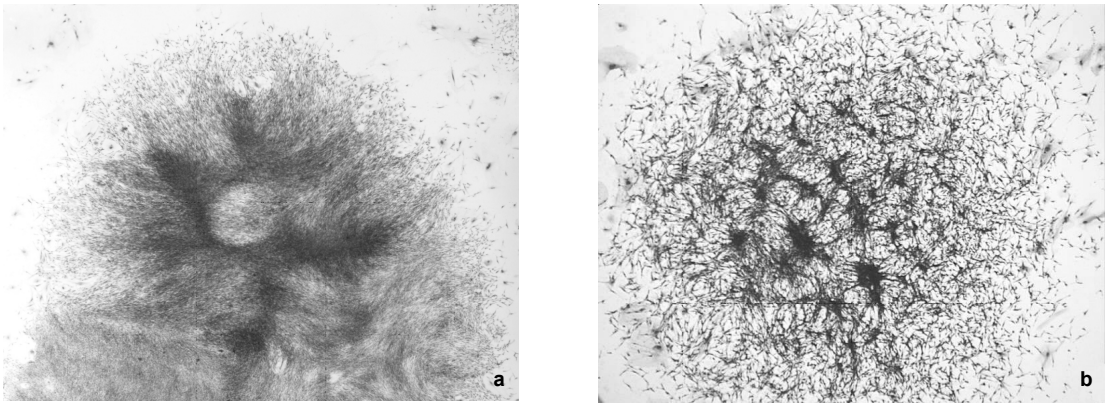


Fig. 1: A non transformed colony (a) and a transformed colony (b) produced upon treatment with benzo(a)pyrene in the SHE cell transformation assay.

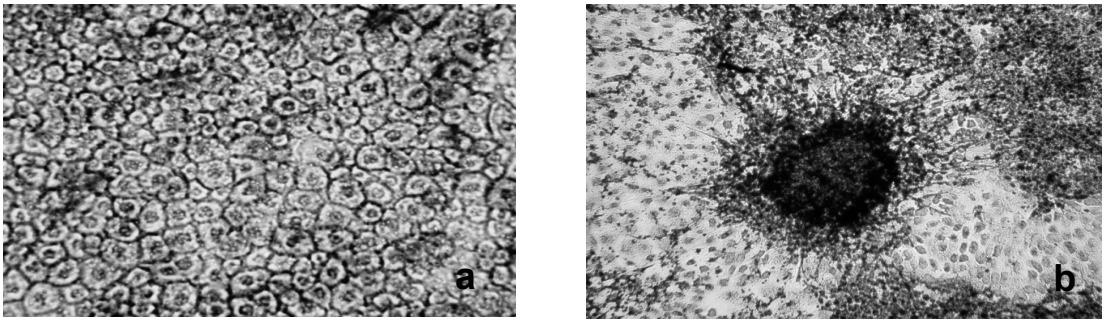


Fig. 2: Non transformed cells (a) and a transformed type III focus (b) in Balb/c 3T3 cells upon treatment with 3-methylcholanthrene.

The Validation Management Team concluded that standardised protocols are now available that should be the basis for future use of CTAs. The SHE pH 6.7, and the SHE pH 7.0 protocols and the assays system themselves are transferable between laboratories, and are reproducible within- and between-laboratories. For the Balb/c 3T3 method, an improved protocol has been developed, which allowed to obtain reproducible results. Further testing of this improved protocol is recommended in order to confirm its robustness. Overall, these results in combination with the extensive database summarised in the OECD DRP31 (OECD, 2007) support the utility of *in vitro* CTAs for the assessment of carcinogenicity potential.

References

Balls, M., Blaauboer, B. J., Fentem, J. H. et al. (1995). Practical Aspects of the Validation of Toxicity Test Procedures. ECVAM Workshop Report 5. *ATLA 23*, 129-147.

Berwald, Y. and Sachs, L. (1965). In vitro transformation of normal cells to tumor cells by carcinogenic hydrocarbons. *J. Natl. Cancer Inst. 35*, 641-661.

Berwald, Y. and Sachs, L. (1963). In vitro transformation with chemical carcinogens. *Nature 200*, 1182-1184.

Combes, R., Balls, M., Curren, R. et al. (1999). Cell transformation assay as predictors of human carcinogenicity. *ATLA 277*, 45-767.

EU (2003). Directive 2003/15/EC of the European Parliament and the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Members States relating to cosmetic products. *Official Journal of the European Union L66*, 26-35.

Hartung, T., Bremer, S., Casati, S., Corvi, R. et al. (2004). A modular approach to the ECVAM principles on test validity. *ATLA 32*, 467-472.

Kakunaga, T. (1973). A quantitative system for assay of malignant transformation by chemical carcinogens using a clone derived from BALB/3T3. *Int. J. Cancer 12*, 463-473.

Kirkland, D., Aardema, M., Henderson, L. and Müller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. *Mutat. Res./Genetic Toxicology and Environmental Mutagenesis 584*, 1-256.

LeBœuf, R. A., Kerckaert, K. A., Aardema, M. and Isfort, R. J. (1999). Use of Syrian hamster embryo and Balb.c 3T3 cell transformation for assessing the carcinogenic potential of chemicals. *IARC Sci. Pub. 146*, 409-425.

Maurici, D., Aardema, M., Corvi, R. et al. (2005). Carcinogenicity. *ATLA 33, Suppl. 1*, 177-182.

OECD (2007). Detailed review on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31.

Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006R1907:EN:NOT

Regulation of the European Parliament and of the Council concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. *EU Official Journal*, in press.

Reznikoff, C. A., Bertram, J. S., Brankow, D. W. and Heidelberger, C. (1973). Quantitative and Qualitative Studies of Chemical Transformation of Cloned C3H Mouse Embryo Cells Sensitive to Postconfluence Inhibition of Cell Division. *Cancer Res. 33*, 3239-3249.

Schechtman, L. M. (1985a). Metabolic Activation of Procarcinogens by Subcellular Enzyme Fractions in the C3H 10T 1/2 and BALB/c 3T3 Cell Transformation Systems. *IARC Sci. Publ. 67*, 137-162.

Schechtman, L. M. (1985b). BALB/c 3T3 Cell Transformation: Protocols, Problems and Improvements. *IARC Sci. Publ. 67*, 165-184.

Correspondence to

Raffaella Corvi
In vitro Methods Unit/European Centre for the
Validation of Alternatives Methods (ECVAM)
Institute for Health and Consumer Protection (IHCP)
JRC of the European Commission TP580
Via E. Fermi 2749
21027 Ispra (Va), Italy
e-mail: raffaella.corvi@jrc.ec.europa.eu

Session BS26: Disease models

Refining Clinical Care in Animal Disease Models

Karl A. Andrutis

Animal Care Services, University of Florida, Gainesville, USA

Summary

Refinement of clinical or supportive care in animal disease models involves the use of therapeutic measures to provide basic physiological support to animals used for biomedical research. Development and implementation of clinical care to research animals represents a significant refinement of animal disease models. Supportive care is routinely provided in clinical situations yet it is uncommon in many animal disease models. The range of supportive care measures to consider extends from the very basics of thermal and fluid support to the more complicated of cardiovascular and respiratory support. Several areas of study have demonstrated that improved supportive care measures do contribute to the refinement of animal disease models. Enhanced development and implementation of clinical care for research animals will not only improve the health and welfare of those animals but will also likely improve the scientific results from animal disease models.

Keywords: refinement, clinical care, supportive care, animal models, disease models

1 Introduction

Refinement of clinical or supportive care in animal disease models involves the use of therapeutic measures to provide basic physiological support to animals used for biomedical research. The aspects of basic physiological support include maintenance of body temperature, hydration status, acid-base and electrolyte balance, cardiovascular support, and respiratory support. Development and implementation of clinical care to research animals represents a significant refinement of animal disease models.

2 Background

Supportive care is the basis of all medical treatment for humans and animals. It is a major component of all disease therapies and provides the foundation for patient stabilization from which specific disease therapies may be applied. Restoration and maintenance of homeostasis with supportive care measures allows the physiology of the patient to function normally as the disease process is addressed by more specific treatments.

The intent of an animal model of disease is to mimic the human or animal disease for which it is a model and to exhibit similar, if not identical, pathophysiological processes of the disease. For models in which therapies are evaluated, these pathophysiological processes are interrupted and, hopefully, corrected to restore the health of the research animal. For these animal models to be truly representative of human or animal clinical disease, the experimental treatments must be applied in a similar clinical care context. If a specific therapy is administered to an animal disease model and the animal's basic physiological needs are not met, then the therapy is not likely to succeed even if it is an effective therapy. Of course, this depends on many things including the original health of the animal, the state and course of the disease, and the timing of treatment administration. However, supportive care is routinely provided in clinical situations yet it is uncommon in many animal disease models. This discrepancy is a matter of concern as we search for new therapies for human and animal diseases and discard therapies that may have failed preclinical studies due to the lack of supportive care of the research animals. It may also be the reason that, in many cases, preclinical study results can-

not be replicated in clinical studies (Dyson and Singer, 2009; Hofstra et al., 2008).

A review of the literature indicates that, in human and veterinary medicine, supportive care is routinely provided as part of any clinical study and treatment evaluation. In contrast, review of the preclinical research study literature shows that few studies describe supportive care measures for their research animals (Dyson and Singer, 2009). Since these same papers do describe their experimental therapies in detail, it seems clear that supportive care measures were not used in these studies; it is not simply that they were excluded from the publication. Of course, there may be instances when specific treatments need to be evaluated individually without supportive care measures. Treatments may be administered prior to the onset of clinical disease or early in the course of disease before physiological functions are significantly disturbed. However, once physiological functions are altered, can one be certain that a treatment would have the same effect if administered to an animal with normal physiology?

Some may argue that providing supportive care to research animals adds a variable to the study design and may confound the interpretation of study results. However, supportive care measures can be controlled similarly to other experimental variables. In addition, the variation in abnormal physiological states in different research animals is usually neither measured nor controlled and may represent an even more significant research variable than supportive care measures (Hofstra et al., 2008). Supportive care is certainly not withheld in human and animal clinical studies in which physiologic homeostasis would seem to outweigh the concern that supportive care represents a significant experimental variable.

3 Supportive care measures

The range of supportive care measures to consider extends from the very basics of thermal and fluid support to more complicated measures such as cardiovascular and respiratory support. Due to the small size and relatively large surface area of rodents, body temperature regulation represents a critical variable in these research animals. Even in larger laboratory animals, maintenance of body temperature is critically important in any disease process. Assistance with the maintenance of body temperature is not difficult to provide, and many creative and effective methods have been developed.

Another relatively simple supportive care measure is fluid therapy. Many disease processes result in fluid shifts within the body and clinical dehydration from reduced consumption or increased loss of fluids. Provision of fluid therapy is not difficult if other aspects of the research animal's physiology allow the use of subcutaneous fluids. Even minor acid-base and electrolyte imbalances can be addressed with subcutaneous fluids. More significant imbalances may need to be addressed with intravascular administration of fluid therapy, which is relatively uncomplicated in large animals but somewhat more complex, yet still possible, in rodents. Another route of administration for fluids in rodents is the intraperitoneal route, which can accommodate significant volumes of fluid. In my experience, hydration status is often overlooked as a component of morbidity and a cause of mortality in rodent disease models. Again, due to their size and metabolic rate, rodents are very susceptible to dehydration, and hydration status should be evaluated in any ill rodent. Clinical assessments of skin turgor and mucous membranes will usually provide a good estimate of hydration status, and measuring hematocrit and total protein should not be dismissed even though blood collection must be limited in rodents.

Provision of cardiovascular and respiratory support can also range from basic to complex. Normalization of hydration status may itself enhance cardiac output and perfusion of organs and peripheral tissues. Specific cardiovascular disturbances may require the use of specific cardiovascular therapies. While drug interactions need to be identified and understood, use of specific cardiovascular therapies may still represent the best clinical disease model as current disease treatments often involve a combination of drugs. Respiratory support in the form of supplemental oxygen is not difficult to provide with some ingenuity with laboratory animal caging and can have a significant effect on physiological processes.

Two recent papers, one using mice (Schuler et al., 2009) and one using swine (Murison et al., 2009), illustrate these points. Schuler and colleagues attempted to optimize surgical techniques and postoperative care to improve survival rates and permit accurate telemetric recording in exercising mice. They established a regimen of pain treatment and fluid therapy administered for 7 days. In addition, supplemental warmth and free access to a high energy food source for 14 days was provided. This regimen led to a substantial decrease in overall morbidity and mortality. Murison and colleagues attempted to develop effective perioperative care for pigs undergoing laryngeal transplantation. They used intensive monitoring and novel circulatory and airway management methods to provide optimal perioperative care and improve survival. These two examples demonstrate that the development and implementation of basic supportive care measures in animal disease models is attainable and has the potential not only to improve animal welfare but to reduce the numbers of animals used and provide more reliable scientific results.

4 Case studies

Several areas of study have demonstrated that improved supportive care measures do contribute to the refinement of animal disease models. Minneci and colleagues have developed a canine model of septic shock which they believe balances animal welfare with scientific relevance (Minneci et al., 2007). Their prospectively determined criteria of a relevant and successful model were to closely simulate the clinical syndrome as encountered in humans and to be applicable to, and consistent with, the best practices of both human and veterinary medicine.

The group used purpose bred beagle dogs, which received increasing doses of *Staphylococcus aureus* directly into the lung. The dogs were anesthetized and instrumented with a tracheos-

tomy, femoral arterial and jugular vein vascular catheters, and a urinary bladder catheter. Mechanical ventilation was initiated and maintained while the dogs received fluids, antibiotics, and vasopressors for 96 hours. Intravascular hemodynamics and cardiac function were measured. Treatment of sepsis was individualized to the hemodynamics, oxygenation, and ventilation needs of each dog, similar to human care. The level of vasopressor and ventilator support was dictated by algorithms and adjusted according to continuously measured oxygen saturation, mean arterial pressure, intermittent pulmonary capillary wedge pressure, and arterial blood gases. They also provided thermal support to maintain core temperature.

This canine model of sepsis incorporated treatments based on algorithms very similar to the titrated care routinely provided to critically ill patients. Supportive care was adjusted to the animal's physiological needs. The authors caution that because it may be more difficult to differentiate the effects of new treatments themselves from those related to differences in the level of support animal receive, greater numbers of animals may need to be studied. However, provision of supportive care may actually reduce the number of animals needed by decreasing the variation between animals. Also, titration of supportive therapies based on the individual needs of the animals is similar to clinical care, making the model potentially more relevant.

Neuromuscular disease is another area of study that has incorporated many supportive care refinements in animal models. These models often affect the animal's ability to ambulate and perform normal bodily functions. There are many factors to consider when caring for these research animals, including the animal's environment, access to food and water, facilitation of normal biological processes, environmental enrichment, social housing, and appropriate veterinary care and monitoring (Wallace and Sikoski, 2007). More specifically, thermal support, caging and soft bedding, food and water on the floor of the cage, bladder expression, and physical therapy or massage may need to be provided. Davis has described the refinement of a mouse model of experimental autoimmune encephalomyelitis in which they were able to develop improved methods for assessment and intervention with supportive care measures (Davis, 1999/2000).

Santos-Benito and colleagues have provided an overview of the proper handling and care of paraplegic adult laboratory animals (rats and monkeys) for long periods (up to one year) (Santos-Benito et al., 2006). They assert that "the use of human treatments in paraplegic animals provides a more realistic model for a later transfer to the clinical arena". They recommend the use of antibiotic and analgesic drugs just before and immediately after surgery. In rats, they prefer to perform all intramuscular injections in the forepaw as paraplegic animals tend to have muscle atrophy in the hind limbs. They provide fluids to prevent hypovolemia and thermal support to prevent hypothermia during surgery. Immediately after surgery, the bladder is expressed manually.

The substrate in the cages should be soft, smooth, and absorbent to minimize pressure ulcers. After surgery, food is provided *ad libitum* until body weight is stabilized, when the quantity of food is adjusted individually to maintain a desired body weight. This is important to be able to compare the motor improvement achieved by different rats. Rats are given a daily health exam, and the bladder is expressed up to four times daily until automatic voidance is reestablished (3-4 weeks). Water intake is controlled to limit the amount of urine in the bladder, and the rats are deprived of water during the night to avoid over-distention of the bladder. Rats are provided with abdominal massage to restore normal intestinal transit. Rats are also provided with physiotherapy and rehabilitation by passive movement of all joints and massage of all muscles below the lesion and they are allowed to freely move in an open field.

The most critical period is immediately after the spinal cord lesion and also during the first 3 months post-injury. The most common complications include loss of weight from dehydration treated with subcutaneous fluids or lack of appetite treated with a nutritional supplement; constipation treated with increased abdominal massages and paraffin oil laxative; urinary tract infections treated with antibiotics; hematuria treated with gentle bladder management, antibiotics, and bladder lavage; pressure ulcers treated with massages, cleaning, ointments, and antibiotics; dysesthesias treated as for ulcers and bandaging; and vaginal infections treated with vaginal suppositories.

These supportive care measures have allowed these researchers to maintain rats with complete spinal cord injury for long periods of time. The medical treatments they employed were adapted from those used in paraplegic people and provide a more realistic animal model to evaluate therapies for spinal cord injury.

5 Conclusion

In summary, it is important to consider supportive care measures as a significant refinement in any animal model in which physiological disturbances occur as a result of the disease process. Restoration and maintenance of normal physiological function is imperative to fully understand the effectiveness of more specific therapies in disease models. Enhanced development and implementation of clinical care for research animals will not only improve the health and welfare of those animals but is also likely to improve the scientific results from animal disease models.

Challenge those who say that supportive care is not warranted or possible with the fact that supportive care is routinely provided to patients in clinical settings. Perhaps we need a change in mind set to one in which our research animals are not considered our research subjects but our research patients.

References

Davis, J. A. (1999/2000). The triple A approach to ensuring animal welfare. *Animal Welfare Information Center Bulletin 10 (3/4),* 1-9.

Dyson, A. and Singer, M. (2009). Animal models of sepsis: why does preclinical efficacy fail to translate to the clinical set-

ting? *Critical Care Medicine 37 (1),* S30-S37.

Hofstra, J-J. H., Juffermans, N. P., Schultz, M. J. and Zweers, M. M. (2008). Pulmonary Coagulopathy as a new target in lung injury – a review of available pre-clinical models. *Current Medicinal Chemistry 15 (6),* 588-595.

Minneci, P. C., Deans, K. J., Hansen, B. et al. (2007). A canine model of septic shock: balancing animal welfare and scientific relevance. *Am. J. Physiol. Heart Circ. Physiol. 293*, H2487-H2500.

Murison, P. J., Jones, A., Mitchard, L. et al. (2009). Development of perioperative care for pigs undergoing laryngeal transplantation: a case series. *Laboratory Animals*, 1-6.

Santos-Benito, F. F., Munoz-Quiles, C. and Ramon-Cueto, A. (2006). Long-term care of paraplegic laboratory animals. *Journal of Neurotrauma 23 (3/4)*, 521-536.

Schuler, B., Rettich, A., Vogel, J. et al. (2009). Optimized surgical techniques and postoperative care improve survival rates and permit accurate telemetric recording in exercising mice. *BMS Veterinary Research 5 (28)*, 1-11.

Wallace, J. M. and Sikoski, P. (2007). Animal care and use issues in movement disorder research. *ILAR Journal 48 (4)*, 317-322.

Correspondence to

Karl A. Andrutis DVM, MS, DACLAM
Clinical Assistant Professor and Assistant Director
Animal Care Services
University of Florida
Gainesville, FL 32610, USA
e-mail: kandrutis@ufl.edu

BS27: Environmental science

Use of Mode and Mechanism of Action Information to Support *In Silico* Prediction of Ecotoxicity

Mark T. D. Cronin

School of Pharmacy and Chemistry, Liverpool John Moores University, UK

Summary

An overview is provided of the use of information from modes and mechanisms of toxic action to predict acute and chronic ecotoxicity effects. The information from modes and mechanisms can be used to provide better in silico estimates through integrated testing strategies. The use of the information is better developed for the prediction of acute rather than chronic toxicity. Mechanistic information can be used to select an appropriate (quantitative) structure-activity relationship (QSAR) or form a category of chemicals to facilitate read across. Relevant approaches to predict, by in silico means, mechanisms and modes of toxic action are discussed. Mechanistic information can be supplemented by the use of results from existing in vivo and further in vitro tests.

Keywords: QSAR, mode of toxic action, mechanism of toxic action, toxicity, integrated testing strategy

1 Introduction

In silico approaches in toxicology attempt to predict harmful effects of compounds from their chemical structure. These approaches include the use of (quantitative) structure-activity relationships ((Q)SARs), category formation and read-across (van Leeuwen et al., 2009). Since the seminal publication of Könemann (1981) the possibility to use *in silico* methods for the prediction of environmental effects has been realised and widely applied and the importance of a mechanistic understanding in developing and applying (Q)SARs has been appreciated (Bradbury et al., 2003; Russom et al., 2003).

Whilst a number of (formal) definitions of mechanism and mode of toxic action exist, there is still confusion within the scientific community. For the purposes of this paper, the following general definitions are applied: In terms of the use of QSAR for the prediction of acute ecotoxic effects, generally a "mechanism of action" has been taken to mean a specific biochemical effect, e.g. non-polar narcosis or polar narcosis. The use of the term "mechanism" in this regard can be supported by techniques such as binary toxicity studies (Hodges et al., 2006); fitting to a known QSAR (Cronin, 2006); and the mechanistic knowledge derived from fish acute toxicity syndromes (McKim et al., 1987a,b,c). "Mode of action for acute toxicity" is a broader definition across mechanisms, e.g. the narcotic mode of action can cross a number of mechanisms (e.g. non-polar; polar; ester; amine) where the same physiological effect is seen. The distinction between "mode" and "mechanisms" for chronic toxicity becomes more blurred. Taking reproductive effects such as endocrine disruption as an example, it becomes complicated by level of exposure, stage of development, etc. It is the author's opinion that whilst definitions and distinctions are helpful, they are not crucial; however if this important information is known, it can be captured into expert system approaches to assist in the making of predictions.

With regard to utilising mechanism and mode of action information to predict toxicity, the prediction of ecotoxicity effects has been better developed for acute rather than chronic aquatic toxicity. There are a number of potential reasons for this. There are more acute aquatic toxicity data publicly available

for modelling than for chronic toxicity. In addition, mechanisms of action are better developed (and possibly more obvious) for acute effects than the subtleties of chronic effects. The prediction of acute aquatic toxicity was also immensely supported by the efforts of the United States Environmental Protection Agency (US EPA) through the development of the fathead minnow (*Pimephales promelas*) acute toxicity database (for a summary of the database see Russom et al., 1997), basic research on mechanisms (Fish Acute Toxicity Syndromes – McKim et al., 1987a,b,c) and the ECOSAR expert system (available free of charge from the US EPA via the web-site www.epa.gov). These efforts have been supported by other initiatives, such as the development of the Tetratox database for *Tetrahymena pyriformis* acute toxicity data, which has extended the QSAR concept and knowledge on domains (Schultz, 1997; Seward et al., 2001).

As new legislation (e.g. REACH) requires new methods to use current, and develop novel, toxicity information (Schaafsma et al., 2009), the concept of integrated testing strategies (ITS) is being developed. These strategies attempt to use existing information and non-test information either to replace or reduce animal testing or to prioritise chemicals for testing. An important component of an ITS is the use of *in silico* techniques and approaches (Grindon et al., 2008b). The utility of ITS to predict environmental effects has been described by Grindon et al. (2008a). These applications include the use of mode and mechanism of action information to support predictions in the following areas:

- Selection of an appropriate (Q)SAR on which to make a prediction. This is particularly powerful when a domain (i.e. the types of chemicals to which it can be applied) is well known and a robust and well characterised QSAR is available. This is the case, for instance, for non-polar narcosis, whereby the domain of the mechanism is becoming increasingly well established and simple (hydrophobicity dependent) QSARs are available.
- Placing a compound into an appropriate category (or grouping). This technique may allow for the related better prediction of chronic effects, i.e. non-acute toxicity a particular fragment of a molecule such as an acid group associated with compounds demonstrating impaired reproductive capability or teratogenicity.
- Development of strategies and (non-animal) testing to deal with current toxicological problems. Mechanistic knowledge will allow for the selection of compounds to test on the basis of the appreciation of how structure will affect activity. A good example is of the development of an expert system for the prediction of oestrogenic activity (Aladjov et al., 2009).

In addition, an understanding of modes and mechanisms of action will allow for "intelligent" testing strategies. This implies that the testing of chemicals can be performed in a rational and logical manner to maximise the information from testing on a mechanistic basis. It should be noted that there is a clear distinction between intelligent and integrated testing strategies. Intelligent testing strategies can be applied to optimise the utilisation of testing resources to gain further knowledge on chemicals. This is achieved by testing chemicals within, and to describe, a mechanistic domain with the minimum of assays. It is very amenable to *in vitro* tests. Integrated testing strategies are the collation of evidence for a single chemical to enable a decision to be made with regard to its toxicity, such that classification and labelling or other regulatory decisions can be made.

The aim of this paper, therefore, was to consider the ability of current QSARs to predict mechanisms and modes of action and utilise the information to estimate levels of toxicity for animals of relevance in environmental science, particularly within an ITS framework. In addition, consideration was given as to what is required to ensure close interaction between experimental biologists and modellers.

2 Uses of information from modes and mechanisms of toxic action to predict acute ecotoxicity

Within an integrated testing strategy for acute fish toxicity assessment, one of the key processes of the *in silico* evaluation is to attempt to filter a compound mechanistically, i.e. to identify the mechanistic domain in which it is placed. If a compound can be placed into an appropriate QSAR based on a mechanism of action, then greater confidence can be assigned to the prediction. There are a number of methods to assign a mechanism of action (in terms of acute ecotoxicity) to a compound, which are outlined below.

A number of "rules" based on structural chemistry have been developed, the most famous and widely applied of these is the Verhaar classification scheme (Verhaar et al., 1992). This provides a scheme for placing a compound into one of four Classes (from non-polar narcotic to specifically acting) or not being able to make a decision (the so-called Class 5). The Verhaar classification scheme has been coded into a number of pieces of software e.g. Toxtree (Pavan and Worth, 2008) and the OECD (Q)SAR Application Toolbox (downloadable from the OECD website, www.oecd.org), thus is freely available. Whilst the Verhaar scheme provides a framework for assigning a compound to an appropriate mechanism of action, it is limited in terms of current knowledge and some attempts have been made to update it (cf. Enoch et al., 2008).

In addition, a number of other approaches to assigning a mechanism exist. These include the Russom rules (Russom et al., 1997) and the OASIS mechanistic filter (see the information contained within Dimitrov et al., 2003a, 2003b, 2004). The latter of these is also available in the OECD (Q)SAR Application Toolbox. Thus, a number of freely available approaches exist for assigning a chemical to a mechanism of action, with the express purpose of assigning it to an appropriate QSAR. The application of these approaches, in terms of their potential use in an integrated testing strategy, is described below.

The formation of chemical categories is seen as a method of facilitating the chemical risk assessment process (van Leeu-

wen et al., 2009). This can be performed in terms of acute aquatic toxicity by grouping chemicals together on the basis of analogues and/or the same structural feature. This may provide evidence if one is able to accept that compounds demonstrating structural similarity are likely to have the same mechanism of action. Such grouping approaches may not be able to direct the user to a particular QSAR, but will assist in the development of local models and/or read-across. The OECD (Q)SAR Application Toolbox facilitates this and also contains many acute aquatic toxicity data.

Other attempts have been made to predict mode or mechanism of action from a chemical structure, mainly using conventional QSAR descriptors and approaches such as discriminant analysis (for examples of these approaches see the papers of Aptula et al., 2002; Nendza and Muller 2001, 2007; Nendza and Wenzel, 2006; Spycher et al., 2004). Unfortunately several of these approaches have been based on a dataset of phenolic compounds, and extrapolations made and claimed outside of the domain of the data matrix. However, they may supply supporting information to complement the use of schemes based on structural features.

Other methods of assigning a chemical to, or (more likely) excluding it from, a mechanism of action are the use of cut-offs from physico-chemical properties. For instance, the solubility of a compound will determine whether it will be toxic in an acute assay (highly hydrophobic compounds may not be toxic at saturation); volatile compounds may be lost from static test systems, etc. Another issue is that for short term aquatic toxicity assays, hydrophobic compounds may not reach equilibrium rapidly enough for a non-narcotic effect to be observed. Thus, at high log P, the toxicity of all compounds may be well modelled by a QSAR for the non-polar narcosis mechanism of action – for instance the toxicity of all compounds with a log P value greater than of 3.5 model as non-polar narcotics to *Tetrahymena pyriformis* (Schultz, 1997). Thus, in terms of predicting acute toxicity, such compounds can be considered to be non-polar narcotics, whilst in reality they may have the potential for another toxic mechanism of action.

3 Uses of information from modes and mechanisms of toxic action to predict chronic ecotoxicity

The use of modes and mechanisms of action is much less developed for chronic than for acute ecotoxicity. Many chronic toxic effects have relatively little supporting information relating to mechanisms of action and how chemical structure can be used to develop mechanistically based categories. A number of applications of mechanistic knowledge can, however, be made. For instance, it is known that for narcotic chemicals acute to chronic ratios may be more predictable and accurate (Ahlers et al., 2006). In addition, low to high dose and interspecies predictions may be more accurate for narcotic chemicals. Thus, if it can be decided, either through the means listed above or by other means, whether a compound is acting by a narcotic mechanism or not, this would be of potentially great advantage.

A clear exception to the assumption that there is limited mechanistic information for chronic toxicity is for endocrine disruption, where there is much high quality research on which to base mechanistic hypotheses. Much of the information on the action molecules on the receptors of the endocrine systems (e.g. oestrogen, androgen, thyroid, etc.) has been derived originally from pharmaceutical studies. In terms of identifying a compound able to act by a non-lethal mechanism of action, many studies will be based around the ability of a compound to bind to a receptor, and if it is able to bind, then the relative binding affinity may provide some estimation of potency. Excellent illustrations for this are provided for the oestrogen receptor (ER), including the use of a four-tier system to identify potential binders to the ER (Shi et al., 2001).

An interesting example of how mechanisms might be applied for chronic effects is in the area of pharmaceuticals. The consequence of low level exposure of environmental species to pharmaceuticals is of concern (Madden et al., 2009). Few data exist for the chronic effects of pharmaceuticals in the environment. Less is known about whether it may be possible to introduce knowledge of mammalian mechanisms of action (pharmacological and toxicological) into the assessment of chronic effects in the environment. This is an intriguing area for further research, particularly in the broader context of whether mammalian mechanisms may be transferable (and how to achieve this) to environmental species.

4 Supplementing integrated testing strategies with input from experimental biology

There are a number of areas where biology can assist in the use of mechanism and mode of action information for predicting ecotoxiological effects through the use of ITS. The key purpose of these efforts is to increase confidence in the ability to predict whether a compound belongs to a particular mechanism or mode of action.

In particular, basic knowledge is required on modes and mechanisms of action for both acute and chronic effects. In terms of assigning a compound to a QSAR for an acute effect, it must be remembered that the majority of our knowledge comes from a very small number of chemicals – this information has been extrapolated very widely. For instance, in terms of *in vivo* knowledge, the gold standard includes the Fish Acute Toxicity Syndromes developed in the 1980s for only a small number of compounds (McKim et al., 1987a). From this, mechanisms were assigned to all chemicals in the fathead minnow database (over 600) and even further beyond. Mechanistic assumptions can be supplemented by binary toxicity testing to identify compounds with additive effects – the assumption being that additive toxicity will occur only for compounds acting by the same mechanism (Hodges et al., 2006).

The need here is not to perform further *in vivo* testing but to use (existing) test results and information better, e.g. to determine symptoms associated with mechanisms. This may even provide an opportunity to optimise *in vivo* testing by identifying key parameters in the test regime associated with the early and better identification of effects. There is also a real need to derive strategies from *in vitro* and -omics assays to support mechanistic interpretation. This may be as simple as using cytotoxicity data to confirm whether a compound is a narcotic or not (i.e. whether it conforms to the so-called baseline effect (Cronin, 2006; Maeder et al., 2004)). Alternatively it may be the application of receptor based assays for specific endpoints relating to acute or chronic toxicity or the development of fingerprints from -omics relating to particular modes of action.

Another area where experimental biology can assist in developing useful information is with regard to metabolism. Metabolism is obviously an important process in toxicology, both for activation and deactivation. Current *in silico* methods can identify potential metabolites but are by no means comprehensive or able to predict accurately the amount or stability of a metabolite. Biological experiments (*in vivo* and *in vitro*) could assist in identifying compounds susceptible to significant metabolism and whether toxicologically important metabolites are formed.

In addition to information from biological tests, potentially useful information may be obtained from non-biological tests. For instance, the so-called *in chemico* tests (Gerberick et al., 2008) identify reactive compounds (i.e. those capable of forming a covalent bond with a biological pseudo-macromolecule). Compounds exhibiting such reactivity may be associated with a number of acute and chronic toxicities and could, for instance, be excluded from narcotic mechanisms. The tests that can be considered include those for electrophilic and nucleophilic reactivity, redox potential and formation of free radicals. Such non-biological test information may be invaluable for identifying specific reactivity or mechanisms not associated with receptor binding.

The concept, therefore, is that a battery of tests can be developed to assist in the identification of a number of modes and mechanisms of action. This type of strategy has been considered previously (Nendza et al., 1995; Wenzel et al., 1997) and can be put together at low cost. The battery approach would provide useful information to assist in the definition of mechanistic domains, i.e. development of *in silico* knowledge.

5 *In silico* predictions in ITS for acute aquatic toxicity – utilising mechanistic information

The key to using mechanism and mode of action information is to form consensus from whatever information may be available. For instance, a possible strategy may be to first consider the evidence of whether a compound is narcotic or non-narcotic. Whilst on first sight this may appear to be trivial and of little utility, it is actually a powerful piece of information. Should a compound be determined as being narcotic, then it can be assigned to a particular mechanism of action and a QSAR and/or inter-species extrapolation can be utilised to make a prediction with high confidence (Dimitrov et al., 2000, 2003b). Should a compound be non-narcotic, it will both indicate that predictions from QSARs will have less confidence associated with them and that further information may be required to make a decision, e.g. non-chemical test or receptor binding information.

The decision of whether a compound is narcotic or not can be derived from a number of information sources. These include the structural fragment-based approaches described above (Verhaar, Russom, OASIS) and implemented in freely available software. In addition, the formation of categories of chemicals will allow for read-across of mechanisms of action. This can be supplemented by *in vitro* information, e.g. whether the compound has toxicity above baseline narcosis. If toxicity is seen above a baseline, then it is indicative of a mechanism of action other than narcosis (Maeder et al., 2004). If cytotoxicity is consistent with baseline, then the compound may be either a true narcotic, or the cytotoxicity is not consistent with a mechanism in the higher species. In this case further evidence would be helpful either from *in silico* investigation or non-biological test information (allowing for inter-species effects if extrapolated to fish (Dimitrov et al., 2000)). If it is determined that a compound is non-narcotic, then QSAR predictions are generally made with lower confidence. Again, this may require further evidence to improve the accuracy of the assessment of toxicity.

The success of using of an ITS to predict acute aquatic toxicity will depend on a number factors including:

- Assigning a compound to a mode and/or mechanism of action with confidence.
- The availability of tools and procedures to assign a compound to a particular mode and/or mechanism of action.
- The linking of QSARs to a mode and/or mechanism of action.
- The ability to correlate lower or *in vitro* toxicity values to higher species.
- A regulatory framework whereby the assumptions and predictions of an ITS are acceptable; this should be supported by clear guidance and case studies.

6 Conclusions

Information on mechanisms and modes of toxicity can play an important role in increasing the confidence placed in the *in silico* predictions of acute and chronic ecotoxicity. There are a number of methods to assign a compound to a mechanism of action, either through the use of structural rules or category formation. This type of information can be supplemented through the use of results from *in vitro* tests and existing *in vivo* data. The information derived for mode and mechanism of action plays an important role in assisting in the derivation of integrated testing strategies for the effects of chemicals.

References

Ahlers, J., Riedhammer, C., Vogliano, M. et al. (2006). Acute to chronic ratios in aquatic toxicity – Variation across trophic levels and relationship with chemical structure. *Environ. Toxicol. Chem. 25*, 2937-2945.

Aladjov, H., Todorov, M., Schmieder, P. et al. (2009). Strategic selection of chemicals for testing. Part I. Functionalities and performance of basic selection methods. *SAR QSAR Environ. 20*, 159-183.

Aptula, A. O., Netzeva, T. I., Valkova, I. V. et al. (2002). Multivariate discrimination between modes of toxic action of phenols. *Quant. Struct-Act. Relat. 21*, 12-22.

Bradbury, S. P., Russom, C. L., Ankley, G. T. et al. (2003). Overview of data and conceptual approaches for derivation of quantitative structure-activity relationships for ecotoxicological effects of organic chemicals. *Environ. Toxicol. Chem. 22*, 1789-1798.

Cronin, M. T. D. (2006). The role of hydrophobicity in toxicity prediction. *Curr. Comp. Aided Drug Des. 2*, 405-413.

Dimitrov, S., Koleva, Y., Schultz, T. W. et al. (2004). Interspecies quantitative structure-activity relationship model for aldehydes: Aquatic toxicity. Environ. *Toxicol. Chem. 23*, 463-470.

Dimitrov, S., Koleva, Y., Lewis, M. et al. (2003a). Modeling mode of action of industrial chemicals: Application using chemicals on Canada's Domestic Substances List (DSL). *QSAR Comb. Sci. 22*, 5-17.

Dimitrov, S. D., Mekenyan, O. G., Sinks, G. D. and Schultz, T. W. (2003b). Global modeling of narcotic chemicals: ciliate and fish toxicity. *J. Mol. Struct.-Theochem. 622*, 63-70.

Dimitrov, S. D., Mekenyan, O. G. and Schultz, T. W. (2000). Interspecies modeling of narcotics toxicity to aquatic animals. *Bull. Env. Contam. Toxicol. 65*, 399-406.

Enoch, S. J., Hewitt, M., Cronin, M. T. D. et al. (2008). Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere 73*, 243-248.

Gerberick, F., Aleksic, M., Basketter, D. et al. (2008). Chemical reactivity measurement and the predictive identification of skin sensitisers. *ATLA 36*, 215-242.

Grindon, C., Combes, R., Cronin, M. T. D. et al. (2008a). Integrated decision tree testing strategies for environmental toxicity with respect to the requirements of the EU REACH legislation. *ATLA 36*, 29-42.

Grindon, C., Combes, R., Cronin, M. T. D. et al. (2008b). Integrated testing strategies for use with respect to the requirements of the EU REACH legislation. *ATLA 36*, 7-27.

Hodges, G., Roberts, D. W., Marshall, S. J. and Dearden, J. C. (2006). Defining the toxic mode of action of esther sulphonates using the joint toxicity of mixtures. *Chemosphere 64*, 17-25.

Könemann, H. (1981). Quantitative structure-activity relationships in fish toxicity studies. 1. Relationships for 50 industrial pollutants. *Toxicology 19*, 209-221.

Madden, J. C., Enoch, S. J., Hewitt, M. and Cronin, M. T. D. (2009). Pharmaceuticals in the environment: Good practice in predicting acute ecotoxicological effects. *Toxicol. Lett. 185*, 85-101.

Maeder, V., Escher, B. I., Scheringer, M. and Hungerbuhler, K. (2004). Toxic ratio as an indicator of the intrinsic toxicity in the assessment of persistent, bioaccumulative, and toxic chemicals. *Environ. Sci. Technol. 38*, 3659-3666.

McKim, J. M., Bradbury, S. P. and Niemi, G. J. (1987a). Fish acute toxicity syndromes and their use in the QSAR approach to hazard assessment. *Environ. Health Persp. 71*, 171-186.

McKim, J. M., Schmieder, P. K., Carlson, R. W. et al. (1987b). Use of respiratory-cardiovascular responses of rainbow trout (Salmo gairdneri) in identifying acute toxicity syndromes in fish. 1. Pentachlorophenol, 2,4-dinitrophenol, tricaine methanesulfonate and 1-octanol. *Environ. Toxicol. Chem. 6*, 295-312.

McKim, J. M., Schmieder, P. K., Niemi, G. J. et al. (1987c). Use of respiratory-cardiovascular responses of rainbow trout (Salmo gairdneri) in identifying acute toxicity syndromes in fish. 2. Malathion, carbaryl, acrolein and benzaldehyde. *Environ. Toxicol. Chem. 6*, 313-328.

Nendza, M. and Muller, M. (2001). Discriminating toxicant classes by mode of action: 2. Physico-chemical descriptors. *Quant. Struct-Act. Relat. 19*, 581-598.

Nendza, M. and Muller, M. (2007). Discriminating toxicant classes by mode of action: 3. Substructure indicators. SAR *QSAR Environ. 18*, 159-168.

Nendza, M. and Wenzel, A. (2006). Discriminating toxicant classes by mode of action: 1. (Eco)toxicity profiles. *Environ. Sci. Pollut. Res. 13*, 192-203.

Nendza, M., Wenzel, A. and Wienen, G. (1995). Classification of contaminants by mode of action based on in vitro assays. *SAR QSAR Environ. 4*, 39-50.

Pavan, M. and Worth, A. P. (2008). Publicly-accessible QSAR software tools developed by the Joint Research Centre. *SAR QSAR Environ. 19*, 785-799.

Russom, C. L., Breton, R. L., Walker, J. D. and Bradbury, S. P. (2003). An overview of the use of quantitative structure-activity relationships for ranking and prioritizing large chemical inventories for environmental risk assessments. *Environ. Toxicol. Chem. 22*, 1810-1821.

Russom, C. L., Bradbury, S. P., Broderius, S. J. et al. (1997). Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem. 16*, 948-967.

Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P. et al. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. *Reg. Toxicol. Pharmacol. 53*, 70-80.

Schultz, T. W. (1997). Tetratox: Tetrahymena pyriformis population growth impairment endpoint – a surrogate for fish lethality. *Toxicol. Meth. 7*, 289-309

Seward, J. R., Sinks, G. D. and Schultz, T. W. (2001). Reproducibility of toxicity across mode of toxic action in the Tet-

rahymena population growth impairment assay. *Aquat. Toxicol. 53*, 33-47.

Shi, L. M., Fang, H., Tong, W. D. et al. (2001). QSAR models using a large diverse set of estrogens. *J. Chem. Infor. Comput. Sci. 41*, 186-195.

Spycher, S., Nendza, M. and Gasteiger, J. (2004). Comparison of different classification methods applied to a mode of toxic action data set. *QSAR Comb. Sci. 23*, 779-791.

van Leeuwen, K., Schultz, T. W., Henry, T. et al. (2009). Using chemical categories to fill data gaps in hazard assessment. *SAR QSAR Environ. Res. 20*, 207-220.

Verhaar, H. J. M., van Leeuwen, C. J. and Hermens, J. L. M. (1992). Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere 25*, 471-491.

Wenzel, A., Nendza, M., Hartmann, P. and Kanne, R. (1997). Test battery for the assessment of aquatic toxicity. *Chemosphere 35*, 307-322.

Acknowledgements

This project was sponsored by Defra through the Sustainable Arable Link Programme. The funding of the European Union 6th Framework OSIRIS Integrated Project (GOCE-037017-OSIRIS) and InSilicoTox Marie Curie Project (MTKD-CT-2006-42328) is also gratefully acknowledged.

Correspondence to

Mark Cronin, Ph.D.
School of Pharmacy and Chemistry
Liverpool John Moores University
Byrom Street
Liverpool
L3 3AF, UK
e-mail: m.t.cronin@ljmu.ac.uk

BS28: Animal welfare science

When do Mammalian Young Become Sentient?

David J. Mellor [1,2], Tamara J. Diesch [1] and Craig B. Johnson [1,3]

[1]Animal Welfare Science and Bioethics Centre, [2]Institute of Food, Nutrition and Human Health and [3]Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand

Summary

Published literature and studies from our laboratory show that neurological development in mammalian young at birth ranged from being exceptionally immature (e.g. newborn marsupials), through moderately immature (e.g. newborn cats, dogs, mice, rabbits, rats) to mature (newborn cattle, deer, goats, sheep, horses, pigs, guinea-pigs). In all cases, brain electrical activity indicates that under normal circumstances none of these young exhibit consciousness before birth. This is the case with exceptionally and moderately immature young because their neurological development is not adequate to support consciousness until several months and several days, respectively, after birth. Neurologically mature newborns do have the neurological capacity for consciousness before birth, but this is usually prevented by the operation of a number of neuroinhibitory mechanisms that are unique to fetal life. After birth these newborns exhibit consciousness within minutes or hours. The evidence for this and the implications for safeguarding the welfare of fetuses and newborns of these different mammals in experimental settings are discussed.

Keywords: fetus, newborn, neurological development, unconsciousness

1 Introduction

Animals considered to be sentient have the ability to experience positive and negative affective states or feelings (Duncan, 2006). These feelings arise via the operation of sensory mechanisms that scan the internal functional state of the animal and its external environment. *Internally-focused* sensory activity contributes to the brain processing that gives rise to a number of experiences including thirst, hunger, breathlessness, pain, nausea, sickness and malaise, and informs the brain when the conditions underlying those feelings are absent; at the same time, *externally stimulated* neural inputs contribute to affective experiences mediated by the sensory modalities of sight, hearing, smell, taste, touch, thermal comfort and others (Dawkins, 2006; Kirkwood, 2006; Mellor et al., 2009a). Although it is philosophically and scientifically problematic to demonstrate unequivocally that higher-order non-human animals do in fact experience such feelings (Dawkins, 2006; Duncan, 2006), for most practical purposes we interact with them as if they do (Dawkins, 2006; Mellor et al., 2009a). Moreover, a presumption of the existence of sentience among higher-order animals generates strong motivation to treat them humanely. This is especially so now that the welfare status of animals is widely understood to reflect the integrated balance between all sensory inputs to the animal's brain that are cognitively processed and emotionally experienced as feelings (Mellor et al., 2009a).

Sentience, therefore, requires the presence of a nervous system that is functionally sophisticated enough to enable the animal to cognitively and emotionally experience its status (its welfare) as good, neutral or bad. This capacity is cautiously attributed to most vertebrates and some invertebrates (Kirkwood and Hubrecht, 2001; Duncan, 2006; Kirkwood, 2006), but perhaps especially to mammals and birds. Allied to sentience, of course, is the requirement that the nervous system be in a state of consciousness, because unconscious animals cannot experience anything (Baars, 2001; Mellor and Diesch, 2006; Boverroux et al., 2008).

In addition to this linkage between the phylogenetic status vis-à-vis brain sophistication of a species and its capacity for sentience and consciousness, it is evident that within those spe-

cies that exhibit these attributes as adults, the first appearance of the capacity for sentience and consciousness also depends on the pattern of neurological development during the early stages of each animal's life cycle (Mellor and Diesch, 2006, 2007). The developmental stages when young of different species first exhibit this capacity are therefore of interest, especially when considering safeguards to be applied during their use in scientific and other procedures that would have negative welfare impacts were they conscious adults. Resolution of this question bears directly on when and how necessary, or safe, it is to employ *refinement strategies* such as the use of anaesthetics or analgesics in developing fetal, newborn and young mammals. We have drawn attention to these issues elsewhere (Diesch et al., 2008, 2009; Mellor et al., 2008). It also bears on the legal definition of an animal framed in animal welfare terms, as discussed elsewhere in these proceedings (Mellor et al., 2010).

The present paper updates understanding of the science underlying the conclusion that consciousness, and therefore the capacity to experience noxious sensations, does not appear until after birth in any mammal examined to date, and that the timing of the onset of consciousness after birth depends on the degree of neurological maturity of the newborn in each species (Mellor et al., 2009b). For the sake of brevity, review articles that fully reference the relevant scientific literature have been used here together with pertinent additional publications.

2 The developmental pattern of brain electrical activity is similar across species

Brain electrical activity, as reflected in the electroencephalogram (EEG), shows distinct patterns during waking consciousness and during the unconsciousness of sleep, general anaesthesia, coma and epileptic "states of absence" (Baars, 2001; Bovereau et al., 2008). On this basis the first appearance of spontaneous EEG activity and the different EEG patterns observed subsequently have been used cautiously to infer when the brain exhibits states of unconsciousness and when consciousness may first appear during neurological development (Mellor and Gregory, 2003; Lee et al., 2005; Mellor et al., 2005, 2009b).

The development of EEG activity appears to follow a similar pattern in all mammalian young examined to date (Ellingson and Rose, 1970; Walker and Berger, 1978; Snead and Stephens, 1983; Mellor et al., 2005; Brusseau and Mashour, 2007). For descriptive convenience this pattern can be divided into six stages categorised as follows:

Stage 1: the EEG is absent, isoelectric or "silent";
Stage 2: a predominance of electrical "silence" is punctuated by very short epochs of low voltage EEG activity;
Stage 3: prolonged periods of electrical "silence" are punctuated by epochs of more sustained low voltage EEG activity that progressively increase in duration;
Stage 4: continuous "mixed" EEG activity containing elements of both rapid-eye-movement (REM) and non-REM sleep-like patterns is present;
Stage 5: continuous EEG activity is differentiated into distinct REM and non-REM patterns with regular cycling between the two states;
Stage 6: sleep-wake cycles, where the EEG and behaviour indicate that periods of sleep alternate with periods of conscious wakefulness, are present.

Progression through these stages is presumed to reflect increasing degrees of functional and structural maturation of the brain (Mellor et al., 2005; Brusseau and Mashour, 2007; Diesch et al., 2009) and is paralleled by increasingly precise control and coordination of movements of the trunk, head, limbs and other body parts (Bolles and Woods, 1964; Tyndale-Biscoe and Janssens, 1988; Mellor and Gregory, 2003). Nevertheless, the characteristics of the EEG suggest that the brain remains in unconscious states throughout stages 1 to 5. Moreover, there is a temporal association between differentiation of REM-non-REM EEG patterns at stage 5 and the establishment of neural connections between the cerebral cortex and underlying brain structures including the thalamus (Ellingson and Rose, 1970; Lee et al., 2005; Mellor et al., 2005; Brusseau and Mashour, 2007). In light of the important functional roles attributed to corticothalamic connections in physiologically, pharmacologically and pathologically induced transitions between consciousness and different states of unconsciousness (Evans, 2003; Bovereau et al., 2008) and evidence for the absence of consciousness in prematurely born infants before the establishment of these neural connections (Lee et al., 2005; Mellor et al., 2005), their appearance during brain development is considered to be a prerequisite for the capacity of young animals to exhibit consciousness (Mellor et al., 2009b). According to this reasoning, therefore, unconscious states would persist until the end of the period of differentiated REM-non-REM patterns at stage 5.

3 Neurological maturity of young at birth varies widely between species

Although the general pattern of neurological development, judged in these terms, appears to be common to many mammals, the stage when birth occurs during this developmental path depends on the species and obviously determines the neurological maturity of the young at that time (Ellingson and Rose, 1970; Tyndale-Biscoe and Janssens, 1988; Mellor et al., 2005; Mellor and Diesch, 2006). This may be illustrated by comparing mammalian young that are neurologically exceptionally immature, moderately immature or mature at birth.

3.1 Neurologically exceptionally immature young

Newborn marsupial joeys fall into this category and, as indicated below, would have electrically "silent" EEGs (stage 1). On this basis, joeys would have no capacity to consciously experience any sensations before, during and for some period after birth.

The tammar wallaby joey (*Macropus eugenii*), which is born after a 28-day pregnancy, has a cerebral cortex that consists of only two layers of cells and resembles that of a 40-

day human or 26-day sheep embryo (Reynolds et al., 1985). After entering the pouch the individual joey remains continuously attached to a teat for about 100 days, and intermittently thereafter, leaving the pouch permanently after approximately 250 days (Tyndale-Biscoe and Janssens, 1988). The EEG of anaesthetised joeys remains isoelectric (stage 1) until about 120 days, after which it develops progressively (through stages 2 and 3) to become continuous by 150-160 days (stage 4) (Diesch et al., 2008, 2010). Anaesthesia precludes observation of differentiated REM-non-REM EEG patterns (stage 5) and also sleep-wake EEG cycles (stage 6). However, opening of the external ear canals at 125-130 days and eye-opening at about 140 days (Hill et al., 1998), and the presence of behavioural signs such as the joey standing unaided by about 160 days, beginning to look out of the pouch at about 180 days and first exiting the pouch at about 190 days (Tyndale-Biscoe and Janssens, 1988), together suggest that the capacity for conscious perception of auditory, visual, tactile, proprioceptive and other sensations would appear in tammar joeys at around 160-180 days. Finally, EEG responses of anaesthetised joeys to noxious stimulation (toe clamping) are consistent with this suggestion, as there was no response at 94-127 days, a minimal response at 142-181 days and a marked response at 187-261 days (Diesch et al., 2010).

Virginia opossum joeys (*Didelphis virginiana*), born after a 13-day pregnancy, enter the pouch where up to 13 of them attach to teats (McManus, 1974). They remain attached to a teat continuously for 55-65 days, then intermittently until 70-90 days, after which they leave the pouch and ride on the mother's back until weaned at 100-110 days (McManus, 1974; Krause and Saunders, 1994; Darlington et al., 1999). The EEG of non-anaesthetised opossum joeys remains mostly electrically "silent" (stage 1) before about 60 days, after which continuous low voltage activity begins to appear (progressing from stage 2 to 4), with early signs of non-REM activity evident by about 65 days and REM-non-REM differentiation (stage 5) by about 75 days (Walker and Berger, 1978). The behaviour of joeys suggests that sleep-wake EEG cycles (stage 6) would appear soon after this. Overall, these behavioural and EEG observations together suggest that opossum joeys would have a well developed capacity for conscious perception around 70-90 days after birth, presumably with the earliest evidence of consciousness occurring before this.

It follows that the earliest postnatal appearance of sentience and consciousness, estimated by reference to EEG patterns, behaviour and/or responses to noxious stimulation, is likely at about 4-5 months in tammar wallaby joeys and about 2 months in Virginia opossum joeys.

3.2 Neurologically moderately immature young

Newborn rat and mouse pups, rabbit kits, kittens and puppies fall into this category. Immediately after birth they exhibit isoelectric, intermittent or continuous but undifferentiated EEGs (stages 1-4), and thereafter it takes several days for REM-non-REM differentiation (stage 5) to occur and sleep-wake cycles (stage 6) to become established (Ellingson and Rose, 1970; Jouvet-Monier et al., 1970; Snead and Stephens, 1983; Daszuta and Gambarelli, 1985).

In rat pups, for instance, EEG stages 2-4 are evident at 5-7 days after birth, the transitions to stage 5, and then to stage 6, begin around 10-12 days and are complete by about 18 days, and adult EEG patterns appear around 18-20 days after birth (Diesch et al., 2009). Pup behaviour develops over the same period (Bolles and Woods, 1964). Pups aged 5-7 days huddle with littermates, suck from their mother, crawl, stretch, yawn and show rudimentary grooming. At 9-10 days, walking, cage exploration and orientation away from the mother and nest appear, and by days 12-13 these activities become more vigorous or prolonged. Increased attention to objects follows eye opening at about 14 days, with the first obvious awareness of the investigator occurring at about 19 days. At 17-20 days, pups are often physically active with fighting between littermates common. Finally, when pups are weaned at 21 days by removal of the mother their physical activity decreases for about one day. These observations together suggest that conscious perception by rat pups is not likely before about 10-12 days of age. EEG responses of anaesthetised pups to noxious stimulation (tail clamping) are consistent with this suggestion, as there was no response at 5-7 days, a marginal response at 12-14 days and a marked response at 21-22 days (Diesch et al., 2009).

If it is conservatively assumed that conscious perception cannot occur any earlier than the postnatal age when stage 5 REM-non-REM differentiation *first* appears, bearing in mind that stage 5 EEGs indicate physiological states of unconsciousness (Baars, 2001; Bovereau et al., 2008), then the published evidence suggests that conscious experience of sensory inputs is likely to occur postnatally only after about 4 days in rabbit kits, 7 days in kittens, 10-12 days in rat and mouse pups, and, perhaps depending on the breed, 4-14 days in puppies (Ellingson and Rose, 1970; Jouvet-Monier et al., 1970; Snead and Stephens, 1983; Daszuta and Gambarelli, 1985).

3.3 Neurologically mature young

Newborn calves, fawns, kids, lambs, foals, piglets and guinea-pig pups, and indeed human infants, are included in this category on the basis of neuroanatomical, neurophysiological and behavioural evidence (Ellingson and Rose, 1970; Jouvet-Monier et al., 1970; Mellor and Gregory, 2003; Mellor and Stafford, 2004; Lee et al., 2005; Mellor et al., 2005; Mellor and Diesch, 2006; Brusseau and Mashour, 2007). They pass through the first five EEG stages before birth, reaching stage 5 (REM-non-REM pattern differentiation) about 75-80% of the way through pregnancy at around the time the critical cortico-thalamic connections are established. As noted above, human infants born prematurely after about 30 weeks of the 40-week pregnancy exhibit increasing behavioural and EEG evidence of conscious perception (Lee et al., 2005), yet in fetal lambs before or during labour, whether premature or not, there are no signs of stage 6 sleep-wake EEG cycles (Rigatto et al., 1986; Mellor et al., 2005; Mellor and Diesch, 2006). Moreover, noxious stimuli that arouse sleeping newborns to conscious wakefulness are not similarly effective in fetuses during stage 4 or 5

sleep-like states of unconsciousness (Mellor et al., 2005). Evidently, therefore, neurologically mature fetuses develop the capacity for sentience and consciousness during late pregnancy, but remain in unconscious states throughout. This is attributed to a unique set of neuroinhibitory factors, with demonstrated effects on the fetal EEG, that function *in utero* to maintain fetal unconsciousness (Mellor et al., 2005; Mellor and Diesch, 2006, 2007). These neuroinhibitors, whose actions have been described in detail elsewhere (Mellor et al., 2005), include adenosine (a potent promoter of sleep and/or unconsciousness), allopregnanolone and pregnanolone (neuroactive steroids with well-established anaesthetic, sedative/hypnotic and analgesic effects, synthesised by the fetal brain), prostaglandin D_2 (a potent sleep-inducing agent synthesised by the fetal brain) and a placental peptide neuroinhibitor, as well as warmth, cushioned tactile stimulation and buoyancy.

Maintaining the fetus in unconscious states may help to ensure an adequate supply of oxygen to its brain. The fetus has no control over the restricted supply of oxygen it receives across the placenta (Mellor and Diesch, 2007), yet its brain in particular is vulnerable to oxygen shortages (Hunter et al., 2003a). Limiting its cerebral oxygen consumption would therefore be beneficial. In adult humans, high cerebral oxygen consumption during consciousness is reduced during sleep, being about 40% and 10% lower during non-REM and REM sleep, respectively (Bovereaux et al., 2008). It follows that the observed presence of sleep-like states in the fetus (Mellor et al., 2005) may reduce its brain requirements for oxygen. There is evidence from fetal sheep to support this: the mean cerebral oxygen consumption is about 80% lower in immature than in mature fetuses (Gleason et al., 1989), and in mature fetuses the mean value is about 40% lower than in the newborn and about 25% lower than in the adult (Rosenberg et al., 1982); also, the mean consumption is about 17% lower during non-REM than REM sleep-like states in mature fetuses (Richardson et al., 1985). Furthermore, oxygen-sparing changes occur in the fetus in apparent preparation for labour, during which oxygen supply may become compromised by uterine contractions. These changes include an increasing incidence of the non-REM EEG state and fetal motor systems becoming largely quiescent (see Mellor and Diesch, 2006). Finally, there is an emergency response mechanism. It ensures that complete occlusion of the umbilical cord during individual labour contractions would lead to an isoelectric EEG within 60-90 seconds (Mallard et al., 1992; Bennet et al., 1999; Hunter et al., 2003b) and to an associated reduction in cerebrocortical oxygen consumption by at least 95% (Hunter et al., 2003a). Although both the EEG and cerebral oxygen consumption would usually return to normal once the contraction-induced cord occlusion has passed, protracted marked oxygen shortages lead to brain damage (Hunter et al., 2003b).

It appears, therefore, that although neurologically mature young develop the capacity for sentience before birth, they remain unconscious while *in utero*, and consciousness first appears only after birth. This occurs through an immediate or progressive postnatal withdrawal of the *in utero* neuroinhibitors noted above, combined with a progressive onset or marked surge in stimulation from a set of potent neuroactivators that includes 17β-oestradiol, noradrenaline, cold exposure, contact with hard surfaces, and auditory, visual and other sensory inputs. Detailed descriptions of the proposed mechanisms have been provided elsewhere (Mellor and Gregory, 2003; Mellor and Diesch, 2006; Mellor et al., 2009b). Critical to the appearance of consciousness after birth, however, is the rapid and successful onset of breathing. This oxygenates brain tissue to well above the highest levels ever achieved before birth, and this, in its turn, is considered to remove an overriding adenosine-induced inhibition of cerebrocortical activity, thereby "permitting" the neuroactivators to act. The result, as assessed behaviourally, is the onset of consciousness over minutes or hours (Mellor and Stafford, 2004; Mellor et al., 2009b).

Interestingly, during the first few days after birth, a persistent, but reduced and decreasing brain synthesis of the previously mentioned "fetal" neuroactive steroids having anaesthetic, sedative/hypnotic and analgesic actions (Mellor and Diesch, 2006; T. J. Diesch, D. J. Mellor, C. B. Johnson and D. W. Walker, unpublished observations) may reduce the newborn's experience of pain due to compression-induced or other birth injuries. Any such analgesic effects are likely to be highest immediately after birth and then wane as the circulating concentrations of these steroids decrease. Consistent with this are lower EEG responses to noxious stimulation (castration) in anaesthetised lambs during the first 1-3 days after birth compared to those of older lambs (Johnson et al., 2009).

4 Concluding remarks

The observations made above are fresh insights derived from recent integrative syntheses of well-demonstrated, yet not well-known, findings in the scientific literature. Although the literature provides a compelling case for persistent unconsciousness before birth, even in neurologically mature fetuses, this proposition is contrary to views held by some biomedical scientists working in the field of fetal pain management (e.g. Anand, 2006, 2007) and contradicts the speculation that brain stem processes are capable of supporting some forms of consciousness (Merker, 2007). Accordingly, a major purpose of presenting this alternative view, together with a summary of the supporting scientific evidence, is to stimulate others to challenge them experimentally. If the present conclusions survive such challenges, they may be adopted more widely. Regardless of the outcome, the evidence presented already exists in the scientific literature and needs to be assimilated into our understanding of developmental processes in mammalian young before and after birth.

References

Anand, K. J. S. (2006). Fetal pain? Pain: *Clin. Updates XIV, No 2*, 1-4.

Anand, K. J. S. (2007). Consciousness, cortical function, and pain perception in nonverbal humans. *Behav. Brain Sci. 30*, 82-83.

Baars, B. J. (2001). There are no known differences in brain mechanisms of consciousness between humans and other mammals. *Anim. Welfare 10 supplement*, S31-S40.

Bennet, L., Rossenrode, S., Gunning, M. I. et al. (1999). The cardiovascular and cerebrovascular responses of the immature fetal sheep to acute umbilical cord occlusion. *J. Physiol. 517*, 247-257.

Bolles, R. C. and Woods, P. J. (1964). The ontogeny of behaviour in the albino rat. *Anim. Behav. 12*, 427-441.

Boveroux, P., Bonhomme, V., Boly, M. et al. (2008). Brain function in physiologically, pharmacologically and pathologically altered states of consciousness. *Int. Anesthesiol. Clin. 46*, 131-146.

Brusseau, R. R. and Mashour, G. A. (2007). Subcortical consciousness: Implications for fetal anesthesia and analgesia. *Behav. Brain Sci. 30*, 86-87.

Darlington, R. B., Dunlop, S. A. and Finlay, B. L. (1999). Neural development in metatherian and eutherian mammals: Variation and constraint. *J. Comp. Neurol. 411*, 359-368.

Daszuta, A. and Gambarelli, F. (1985). Early postnatal development of EEG and sleep-waking cycle in two inbred mouse strains. *Dev. Brain. Res. 22*, 39-47.

Dawkins, M. S. (2006). Through animal eyes: What behaviour tells us. *Behav. Sci. 100*, 4-10.

Diesch, T. J., Mellor, D. J., Johnson, C. B. and Lentle, R. G. (2010). Developmental changes in the electroencephalogram and responses to a noxious stimulus in anaesthetized tammar wallaby joeys (Macropus eugenii eugenii). *Lab. Anim. 44(2),* 79-87.

Diesch, T. J., Mellor, D. J., Johnson, C. B. and Lentle, R. G. (2009). Electroencephalographic responses to tail clamping in anaesthetised rat pups. *Lab. Anim. 43*, 224-231.

Diesch, T. J., Mellor, D. J., Johnson, C. B. and Lentle, R. G. (2008). Responsiveness to painful stimuli in anaesthetised newborn and young animals of varying neurological maturity (wallaby joeys, rat pups and lambs). *AATEX J. 14, Special Issue*, 549-552.

Duncan, I. J. H. (2006). The changing concept of animal sentience. *Appl. Anim. Behav. Sci. 100*, 11-19.

Ellingson, R. J. and Rose, G. H. (1970). Ontogenesis of the electroencephalogram. In W. A. Himwich (ed), *Developmental Neurobiology* (441-471). Springfield: Charles C Thomas.

Evans, B. M. (2003). Sleep, consciousness and the spontaneous and evoked electrical activity of the brain. Is there a cortical integrating mechanism? *Clin. Neurophysiol. 33*, 1-10.

Gleason, C. A., Hamm, C. and Jones, M. D. (1989). Cerebral blood flow, oxygenation, and carbohydrate metabolism in immature fetal sheep in utero. *Am. J. Physiol. 256 (Regulatory Integrative, Comp. Physiol. 25),* R1264-R1268.

Hill, K. G., Cone-Wesson, B. and Liu, G. B. (1998). Development of auditory function in the tammar wallaby Macropus eugenii. *Hearing Res. 117*, 97-106.

Hunter, C. J., Blood, A. B. and Power, G. G. (2003a). Cerebral metabolism during cord occlusion and hypoxia in fetal sheep: a novel method of continuous measurement of basal heat production. *J. Physiol. 552*, 241-251.

Hunter, C. J., Bennett, L., Power, G. G. et al. (2003b). Key neuroprotective role for endogenous adenosine A1 receptor activation during asphyxia in the fetal sheep. *Stroke 34*, 2240-2245.

Johnson, C. B., Sylvester, S. P., Stafford, K. J. et al. (2009). Effects of age on the electroencephalographic response to castration in lambs anaesthetised using halothane in oxygen from birth to 6 weeks old. *Vet. Anaesth. Analg. 36*, 273-279.

Jouvet-Mounier, D., Astic, L. and Lacote, D. (1970). Ontogenesis of the states of sleep in rat, cat, and guinea pig during the first postnatal month. *Dev. Psychobiol. 2*, 216-239.

Kirkwood, J. K. (2006) The distribution of the capacity for sentience in the animal kingdom. In J. Turner and J. D'Silva (eds), *Animals, Ethics and Trade: The Challenge of Animal Sentience* (12-26). London: Earthscan.

Kirkwood, J. K. and Hubrecht, R. (2001). Animal consciousness, cognition and welfare. *Anim. Welfare 10 Suppl.*, S5-S17.

Krause, W. J. and Saunders, N. R. (1994). Brain growth and neocortical development in the opossum. *Ann. Anat. 176*, 395-407.

Lee, S. J., Peter Ralston, H. J., Drey, E. A. et al. (2005). Fetal pain: a systematic multidisciplinary review of the evidence. *Journal of the American Medical Association 294*, 947-954.

Mallard, E. C., Gunn, A. J., Williams, C. E. et al. (1992). Transient umbilical cord occlusion causes hippocampal damage in the fetal sheep. *Am. J. Obstet. Gynecol. 167*, 1423-1430.

McManus, J. J. (1974). Didelphis virginiana. Mamm. *Species 40*, 1-6.

Mellor, D. J., Diesch, T. J. and Johnson, C. B. (2010). Should mammalian fetuses be excluded from regulations protecting animals during experiments? *ALTEX*, present proceedings.

Mellor, D. J., Patterson-Kane, E. and Stafford, K. J. (2009a). Animal welfare, grading compromise and mitigating suffering. In *The Sciences of Animal Welfare* (71-94). Oxford: Wiley-Blackwell Publishing.

Mellor, D. J., Patterson-Kane, E. and Stafford, K. J. (2009b). Integrated perspectives: sleep, developmental stage and animal welfare. In *The Sciences of Animal Welfare* (161-185). Oxford: Wiley-Blackwell Publishing.

Mellor, D. J., Diesch, T. J., Gunn, A. J. and Bennet, L. (2008). Fetal "awareness" and "pain": what precautions should be taken to safeguard fetal welfare during experiments? *AATEX J. 14, Special Issue*, 79-83.

Mellor, D. J. and Diesch, T. J. (2007). Birth and hatching: key events in the onset of "awareness" in lambs and chicks. *New Zeal. Vet. J. 55*, 51-60.

Mellor, D. J. and Diesch, T. J. (2006). Onset of sentience: The potential for suffering in fetal and newborn farm animals. *Appl. Anim. Behav. Sci. 100*, 48-57.

Mellor, D. J., Diesch, T. J., Gunn, A. J. and Bennet, L. (2005). The importance of "awareness" for understanding fetal pain. *Brain Res. Rev. 49*, 455-471.

Mellor, D. J. and Stafford, K. J. (2004). Animal welfare implications of neonatal mortality and morbidity in farm animals. *Vet. J. 168*, 118-133.

Mellor, D. J. and Gregory, N. G. (2003). Responsiveness, behavioural arousal and awareness in fetal and newborn lambs:

experimental, practical and therapeutic implications. *New Zeal. Vet. J. 51*, 2-13.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behav. Brain Sci. 30*, 63-81.

Reynolds, M. L., Cavanagh, M. E., Dziegielewska, K. M. et al. (1985). Postnatal development of the telencephalon of the tammar wallaby (Macropus eugenii). *Anat. Embryol. 173*, 81-94.

Richardson, B. S., Patrick, J. E. and Abduljabber, H. (1985). Cerebral oxidative metabolism in the fetal lamb: Relationship to electrocortical state. *Am. J. Obstet. Gynecol. 153*, 426-431.

Rigatto, H., Moore, M. and Cates, D. (1986). Fetal breathing and behavior measured through a double-wall Plexiglas window in sheep. *J. Appl. Physiol. 61*, 160-164.

Rosenberg, A. A., Jones, M. D., Traystman, R. J. et al. (1982). Response of cerebral blood flow to changes in Pco2 in fetal, newborn, and adult sheep. *Am. J. Physiol. 242 (Heart 11)*, H862-H866.

Snead, O. C. and Stephens, H. I. (1983). Ontogeny of cortical and subcortical electroencephalographic events in unrestrained neonatal and infant rats. *Exp. Neurol. 82*, 249-269.

Tyndale-Biscoe, C. H. and Janssens, P. A. (1988). Introduction. In C. H. Tyndale-Biscoe and P. A. Janssens, *The Developing Marsupial: Models for Biomedical Research* (1-7). Berlin: Springer-Verlag.

Walker, J. M. and Berger, R. J. (1978). The ontogenesis of sleep states, thermogenesis, and thermoregulation in the Virginia opossum. *Dev. Psychobiol. 13*, 443-454.

Acknowledgements

We are grateful to many colleagues for their helpful input during the development of these ideas. We also thank the Agricultural and Marketing Research and Development Trust for providing PhD funding for one of us (TJD), as well as the Geoffrey Gardiner Foundation (Australia), MAF Biosecurity New Zealand, and the Animal Welfare Science and Bioethics Centre and the Riddet Institute at Massey University for providing financial support for our related research.

Correspondence to

Prof. David J. Mellor, PhD
Animal Welfare Science and Bioethics Centre, IFNHH
College of Sciences, Massey University
Palmerston North 4442, New Zealand
e-mail: D.J.Mellor@massey.ac.nz

International Perspectives on Animal Welfare: Initiatives to Harmonise Approaches when Animals are Used for Scientific Purposes

Margaret Rose

Prince of Wales Clinical School, University of NSW, Sydney, Australia

Summary

The notion of animal welfare is captive to the complexity of the relationship between human beings and other animals. Animal welfare can be seen as a judgement about both an animal's experiences as well as a benchmark that defines the actions and duties of human beings towards other animals; a judgement that is honed by cultural, social and personal values and beliefs. Thus legislation and government policies that refer to animal welfare will vary between countries reflecting cultural needs, values and expectations. Controversy over the use of animals in research is longstanding and marked by robust public debate. In these circumstances, many countries have enacted legislation seeking to promote the public good by recognising and seeking to balance the competing claims of animal welfare and the potential benefits to society from such research activities.

Although in the legislative context, notions of animal welfare may differ, research is not constrained by national borders; it is a global activity. Consequently, in recent years there have been a number of initiatives whereby common ground has been sought so as to develop a harmonised, international framework to promote the welfare of animals used in research. In these endeavours, Replacement, Reduction and Refinement have been the unifying principles.

This paper will provide an overview of these international initiatives. Through several case studies, the seminal role of animal welfare science to inform the development and implementation of such activities will be argued and potential barriers to achieving an international approach will be discussed.

Keywords: international policies, animal welfare science, 3Rs

1 Introduction

Through recent initiatives of organisations such as the World Organisation for Animal Health (OIE) and the Food and Agricultural Organisation of the United Nations (FAO), animal welfare is being recognised by governments as a global issue (OIE, 2005; FAO, 2009). Although these initiatives have been directed towards farm animal welfare (Fraser, 2008), they have wider implications to animal welfare issues in other spheres of activity. Importantly, in their statement of general animal welfare principles, the OIE highlights the importance of the principles of Replacement, Reduction and Refinement in promoting the welfare of animals used in science. Of note, the FAO has adopted the definition used by the OIE for "animal welfare", emphasising that this relates to the state of the animal and evidence as to how it is coping with the conditions in which it lives; an evidence-based approach where animal welfare science is integral to achieving outcomes.

In recent years we have seen increasing globalisation of many activities including the ways we undertake scientific research that has resulted in a significant shift in the scope and kind of international activities around the business of science. Whereas for many years it has been the practice to disseminate research findings through international publications and presentations to scientific societies, today we are seeing the globalisation of all aspects of research activities with the development of multinational research programs often promoted through inter-government agreements. Such developments provide opportunities previously unimagined but do present a number of logistic challenges, not the least when those research activities involve the use of animals.

The use of animals for scientific purposes, be that research, teaching or product testing, is one of the few areas where the conduct of and decisions about scientific activities are regulated. There is a long history of public controversy about such use of animals, albeit with differing levels and kinds of concerns in various countries. Governments have responded to these issues in a range of ways, so that today many countries have legislation that regulates such use of animals. Although there are common elements, we see different legislative approaches reflecting historical experiences, social needs and cultural and political influences. Consequently, these legislative differences can present barriers to achieving multi-national research programmes; varying requirements for the governance of such research activities

and differences in accepted standards of practice, including housing and care of animals, being issues frequently cited.

Government regulation also impacts on the use of animals in science particularly through the requirements to use animals in regulatory testing for the safety and efficacy of human and veterinary therapeutic goods and devices, for identifying risks and setting standards for the use of chemicals and hazardous substances and for assessing the impact of human activity on the environment. In these circumstances, from the perspective of animal welfare, there have been criticisms of unnecessary use of animals, especially in relation to duplication of tests and a failure to adopt alternative non-animal methods. Government regulations prescribe the kind of testing required often with different testing protocols for the same product in different countries. This, together with non-acceptance by governments of testing in another jurisdiction, has been seen as a major impediment to achieving animal welfare goals in this area. As will be discussed below, seeking a harmonised, international approach to the use of animals in these circumstances has been the focus of significant efforts by both government and non-government agencies in the past 10-15 years. Much has been achieved but more needs to be done.

Whilst the globalisation of science and the barriers presented by differing government regulations have been recent drivers to initiatives to harmonise approaches around the use of animals for scientific purposes, prior to this, animal welfare concerns have driven a number of activities directed towards international harmonisation of policies and guidelines concerning decisions as to if and how animals are used in these circumstances. The Principles of the 3Rs – Replacement, Reduction and Refinement – have been the cornerstone for these developments; undoubtedly, despite differences in government policies, the universal acceptance of these Principles, has been integral to the outcomes achieved and will continue to be so for the success of future endeavours.

This paper will describe recent initiatives to foster a harmonised, international approach to promoting the welfare of animals used for scientific purposes, including regulatory testing, and in discussing the challenges to the effective implementation of these initiatives, argue for the critical role of animal welfare science in achieving the goals sought.

2 International initiatives

At an international level, efforts to promote the welfare of animals used in science have sought to bring about a harmonisation of approaches in different countries through the identification of and agreement upon common principles. In this way, international agreements have been achieved in relation to the general principles governing the use of animals in science as well as to approaches to particular methods, including those used in regulatory testing, so as to reduce or minimise the impact on animal welfare. In this regard, the achievements of the Council for International Organisations of Medical Sciences (CIOMS), the International Council for Laboratory Animal Science (ICLAS) and the Organisation for Economic Cooperation and Development (OECD) and the agreement for International Cooperation on Alternative Test Methods (ICATM) are highlighted. However, it also should be noted that national policies and guidelines, such as those published by the National Institutes of Health (NIH), the Institute for Laboratory Animal Research (ILAR), the Canadian Council on Animal Care (CCAC), the Joint Working Party on Refinement (BVA/FRAME/RSPCA/UFAW) and Australia's National Health and Medical Research Council (NHMRC), have influence beyond national borders and have contributed to and facilitated the development and implementation of a harmonised, international approach.

2.1 Council for International Organisations of Medical Sciences (CIOMS)

The CIOMS is an international, non-government organisation representing the biomedical research community that for many years has fostered consideration of the ethical issues involved in these research activities; many of its publications and policies concern ethical issues associated with the use of human subjects.

In 1985, following extensive consultation over a two-year period involving both international and interdisciplinary representatives, the CIOMS published the *International Guiding Principles for Biomedical Research Involving Animals* (CIOMS, 1985). This document was developed to provide a conceptual and ethical framework that could be used irrespective of the regulatory systems in place in various countries or the policies of different scientific societies. Whilst strongly emphasising the importance of the use of animals for biomedical research and regulatory testing, these Principles argued that scientists should be ever mindful of their moral obligations towards the use of animals, should, as far as possible, prevent pain and discomfort and be ever alert to the possibility of achieving their scientific goals without using animals. In addition to a set of Basic Principles, special requirements for the acquisition, transport, housing, husbandry and monitoring of animals are noted. Although not stated explicitly, these principles reflect the stated goals of the 3Rs.

The international impact of this document has been significant. It was endorsed by the WHO Advisory Committee on Medical Research and has underpinned government policy in many countries, for example, the US Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research and Training, are closely aligned with the CIOMS document.

In 2008, it was agreed to revise this document. This revision is now under way under the joint auspices of CIOMS and ICLAS.

2.2 International Council for Laboratory Animal Science (ICLAS)

ICLAS is the one umbrella, international organisation with a primary focus on the use of animals in science; its aims include the promotion of collaboration and dissemination of information in laboratory animal science and the humane use of animals based on ethical principles and promotion of the 3Rs. Since its foundation, ICLAS has sponsored programs that promote the quality of scientific outcomes through the definition and monitoring of the genetic status and health quality of laboratory ani-

mals and the education and training of animal care personnel and scientists.

In 2004, ICLAS organised its first meeting on the Harmonization of Guidelines on the Use of Animals in Science. Recognising the need for various countries to maintain their existing mechanisms for the oversight of animal research, but cognisant of the globalisation of research activities, the goal for ICLAS has been to develop guidance documents that reflected a harmonised approach to animal care and use. ICLAS has stated unequivocally that it does not intend such documents to be used as to standards.

ICLAS has sustained this initiative through its Working Group on Harmonisation, which to date has prepared four international guidelines with a fifth in preparation; all are based on a set of agreed general principles drawn from existing documents. In so doing key documents are cited and recognised as international references. For example, the first two documents set out principles for the establishment of humane endpoints and for animal euthanasia. The former cites key references as documents produced by the OECD and the CCAC, the latter draws on guidelines that have been published by the American Veterinary Medical Association and by the European Commission (Demers et al., 2006). Documents detailing international harmonisation of guidelines on animal user training programmes and the review of research protocols have been approved by ICLAS; the fifth guideline will address issues related to the production and use of genetically modified animals.

Through this programme ICLAS has demonstrated the potential opportunities for and benefits from a harmonised approach to promoting animal welfare, and collaborations with international organisations such as CIOMS and OIE will enable ICLAS to promote the implementation of these guidelines to a wider audience.

2.3 Organisation for Economic Cooperation and Development (OECD)

The OECD has been proactive in the international harmonisation of guidelines relating to the testing of chemicals. The Council's decision in 1981 concerning the mutual acceptance of data when those data were generated in accordance with OECD Test Guidelines and the Principles of Good Laboratory Practice (GLP) was seen as a way to reduce the numbers of animals used in chemical testing and clearly has had significant long-term implications for animal welfare in other areas of regulatory testing. At that time, this decision referred to activities in member countries only but, since 1997, non-member countries have been able to participate in this programme (Koëter, 2003).

The OECD has pursued a policy to implement the 3R principles wherever possible and to that end has revised Test Guidelines where possible, removing the requirement for tests with a high impact on animal welfare or introducing amendments to reduce that impact. As of May 2009 the OECD had 24 Test Guidelines in preparation or under revisions with relevance to the goals of the 3Rs.

Two Guidance Documents with significant implications for animal welfare are on the *Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation* (OECD, 2000) and the Validation and International Acceptance of New and Updated Test Methods for Hazard Assessment (OECD, 2005). As mentioned above, the guideline on humane endpoints has general application to the use of animals in research. It was developed to provide a practical guide to applying the 3Rs with an emphasis on Refinement and highlights the importance of planning a study to identify the earliest possible endpoints, the use of pilot studies, ways to recognise and assess pain, distress and suffering and provides practical guidance to making an informed decision to humanely kill animals.

The Guidance Document on the international acceptance of validated new methods has application to the general area of toxicity testing. Two other organisations that are working towards harmonising international validation of toxicity test methods and associated criteria for regulatory acceptance of these methods, namely the European Centre for the Validation of Alternative Methods (ECVAM) and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) were involved in the development of the OECD document; each agency has adopted similar validation criteria.

2.4 International Cooperation on Alternative Test Methods (ICATM)

International cooperation in the application of the 3Rs to regulatory toxicity testing has been at the forefront of developments to harmonise guidelines that promote the welfare of animals used for scientific purposes. Agencies such as ICCVAM and ECVAM have provided a focus for such research activities and have been strong advocates for the development and validation of alternatives, particularly in relation to toxicity testing. With responsibility for evaluating the validation status of new and alternative test methods and coordinating these activities over fifteen (15) Federal agencies, ICCVAM plays a critical role in the US. Similarly, ECVAM plays a pivotal role in the European Commission. Both organisations hold within their charter to facilitate international harmonisation of toxicology test protocols that promote the 3Rs.

In April 2009, ICCVAM, ECVAM, the Japanese Centre for the Validation of Alternative Methods (JaCVAM) and Health Canada, entered into an agreement establishing the International Cooperation on Alternative Test Methods (ICATM). This agreement establishes an international framework for enhanced cooperation and coordination in validation studies and peer review. In so doing, the goal is to provide a strong, transparent, scientific basis for recommendations to harmonise test protocols to ensure wider acceptance of alternative methods and strategies that promote the 3Rs.

3 Achieving the 3Rs

It is irrefutable that the principles of the 3Rs have been the guiding force in these developments, not only providing goals against which progress in promoting animal welfare outcomes can be measured but also as a unifying platform that facilitates the identification of common aspirations. However, it is salutary

to realise that these principles were put forward many years before Russell and Burch published *The Principles of Humane Experimental Technique* (1959). For example, in 1847, the physiologist Marshall Hall argued that animals should not be used in an experiment unless there was a good scientific reason to do so, that such experiments should not be repeated needlessly and animals should experience the least possible infliction of suffering. What distinguishes Russell and Burch is that they advanced the scientific rationale for implementing these principles, arguing the inextricable link between animal welfare and scientific outcomes.

The impact of these international initiatives in promoting the welfare of animals used in science will be evident in the extent to which the goals of Replacement, Reduction and Refinement are achieved throughout the world. However, international guidelines will be based on agreed principles, and it will be the scientific underpinning of the implementation of those principles that will enable or limit potential outcomes. This is evident in the success that has been achieved in the Replacement of animals in toxicity testing and the harmonisation of toxicity testing protocols with the concomitant effect of a Reduction in the number of animals used. The constraints to the implementation of such guidelines where scientific evidence is limited are noted by Demers and colleagues (2006) in their discussion of the implementation of the ICLAS document on *International Harmonization of Guidelines on Euthanasia*. Although the general principles for euthanasia could be drawn from the cited key references, in those documents there are differences in the assessment of and recommendations for the use of various techniques, indicating areas where further research is needed.

The Principle of Refinement has broad implications to achieving the goals of any international guideline concerning the use of animals for scientific purposes; strategies to recognise and alleviate pain and distress being a primary focus. Russell and Burch identified our lack of understanding of the mechanisms of pain and distress as presenting a major barrier to the development and implementation of strategies to achieve the goal of Refinement. Since that time, there have been significant advances in our understanding of the physiological and behavioural basis for these phenomena in animals as well as the development of strategies to recognise and manage these experiences in the most common laboratory species. However, as argued in two recent major reviews conducted by ILAR (National Research Council, 2008 and 2009), there are significant gaps in our knowledge that limit our ability to effectively manage these experiences in animals. Both reviews identify the 3Rs as setting the benchmark against which strategies to recognise and alleviate pain and distress must be judged but also identify the risk of uncontrolled pain or distress confounding the interpretation of research data.

Despite the availability and knowledge of the efficacy of analgesics for use in laboratory species, including rodents, surveys of the use of such agents after surgical procedures suggests a relatively low level of use (Richardson and Flecknell, 2005; Stokes et al., 2009). Further, in these studies, both the use of analgesics and of anaesthetics with the potential to effect post-operative analgesia were included, and evidence of effective pain management such as monitoring outcomes were not considered. Of note, using the same survey method, this research group reported a higher use of analgesics in other laboratory species such as rabbits, pigs, sheep, dogs and non-human primates. Based on a roundtable discussion at the 2003 National AALAS Meeting, Karas (2006) identified a number of barriers to the assessment and treatment of pain in laboratory animals, including a lack of knowledge of effective techniques to assess, monitor and treat pain, the lack of information published on effective techniques, limited scientific evidence to support published methods and concerns about the risk of drugs confounding research results.

Both the ILAR Report on the Recognition and Alleviation of Distress (National Research Council, 2008) and Karas (2006) identified the need for research into the impact of an animal's living conditions, not only as a potential cause of distress but also causing unexpected variability in physiological and behavioural responses and so affecting the reliability of research data. In November 2003, ILAR hosted an international workshop to discuss the development of evidence-based guidelines for laboratory animal care (National Research Council, 2004). Recognising the need to critically review and apply the available scientific evidence, the workshop participants identified a number of significant knowledge gaps in relation to species-specific data. Similarly, a lack of evidence supporting the current use of various enrichment strategies for laboratory species was identified in a report by a FELASA Working Group (FELASA, 2006). Such data would not only inform animal care practices but, as has been highlighted in recent reports of phenotypic variability in genetically modified mouse lines, an understanding of the complexity and subtlety of gene-environment interactions has implications for the transferability of data from animal models.

Clearly, the availability of scientific evidence to support the development and application of strategies to identify and manage pain or distress is essential to achieve the goals of Refinement and the lack of such data may limit the impact of initiatives to develop a harmonisation of animal welfare guidelines. However, as has been outlined above, research addressing these questions has implications for both animal welfare and scientific outcomes.

4 Animal Welfare Science

Russell and Burch argued that what they described as "the psychosomatics of experimental animals" was the most important single factor influencing the achievement of the humane experimental techniques. Today, the study and interpretation of data in relation to the experience of emotional states in animals and their cognitive abilities are controversial. However, such evidence is central to the concept of animal welfare and fundamental to the discipline of animal welfare science (Dawkins, 2006). Evidence of an animal's emotional state, both negative and positive, is integral to our interpretation of how it feels in a given circumstance. That such experiences are subjective and not available to critical analysis, led the authors on the ILAR reports on distress and pain, to discount evaluation of the emotional component of these experiences. However, an alternative view is that such

experiences can be measured, albeit with acknowledgement that this is an emerging and challenging area of science (Dawkins, 2006; Boissy et al., 2007; Mendl et al., 2009).

Whether in developing more sophisticated and reliable ways to diagnose pain and distress or in assessing an animal's responses to its environmental conditions, the development and validation of methods that will enable our understanding of the experiences of animals in the research setting has implications to achieving the goals of Refinement. This is an area where the application of the knowledge and methods from animal welfare science will not only advance animal welfare but also scientific outcomes.

References

Boissy, A., Manteuffel, G., Jensen, M. B. et al. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior 92*, 375-397.

CIOMS (1985). *International Guiding Principles for Biomedical Research Involving Animals*.

Coullter, C. A., Flecknell, P. A. and Richardson, C. A. (2009). Reported analgesic administration to rabbits, pigs, sheep, dogs and non-human primates undergoing experimental surgical procedures. *Laboratory Animals 43*, 232-238.

Dawkins, M. S. (2006). A user's guide to animal welfare science. *TRENDS in Ecology and Evolution 21*, 77-82.

Demers, G., Griffin, G., De Vroey, G. et al. (2006). Harmonization of Animal Care and Use Guidance. *Science 312*, 700-701.

FAO (2009). *Capacity Building to Implement Good Animal Welfare Practice*. Rome Italy. Available at: ftp://ftp.fao.org/docrep/fao/011/i0483e/i0483e00.pdf (Accessed August, 2009).

FELASA (2006). *FELASA Working Group Report Standardization of Enrichment*. http://www.lal.org.uk/pdffiles/FELASA_Enrichment_2006.pdf (Accessed August 2009).

Fraser, D. (2008). Toward a global perspective on farm animal welfare. *Applied Animal Behaviour Science 113*, 330-339.

Hall, M. (1847). A critical and experimental essay on the circulation of the blood. Reprinted in *Lancet 135*, 58-60.

Karas, A. Z. (2006). Barriers to assessment and treatment of pain in laboratory animals. *Lab. Animal 35*, 1-8.

Koëter, H. B. (2003). Mutual acceptance of data: Harmonized test methods and quality assurance of data – the process explained. *Toxicol. Lett. 140-141*, 11-20.

Mendl, M., Burman, O. H. O., Parker, R. M. A. and Paul, E. S. (2009). Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science 118*, 161-181.

National Research Council (2004). *The Development of Science-based Guidelines for Laboratory Animal Care*. Washington DC: National Academies Press.

National Research Council (2008). *Recognition and Alleviation of Distress in Laboratory Animals*. Washington DC: National Academies Press.

National Research Council (2009). *Recognition and Alleviation of Pain in Laboratory Animals*. Washington DC: National Academies Press.

OECD (2000). *Guidance Document No.19 Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation*. Paris, France.

OECD (2005). *Guidance Document No. 34, Validation and International Acceptance of New and Updated Test Methods for Hazard Assessment*. Paris, France.

OIE (2005). *Terrestrial Animal Health Code*. Paris, France: World Organisation for Animal Health (OIE).

Richardson, C. A. and Flecknell, P. A. (2005). Anaesthesia and post-operative analgesia following experimental surgery in laboratory rodents. *ATLA 33*, 119-127.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen & Co Ltd.

Stokes, E. L., Flecknell, P. A. and Richardson, C. A. (2009). Reported analgesic and anaesthetic administration to rodents undergoing experimental surgical procedures. *Laboratory Animals 43*, 149-154.

Correspondence to

Margaret Rose
Room G6, Level 2
Clinical Sciences Building
Prince of Wales Hospital, High Street
Randwick, NSW 2031
Australia
e-mail: m.rose@unsw.edu.au

Planning for Refinement and Reduction

Derek Fry[1], Rose Gaines Das[1], Richard Preziosi[2] and Michelle Hudson[1]
[1]FRAME, [2]University of Manchester, UK

Summary
Experiments using laboratory animals do not normally occur as isolated 'one-off' studies and there is often considerable scope for reducing overall severity and the number of animals used by careful planning of the complete programme. The paper illustrates this with examples and provides a commentary on the flowchart for planning a research programme developed by the FRAME Reduction Steering Committee. The flowchart reminds researchers that programme planning needs well-specified objectives and research into different ways of achieving them, and into the severity of the methods involved. The least severe sequence can then be chosen and the series planned to identify unexpected adverse effects and good endpoints early on, so later experiments can minimise severity. Starting with low-severity work can avoid unnecessary higher-severity studies and individual experiments can be designed to minimise numbers and severity. The flowchart should help experimenters plan minimal-severity programmes and be useful for ethical evaluation.

Keywords: planning, refinement, reduction, severity, experimental design

1 Introduction

A need for better training in the design of experiments and the strategy for experimental programmes became apparent from discussion at a number of meetings (see Howard et al., 2009). In response, the FRAME Reduction Steering Committee (FRSC) has been running training courses on this topic for postgraduates. In looking for material to bring together the various aspects of planning and carrying through an experimental programme, the Committee found the available flowcharts insufficient and has developed its own (Gaines Das et al., 2009). This covers not just the planning and design of individual experiments but also the strategy for the whole programme. The strategy is important, as experiments using laboratory animals do not normally occur as isolated 'one-off' studies and there is often considerable scope for decreasing overall severity and reducing the number of animals used by careful planning and design in the context of the complete programme. Gaines Das et al. (2009) concentrated on planning to reduce overall numbers: this communication illustrates how this approach could be used to reduce the severity of an experimental programme.

Fifty years ago Russell and Burch (1959) recognised that "One general way in which great reduction may occur is by the right choice of strategies in the planning and performance of whole lines of research." They also recognised the ethical imperative to "reduce to an absolute minimum the amount of distress imposed". Reduction in animal usage reduces overall suffering by exposing fewer animals to adverse effects, but good programme planning can also minimise overall severity. Unfortunately, experimental design texts usually provide no guidance on how to design an individual experiment to minimise severity and are silent on how to organise a sequence of experiments. Ethical evaluation processes that judge only protocols may well miss possibilities for minimising overall severity through a suitable strategy for the whole programme. It is also a topic missing in FELASA's suggested syllabus for the training of researchers (see FELASA 1995). The UK has had an advantage in developing ideas in this area, as researchers have been obliged since 1987 to apply for projects covering an experimental programme of up to five years and a key section of the project licence application form has been the plan of work (see http://scienceandresearch.homeoffice.gov.uk/animal-research/publications-and-reference/publications/licences/project-licences/ for the latest example of this on the application form and notes with the form on planning and refinement). Many of the ideas developed over 50 years within the UK animal science community and by the UK Animals Scientific Procedures Inspectorate and its predecessor can be found in the fact sheet prepared by Morton (1998) and how they are encouraged in practice can be seen in the first report of the Animals Scientific Procedures Inspectorate (Home Office, 2004). Strategy for reduction is covered to some extent in Festing et al. (1998), the value of carefully-specified objectives in Fry (2004) and a step-wise approach to refining an experimental programme in Fry and Morton (2000).

In Europe the replacement of Directive 86/609, with its introduction of "projects", could place more emphasis on ethical evaluation of the overall severity of the programme. In the proposal put to the European Parliament in November 2008 Article 37 requires the assessment of whether "the project is designed so as to enable procedures to be carried out in the most humane ... manner." This could translate into closer scrutiny of the combined severity of a series of individual experiments. There is also an emphasis on retrospective review – another feature in the FRSC flowchart. Generally, wider appreciation that good planning avoids wasteful use of animals and can reduce the extent of animal suffering caused by an experimental programme

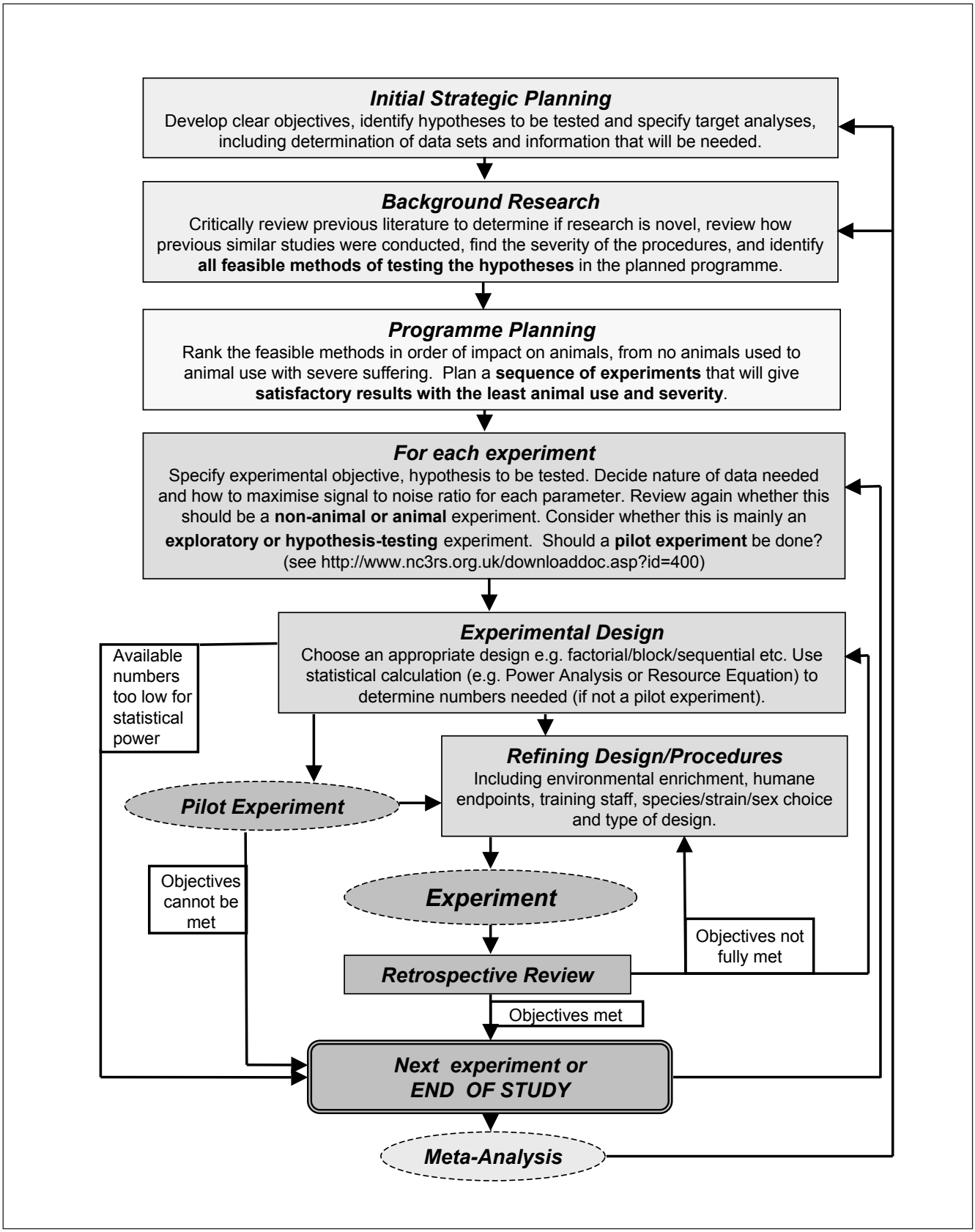


Fig. 1: FRAME Reduction Steering Committee Strategic Planning Flowchart. Slightly modified and reproduced from Gaines Das et al. (2009) with permission from ATLA.

is likely to raise ethical awareness of the value of looking beyond the individual protocol. The FRSC flowchart, reproduced in Figure 1 with an addition on researching the severity of procedures, is designed as a framework to help both researchers and those assessing programmes achieve good scientific outputs with minimal animal use and suffering.

2 Planning a programme for minimal severity

2.1 Initial planning

The first and most important parts of any strategic plan are a) setting the aims or general objectives and b) doing the relevant background research pertaining to these objectives. When the broad objectives are clearly specified, an appropriate programme can be planned within which possibilities for studies not involving animals can be identified and individual animal experiments can be considered. It is important to distinguish between the general aims of the programme and the hypotheses the individual experiments would test. Without clarity at both of these levels it is difficult to consider all feasible methods of testing the hypotheses of interest without using animals, or to see non-animal methods that could be developed for the purpose. In the severity context specified objectives are needed for setting objective-dependent endpoints for the individual experiments (see below).

In the initial strategic planning stage, decision points in the programme should also be considered. At a decision point the progress so far is reviewed and decisions taken on whether in the interests of minimal animal use and severity the plans should be changed, additional monitoring or revised procedures incorporated or the programme abandoned. If the first part of the programme involves development of a new animal model, for example, it may be necessary to accept that a satisfactory model is illusory and save further suffering by taking a different approach.

2.2 Background research

The flowchart suggests that background research should not be limited to checking to avoid duplication and seeking one practicable route to pursuing the aim of the programme, but should include

- identifying the range of possible routes and both their relative need for use of animals and their relative severities
- obtaining information on the conduct of experiments similar to those envisaged and any adverse effects encountered
- gathering information on the likely constraints, such as time taken for procedures and availability of people and of animal accommodation and
- researching the techniques and sampling methods proposed, the severity and possible adverse effects involved and their limitations.

Articles on specific techniques or procedures may identify adverse effects and there may also be clues in experimental reports. Unexplained differences in group sizes, for example, may indicate that some animals unexpectedly died or had to be withdrawn from a study on humane grounds.

2.3 Programme planning

Using the information obtained from experience and background research, a series of studies likely to use minimal numbers and involve least severity can be structured around the planned decision points. As pointed out in Gaines Das et al. (2009), this programme planning should also consider and take into account constraints that may frustrate achieving the objectives or involve unnecessary animal use or suffering. These include limitations on the resources that can be drawn upon, availability of accommodation, restrictions on sample collection and handling and how much data can be obtained in a given time period.

For minimising severity, a critical element is planning the sequence of experiments so that at an early stage adverse effects, and how to avoid or control them, are identified and humane endpoints are set and reviewed in practice. Any programme is likely to have unpredictable unknowns, such as technical difficulties with a published procedure new to the laboratory, unanticipated adverse effects of a substance or procedure or combination of experimental treatments. An objective of the first experiments, or the first experiment in each subordinate series, should be to identify such unknowns and suitable humane endpoints.

When developing a new animal model the background research should enable the planner to pre-set criteria by which the model will be judged as successful, and a decision point at which the results with the model are assessed against these criteria should be planned into the programme.

3 Examples of planning for minimal severity

In vaccine studies the aim of a programme may be to test new vaccines, with different studies on efficacy and safety. Animals are likely to be subjected to unnecessary suffering if the safety studies, which typically involve giving double or more the expected dose of the vaccine, are optimistically planned to start before the results of efficacy studies are known, since a proportion of the vaccine tested will show insufficient efficacy.

A programme may be staged to achieve all the objectives that can be met at low severity first, only proceeding to higher severity studies when the low severity approach is exhausted. The sequence of questions could be -

Question	Example of response
What is the overall objective?	To determine the effects of hypoxia on sympathetic nerve activity
What can be done without animals?	No relevant studies – an intact vascular nerve network is needed
What can be done under terminal anaesthesia?	All the studies on the effects of hypoxia up to 24 h

Question	Example of response
What can be done with only mild severity?	Studies on the effects of prolonged mild hypoxia in which the experiments involve >24 h exposure, then terminal anaesthesia
What can only be done at more than mild severity?	Investigation of effects of prolonged, more severe hypoxia
	Confirmation studies with implanted electrodes in conscious animals

False assumptions are then exposed and basic information obtained without subjecting animals to the more distressing procedures. In some cases the results from the milder studies may well indicate that the more severe ones are not needed.

The development of a mouse model of acute pancreatitis illustrates how planning to minimise severity could operate to save much animal suffering. Acute pancreatitis is a very painful condition in humans. It carries substantial morbidity and can be fatal. It is worse in those who are obese and the reasons for this are unknown. Deficiencies in current treatment and management of patients with pancreatitis, the lack of understanding of the mechanism for the higher severity in obesity and the need for the interactions of multiple body systems to mimic the condition justify use of animals to study how the condition can be ameliorated. The objective of the programme is to investigate why acute pancreatitis is worse in obesity by first developing an obese animal model, then using pharmacological dissection to detect mechanisms and compare these with those found in non-obese animals. Background research indicates that intraperitoneal injection of a combination of two interleukins should induce pancreatitis and that a genetically obese mouse should be a good prospective model, but also that acute pancreatitis in rodents produces substantial suffering. In acute pancreatitis in man, and non-obese rodents, there are early changes in serum amylase and other enzymes and marked histological alterations to pancreatic cells. Certain serum components are good early predictors of the severity of the pancreatitis. Reasonable criteria for the obese mouse model could be the demonstration of comparable serum and histological changes.

The programme sequence could be a pilot experiment with a dose and combination of the agents expected to produce pancreatitis, with serial blood sampling and post mortem histology as measures, and signs of abdominal pain or a set time after the injection to induce pancreatitis as an endpoint. This should give a good indication of whether the model has prospects and whether the severity controls and endpoint can be refined. It could be followed by a factorial experiment using different interleukin doses and combinations to determine the optimal dose combination for producing raised serum levels and the characteristic histological changes. A further experiment could determine the time course of the enzyme and other serum changes more precisely, and then one with the animals killed at set time points to follow the histo-pathological changes in the pancreas. Finally, it might be necessary to allow progression to full blown acute pancreatitis in a few animals to confirm that the demonstrated signs are genuinely those of early development of the condition. This would establish the model and the pharmacological studies could then use the optimal induction and sampling arrangements determined from these experiments. The pharmacological experiments could routinely end at the time when definitive early changes were reliably detectable, sparing animals the further suffering of a progression to later stages of the disease.

The approach taken by Sennello et al. (2008) to developing an obese mouse model of acute pancreatitis seems to have been very different. The first experiment reported in the paper is a survival study, with the number of mice that died after the interleukin injection as the key parameter. When all mice in the obese group of 10 mice died but none in the non-obese group, and this was confirmed in a repeat experiment, the dose was increased for groups of 10 of the latter until 30% died. A further experiment confirmed that multiple organ failure occurred. Subsequent experiments then followed the time course of the serum changes and the pancreatic histology. Clear changes were detectable two to six hours after injection, whereas mice took 24-48 hours to die. This sequence is quite different in severity from that outlined above.

4 Achieving minimal severity at the experiment level

The lower half of the flowchart mainly concerns the individual experiments. The flowchart assumes researchers will be familiar with the important points on refining experimental procedures and designing experiments for efficient use of animals well covered elsewhere (For the former see the UK National Centre for the 3Rs Information Portal at http://www.nc3rs.org.uk/landing.asp?id=38, Morton 1998, or the section on *Avoiding or Minimizing Distress in Laboratory Animal Use* in the report of the US National Academy of Sciences 2008, and for the latter see Festing et al., 2002). There are some additional considerations for designing for minimal severity, however.

4.1 Setting clear objective(s) for each experiment

As pointed out elsewhere (Gaines Das et al., 2009; Fry, 2004), this is important for formulating a design that uses the right number of animals. In the context of refinement it is crucial to setting objective-related endpoints, i.e. the points when individual experiments have met the objective or clearly cannot achieve it. Continuing beyond this endpoint involves risk of animal suffering or distress, for which there is no justification (see Fry, 1998).

4.2 Pilot experiments or dose-setting procedures

The flowchart highlights the value of incorporating pilot experiments at suitable points in the programme and provides the link to a National Centre for the 3Rs online document on the subject. These preliminary experiments are likely to be worthwhile use of animals to establish proof-of-concept or to provide use-

ful information on technical problems, the time course of an experimental outcome, or how much resource is needed for a full-size experiment. In regulatory toxicology studies and some pharmaceutical work the dose-setting procedure is essentially a type of pilot study.

Research pilots and dose-setting runs should also be planned to provide information to help minimise severity. This means having observation schedules that will detect adverse effects and indicate their severity and duration, so that in the definitive studies suitable monitoring arrangements can be made, opportunities for refinements considered and incorporated, and severity-related humane endpoints identified. Noting the time course of the experimental effect can help with setting objective-related humane endpoints.

4.3 Designs to minimise severity

Factorial designs

These are recommended for their efficient use of animals (Shaw et al 2002), but can also be used to minimise severity by pointing to optimal conditions for a series of studies where the conditions carry significant severity. For example, a group needs to set up an hypoxia model to study alterations in sympathetic nerve activity produced by exposure to low oxygen tensions. A mouse strain is known to show the effect but is sensitive to low oxygen tensions and can go into respiratory distress. Good data on the minimal extent and duration of hypoxia are not available. A factorial design for the exploratory experiment, in which several levels of oxygen deprivation and different durations are used in various combinations in the same experiment, will be the most efficient way of seeing the minimal level and duration of hypoxia needed to produce an effect large enough to study. This combination can then be used for performing the subsequent series of experiments with minimal adverse effects from the hypoxia.

Sequential designs

Sequential experimental designs, which allow the cumulative analysis of data, are efficient in that the study only continues until the objective is reached. Compared to other designs, where estimates of the number of animals needed have to be made in advance and risk either overestimating or using too few (and thus wasting them), severity is reduced as only the number of animals actually needed experience distress. Definitive experiments of this nature (e.g. Waterton et al., 2000) need careful planning and analysis as assumptions have to be made about the comparability of the conditions for successive groups of animals and the risk of time bias. The help of a statistician is highly advisable.

However, a modified sequential design, the "up and down" approach, is one of the standard methods for determining acute oral toxicity in regulatory studies (Bruce 1985; Rispin et al. 2002), and a similar approach (but without lethality or predicted lethality as the endpoint) could be considered when planning research studies for which it is suitable. For example, a dose of compound expected to show some effect if the compound is efficacious is given to two animals and they are observed for a set period. (Using a pair of animals gives some control for variability, but if the likely effect is severe, as in acute oral toxicity tests, then only one animal at a time should be used). If no or little effect is seen, the dose is increased for the next pair and so on until a useful efficacy of the substance can be determined or a cut-off point at which it can be deemed ineffective is reached. If the first pair shows higher effects than needed, the dose for the second pair is decreased. This continues until a minimal useful effect level is found. Using such an approach for research work where a substance of unknown effect is being given, or an agent is being tried on a new genetically modified mouse line, should help minimise severity by minimising the numbers exposed to adverse effects of unknown or unpredictable severity.

Use of unequal group sizes

Where the experimental procedure involves considerable severity, distributing the animals into a large control group and small experimental group or groups can provide as powerful a design as one using equal group sizes, with only a modest increase in total number required (Ruxton, 1998). Again severity is reduced because many fewer animals are exposed to the suffering of the experimental procedure. Planning for this means the extra animals required are ordered in time, the arrangement of any extra cages for the large control group is properly set out and the analysis and interpretation of the results are considered.

5 Conclusion

Good overall planning allows early identification of factors that may affect severity, such as unanticipated adverse effects and, where possible, stages the experiments in a programme to achieve objectives at the lowest level of severity before proceeding to higher severity. Careful consideration of the steps and experiments needed to meet well-specified objectives and adjusting the sequence of experiments to involve minimal severity is necessary to "reduce to an absolute minimum the amount of distress imposed ..." (Russell and Burch, 1959). The FRAME flowchart brings together the ideas on planning for reduction and refinement into a single sheet and includes points not made in previous overviews. It should help experimenters plan programmes of minimal severity and may also be a useful checklist for those doing ethical evaluation.

References

Bruce, R. D. (1985). An up-and down procedure for acute toxicity testing. *Fundam. Appl. Toxicol. 5*, 151-157.

FELASA (1995). Recommendations on the education and training of persons working with laboratory animals: Categories A and C. *Laboratory Animals 29*, 121-131.

Festing, M. F. W., Baumans, V., Combes, R. D. et al. (1998). Reducing the use of laboratory animals in biomedical research: problems and possible solutions. *ATLA 26*, 283-301.

Festing, M., Overend, P., Gaines Das, R., et al. (2002). *The Design of Animal Experiments: Reducing the Use of Animals in Research Through Better Experimental Design* (39). London: Royal Society of Medicine Press Limited.

Fry, D. J. (2004). Reduction by well-defined objectives. *ATLA 32 (Suppl. 1)*, 241–244.

Fry, D. J. (1998). Relating criteria for humane endpoints to objectives. In C. F. M. Hendriksen & D. Morton (eds.) *Humane Endpoints in Animal Experiments for Biomedical Research* (54-57). London: Royal Society of Medicine Press Limited.

Fry, D. J. and Morton, D. B. (2000). Non-statistical experimental design as an aid to refinement. In M. Balls, A. M. van Zeller & M. Halder (eds.) *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (1687-1690). Amsterdam: Elsevier.

Gaines Das, R., Fry, D. J., Preziosi, R. and Hudson, M. (2009). Planning for reduction. *ATLA 37*, 27-32.

Home Office (2004). Animal (Scientific Procedures) *Inspectorate Annual Report 2004* (12-13) available at http://scienceandresearch.homeoffice.gov.uk/animal-research/publications-and-reference/publications/reports-and-reviews/annual-report?view=Binary

Howard, B., Hudson, M. and Preziosi, R. (2009). More is less: reducing animal use by raising awareness of the principles of efficient study design and analysis. *ATLA 37*, 33-42.

Morton, D. B. (1998). The importance of non-statistical design in refining animal experimentation. ANZCCART Facts Sheet. *ANZCCART News 1*, 12.

Rispin, A., Farrar, D., Margosches, E., et al. (2002). Alternative methods for the median lethal dose (LD(50)) test: the up-and-down procedure for acute oral toxicity. *ILAR J. 43*, 233-243.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen.

Ruxton, G. D. (1998) Experimental design: minimising suffering may not always mean minimising number of subjects. *Animal Behaviour 56*, 511-512.

Sennello, J. A., Fayad, R., Pinin, M. et al. (2008). Interleukin-18, together with interleukin-12, induces severe acute pancreatitis in obese but not in non-obese leptin-deficient mice. *Proc. Natl. Acad. Sci. U.S.A. 105*, 8085–8090.

Shaw, R., Festing, M. F. W., Peers, I. and Furlong, L. (2002) Use of factorial designs to optimize animal experiments and reduce animal use. *ILAR J. 43*, 223-232.

US National Academy of Sciences (2008). *Recognition and Alleviation of Distress in Laboratory Animals* (63-86). The National Academies Press: Washington.

Waterton, J. C., Middleton, B. J., Pickford, R. et al. (2000). Reduced animal use in efficacy testing in disease models by the use of sequential experimental designs. In M. Balls, A. M. van Zeller & M. Halder (eds.) *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (737-745). Amsterdam: Elsevier.

Correspondence to

Michelle Hudson MbiolSci
Scientific Officer
FRAME
96-98 North Sherwood Street
Nottingham, NG1 4EE
UK
e-mail: science@frame.org.uk

BS29: Immunology

Innate Immune Mechanisms in Contact Dermatitis and the Resulting T Cell Responses

Stefan F. Martin and P. R. Esser

Allergy Research Group, Department of Dermatology, University Medical Center Freiburg, Germany

Summary

Allergic contact dermatitis is a chemical-induced inflammatory skin disease that is mediated by T cells reacting to chemical contact allergens. A key event in this disease, both in the sensitisation and elicitation phase, is the induction of inflammation. The molecular mechanisms by which chemicals induce skin inflammation are poorly understood. It is becoming clear that contact allergens activate innate immune and stress responses, and analogies with infections are observed. Recent studies are beginning to elucidate the pathways of activation of the innate mechanisms involved in the allergic sensitisation process and the complex interplay of innate immune and stress responses. The progress made in basic research in this field has great potential to exploit this knowledge for the development of better treatment strategies for allergic contact dermatitis as well as for the development of in vitro assays for contact allergen identification and potency assessment to replace animal testing.

Keywords: contact dermatitis, innate immunity, Toll-like receptor, NOD-like receptor, T cell, dendritic cells

1 Innate immune and stress responses to pathogens

The immune system is constantly challenged by infections with pathogenic bacteria, viruses, fungi or protozoa. The first line of defence against these threats is the innate immune system and innate stress responses such as oxidative stress. The main benefit of these responses is their rapid kinetics and their efficient induction of an inflammatory response, which results in the activation of tissue cells and in the recruitment of effector cells of innate immunity, such as neutrophils, mast cells or NK cells. There is a broad understanding of the mechanism by which pathogens induce these responses, which are a prerequisite for the activation of the adaptive immune system (Medzhitov, 2007). The T and B cells of the adaptive immune response are activated only within days of an infection. T cell priming occurs when dendritic cells are activated in the inflammatory milieu of the infected peripheral tissue and migrate to the draining lymph nodes, where they present antigens derived from pathogens in the form of peptides on MHC molecules to naïve T cells. The antigen-specific T cells, once activated, start to expand by proliferation and differentiate to potent effector and memory T cells that have received the information about the site of the infection by the dendritic cells and enter the blood circulation. We and others have previously shown that dendritic cells immigrating from the peripheral tissues induce a specific combination of homing receptors, i.e. adhesion molecules and chemokine receptors, which allow the circulating effector and memory T cells to enter the infected tissue due to the homing signals, chemokines and adhesion molecules, e.g. on endothelial cells of the blood vessels, that are produced during the inflammatory response (Dudda and Martin, 2004; Mora and von Andrian, 2006). Pathogens harbour so-called pathogen-associated molecular patterns (PAMPs) that are recognised by pattern recognition receptors (PRR) of the innate immune system (Kawai and Akira, 2007). These PAMPs are often molecular patterns such as lipopolysaccharide (LPS) from the cell wall of Gram-negative bacteria or nucleic acids such as viral double-stranded RNA (dsRNA). The PRR that recognise these PAMPs, also called danger signals, belong to the Toll-like receptor (TLR) family, which comprises about 13 members with different PAMP specificities. TLR are expressed on tissue cells, innate and adaptive immune cells, and they are localised

in the plasma membrane or in the endosomal membrane in the case of the nucleic acid-recognising TLR3, 7, 8 and 9. As an example for PAMPs, LPS activates TLR4, viral dsRNA TLR3. The consequence of TLR triggering is the activation of NF-κB signalling, MAP kinase signalling, and the production of type I interferons (IFNs). Eventually, cytokines, chemokines and antimicrobial peptides are produced which orchestrate the innate inflammatory response.

Another family of PRR are the NOD-like receptors (NLR). They are also expressed in a variety of cell types and have different functions (Martinon et al., 2009). Interestingly, they seem to have a much broader specificity for activating ligands. Thus, it is not yet clear how they are activated. One member of this family is NLRP3, which is a component of the inflammasome. This cytosolic protein complex consisting of NLRP3, the adaptor protein ASC and pro-caspase-1 activates the caspase, which then processes the pro-forms of IL-1β and IL-18, which are produced in response to TLR triggering and play an important role in the inflammatory response.

In addition to TLR and NLR triggering, pathogens can activate the production of reactive oxygen species (ROS), e.g. in response to phagocytosis of the pathogen. ROS are now recognised as not purely toxic molecules but as signalling molecules, which can directly or indirectly regulate transcription factor activities and produce oxidised lipids as important signalling molecules (Lambeth, 2004).

TLR, NLR and oxidative stress pathways as well as the antioxidant phase 2 response mediated by the cytosolic sensor molecule Keap-1 and the transcription factor Nrf2 (Dinkova-Kostova et al., 2005) interact with each other by intensive crosstalk. The resulting innate and adaptive immune response can then protect against infection or, in the case of a dysregulated response, cause immunopathology.

2 Innate immune and stress responses to contact allergens

Low molecular weight chemicals, like electrophilic organic compounds and metal ions, can cause allergic contact dermatitis (ACD), a T cell mediated inflammatory skin disease (Martin, 2004). Similar mechanisms are found for drugs that induce T cell-mediated drug hypersensitivity. It is up to now unclear how these chemicals activate the innate immune system in order to produce a pathogenic T cell response in the skin. It has been shown that T cells recognise contact allergens as classical haptens, either covalently bound to peptides presented on MHC molecules of antigen presenting cells or complexed to peptide and/or MHC and T cell receptor in the case of metal ions. As mentioned above the innate immune response is crucial for the activation of naïve T cells. Recent research in the mouse model of ACD, the contact hypersensitivity (CHS) model, has provided evidence that contact allergens may trigger TLR, NLR and induce oxidative stress and anti-oxidant responses as seen in response to infection (Freudenberg et al., 2009). Looking at this striking analogy, the fascinating question to be answered in the future is how protein reactivity as a hallmark of contact allergens is translated into the activation of signalling pathways. So far only hypotheses can be put forward to explain the potential mechanisms (Martin et al., 2006; Freudenberg et al., 2009). Contact allergens may modify proteins in such a way that they change their function, or they may induce danger signals that can serve as triggers for TLR and NLR. It is known that contact allergens can induce ROS production, but how they do that is unclear. As mentioned above, ROS engage in important signalling processes that may be relevant in ACD. We hypothesise that contact allergen modification of proteins can be perceived in some cases, like conventional post-translational modifications such as phosphorylation or glycosylation. This could directly induce signalling function in innate immune receptors. Alternatively, as recently shown by us (Martin et al., 2008), contact allergens may induce the breakdown of extracellular matrix (ECM) components such as hyaluronic acid. Such HA fragments are known to trigger TLR2 and TLR4, which we have shown to be important in the CHS model. Mice lacking both of these receptors are resistant to CHS induction. In summary, contact allergens may themselves serve as danger signals or induce the production of such endogenous danger signals, e.g. by causing damage to the ECM or to cells, which may explain the TLR4 dependent role of heat shock proteins such as HSP27 and HSP70 in CHS (Yusuf et al., 2009). In analogy to PAMPs such endogenous danger signals are called damage-associated molecular patterns (DAMPs) (Seong and Matzinger, 2004). Future research involving proteomics and chemical biology should identify hapten modified proteins and identify modifications that alter protein function, localisation, interaction with other proteins or proper folding.

An interesting development is the demonstration that cysteine reactive contact allergens bind covalently to Keap-1 and trigger the antioxidant response via Nrf2 mediated induction of enzymes that replenish the cellular antioxidant pool, such as catalase, heme oxygenase, and NADPH quinone oxidoreductase and glutathione reductase (Ade et al. 2009; Natsch and Emter, 2008). This identifies Keap-1 as a very important functional target for contact allergens.

3 Implications for therapeutic approaches in allergic contact dermatitis and replacement of animal testing for contact allergen identification and risk assessment

The realisation that contact allergens, as non-infectious and non-replicative agents, activate either directly or indirectly the same innate immune and stress pathways as infectious agents has a number of important implications. Polymorphisms in innate immune receptors are often associated with autoimmune or allergic diseases. Thus, it is possible that TLR or NLR polymorphisms are important predisposition factors in ACD. The fact that the same pathways can be triggered by infectious agents and contact allergens opens the possibility that infection may contribute to ACD as a trigger factor. In our study we were able to render CHS resistant mice susceptible to CHS by triggering their endogenous

danger signal incompetent DC via TLR9, thus mimicking an infection. The detailed knowledge of the inflammatory pathways involved in ACD allows us to target these pathways, e.g. by interference with TLR, NLR and ROS signalling.

The identification of the mechanisms underlying the activation of innate immune, stress and antioxidant responses will result in new biomarkers that can be used for the development of *in vitro* assays for contact allergen identification and risk assessment. A great example is the Keap-1/Nrf2 system that seems suited to differentiate between contact allergens and irritants but is limited to cysteine reactive compounds. Lysine reactive compounds must be identified by other methods. These and other studies allow an optimistic outlook. It seems possible to recapitulate some of the very complex *in vivo* mechanisms that are happening in the skin and secondary lymph nodes in ACD using *in vitro* systems. Therefore, it should be possible by intensive research to develop a series of suitable assays covering different mechanistic steps of ACD that in their sum will allow the identification of contact allergens as well as potency assessment (Jowsey et al., 2006) and, most importantly, allow to reduce and eventually replace animal testing without compromising product and consumer safety (Basketter, 2008; dos Santos et al., 2009).

References

Ade, N., Leon, F., Pallardy, M. et al. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. *Toxicol. Sci. 107*, 451-460.

Basketter, D. A. (2008). Skin sensitization: strategies for the assessment and management of risk. *Br. J. Dermatol. 159*, 267-273.

dos Santos, G. G., Reinders, J., Ouwehand, K. et al. (2009). Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound. *Toxicol. Appl. Pharmacol. 236*, 372-382.

Edele, F., Esser, P. R., Lass, C. et al. (2007). Innate and adaptive immune responses in allergic contact dermatitis and autoimmune skin diseases. *Inflamm. Allergy Drug Targets 6*, 236-244.

Freudenberg, M. A., Esser, P. R., Jakob, T. et al. (2009). Innate and adaptive immune responses in contact dermatitis: analogy with infections. *G. Ital. Dermatol. Venereol. 144*, 173-185.

Dinkova-Kostova, A. T., Holtzclaw, W. D. and Kensler, T. W. (2005). The role of Keap1 in cellular protective responses. *Chem. Res Toxicol. 18*, 1779-1791.

Dudda, J. C. and Martin, S. F. (2004). Tissue targeting of T cells by DCs and microenvironments. *Trends. Immunol. 25*, 417-421.

Jowsey, I. R., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). *J. Appl. Toxicol. 26*, 341-350.

Kawai, T. and Akira, S. (2007). TLR signaling. *Semin. Immunol. 19*, 24-32.

Lambeth, J. D. (2004). NOX enzymes and the biology of reactive oxygen. *Nat. Rev. Immunol. 4*, 181-189.

Martin, S. F., Dudda, J. C., Bachtanian, E. et al. (2008) Toll-like receptor and IL-12 signaling control susceptibility to contact hypersensitivity. *J. Exp. Med. 205*, 2151-2162.

Martin, S. F., Merfort, I. and Thierse, H. J. (2006). Interactions of chemicals and metal ions with proteins and role for immune responses. *Mini. Rev. Med. Chem. 6*, 247-255.

Martin, S. F. (2004). T lymphocyte-mediated immune responses to chemical haptens and metal ions: implications for allergic and autoimmune disease. *Int. Arch. Allergy Immunol. 134*, 186-198.

Martinon, F., Mayor, A. and Tschopp, J. (2009). The inflammasomes: guardians of the body. *Annu. Rev. Immunol. 27*, 229-265.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature 449*, 819-826.

Mora, J. R. and von Andrian, U. H. (2006). T-cell homing specificity and plasticity: new concepts and future challenges. *Trends Immunol. 27*, 235-243.

Natsch, A. and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci. 102*, 110-119.

Seong, S. Y. and Matzinger, P. (2004). Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nat. Rev. Immunol. 4*, 469-478.

Yusuf, N., Nasti, T. H., Huang, C. M. et al. (2009). Heat shock proteins HSP27 and HSP70 are present in the skin and are important mediators of allergic contact hypersensitivity. *J. Immunol. 182*, 675-683.

Acknowledgments

The work summarised here was supported by a grant from the European Commission as part of the project "Novel Testing Strategies for In Vitro Assessment of Allergens (Sens-it-iv)", LSHB-CT-2005-018681 and by the Research Commission of the Medical Faculty, University Medical Center Freiburg.

Correspondence to

Stefan F. Martin, Ph.D.
Allergy Research Group
Department of Dermatology
University Medical Center Freiburg
Hauptstrasse 7
79104 Freiburg
Germany
e-mail: stefan.martin@uniklinik-freiburg.de

BS30: Neuroscience

In Vitro Models to Study Cell-Cell Interactions that Influence Developmental Neurotoxicity

Lucio G. Costa[1,2], Gennaro Giordano[1] and Marina Guizzetti[1]

[1]Dept. of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA; [2]Dept. of Human Anatomy, Pharmacology, and Forensic Science, University of Parma Medical School, Italy

Summary

Glial cells, astrocytes in particular, play an important role in neuronal development, functioning and survival, and in vitro co-culturing systems facilitate the study of astrocyte-neuron interactions. Here we describe how cerebellar astrocytes, co-cultured with cerebellar granule neurons, provide protection against oxidative stress-mediated neurotoxicity caused by a polybrominated diphenyl ether flame retardant. Protection is due to the ability of astrocytes to supply glutathione to neurons, thereby increasing their defense against oxidative stress. Astrocytes can also foster neuronal differentiation, as evidenced by their ability to promote axon and dendrite elongation in hippocampal neurons. This effect is mediated by the release by astrocytes of neuritogenic factors such as the extracellular matrix protein fibronectin. Certain neurotoxicants, such as ethanol and manganese, can "indirectly" affect neuronal differentiation by impairing astrocyte-neuron communication. These examples indicate that in vitro co- culturing systems such as those described provide useful tools to investigate the nature of such interactions and the underlying mechanisms.

Keywords: astrocytes, neurons, cell-cell interactions, developmental neurotoxicity

1 Introduction

The human nervous system is one of the most complex organ systems in terms of both structure and function. It contains billions of neurons, each forming thousands of synapses leading to a very large number of connections. It also contains perhaps ten-times more glial cells (astrocytes, oligodendrocytes, microglia) than neurons, which play important roles in the overall development and functioning of the nervous system (Barres, 2008).

Neurotoxicity can be defined as any adverse effect on the chemistry, structure, or function of the nervous system, during development or at maturity, induced by chemical or physical influences (Costa, 1998). This definition of neurotoxicity indicates a potential difference between the developing and the mature nervous system, to underscore the fact that developmental neurotoxicity is an important aspect of neurotoxicology. Most known human neurotoxicants are indeed developmental neurotoxicants (Grandjean and Landrigan, 2006). In most, but not all cases, the developing nervous system is more sensitive to adverse effects than the adult nervous system, as indicated, for example, by the most deleterious effects of ethanol, methylmercury, or lead when exposure occurs *in utero* or during childhood. Furthermore, the blood-brain-barrier (BBB), which protects the mature nervous system from the entry of a number of substances, appears to be poorly developed at birth and during the first few years of life.

The nervous system undergoes gradual development that continues well after birth in both animals and humans. While on one hand, the developing nervous system may more readily adapt to, or compensate for, functional losses as a result of a toxic insult, on the other hand, damage to the nervous system during key periods of brain development may result in long-term, irreversible damage (Costa, 1998). Evidence that developmental exposure to chemicals and drugs may alter behavioral functions in young animals began to emerge in the early 1970s. The field of developmental neurotoxicology thus evolved from the disciplines of neurotoxicology, developmental toxicology, and experimental psychology. In response to this issue, developmental neurotox-

icity testing guidelines were developed both in the USA and in Europe. Exposure of rodents to the test chemicals is from gestational day 6 to postnatal day 10 or 21 to the mother, thus ensuring exposure *in utero* and through maternal milk. Tests involve measurements of developmental landmarks and reflexes, motor activity, auditory startle test, learning and memory tests, and neuropathology.

In the past several years, the need to develop acceptable alternatives to conventional animal testing has been increasingly recognized by toxicologists. The concerns include the need to address problems related to the escalating costs and time required for toxicity assessments, the increasing number of chemicals being developed and commercialized, the need to respond to recent legislations (e.g. REACH (Registration Evaluation and Authorization of Chemicals) and the Cosmetics Directive (76/768/EEC) in the EU), and efforts aimed at reducing the number of animals used for toxicity testing (Costa, 1998; Bal-Price et al., 2008). Hence, efforts have been directed toward the development of alternative models, utilizing either mammalian cells *in vitro* or non mammalian model systems, which could serve as tools for neurotoxicity and developmental neurotoxicity testing.

In vitro testing procedures utilizing mammalian cells have two main purposes: a) investigate mode and/or mechanism of action of chemicals, particularly related to early, upstream events in the neurotoxic process; and b) screening of chemicals of unknown toxicity to flag compounds for further *in vitro* and *in vivo* studies. Several issues need to be considered when exploring potential *in vitro* models for neurotoxicity and developmental neurotoxicity. First, the nervous system comprises several types of cells (neurons, astrocytes, oligodendrocytes, Schwann cells, microglia and neural stem (progenitor) cells) (Costa et al., 2009). There are also different models that can be used; in increasing level of complexity they are immortalized cell lines, primary cells, cells in co-culture, aggregating cell cultures, and brain slices. Each model has its own advantages and disadvantages. For example, cell lines provide a defined and homogenous population of cells (usually clonal) derived from tumors or using oncogene-containing retroviruses, are easy to grow, divide rapidly, are available from various animal species including humans, and can be induced to differentiate. On the other hand, transformed cell lines may not exhibit the same phenotype of primary cells or may represent a specific cell sub-population. There is also increased genetic instability with increased number of passages; neurites may not represent true axons or dendrites, and, most importantly, cell-cell interactions are missing. More complex systems, such as aggregating brain cell cultures, have the advantage of providing a three-dimensional cell system containing all cell types and allowing cell-cell interactions, They also allow testing of multiple end-points in different cell types, including, for example, inflammatory responses. However, such cultures are difficult to prepare and maintain, there is a notable degree of variability between aggregates, and the anatomical organization of the tissue is missing (Costa et al., 2009). A compromise between *in vitro* systems utilizing only one specific cell type and more complex systems is represented by co-cultures of cells, most often neurons and glial cells though other co-cultures, including endothelial cells, for example, also are possible. Such co-culturing systems allow the study of cell-cell interactions, which are key to the development and functioning of the nervous system and which can be targeted by developmental neurotoxicants.

2 Roles of astrocytes

The name "glia" originates from Greek for glue. We now know, however, that glia do not just hold together neurons; instead, they play active and essential roles in brain development and function (He and Sun, 2007; Barres, 2008; Allen and Barres, 2009; Pfrieger, 2009). Astrocytes, the most abundant glial cells, have irregular star-shaped cell bodies and broad end-feet processes. They interact extensively with neurons, providing them with important structural and metabolic support (He and Sun, 2007). Astrocytes are broadly divided into two groups, protoplasmic astrocytes, found in the grey matter, and fibrous astrocytes in the white matter. The former are intimately associated with neuronal cell bodies and synapses, whereas the latter are associated with neuronal axons. There are also differences among astrocytes in a single grey matter region, but a complete understanding of the functional differences between the various types of astrocytes remains elusive (Allen and Barres, 2009).

One well known function of astrocytes is to help form, with endothelial cells, the blood-brain-barrier that prevents toxic substances in blood from entering the brain. A second important function of astrocytes is that of controlling the levels of certain neurotransmitters. For example, astrocytes facilitate rapid removal of synaptic glutamate following its release from the pre-synaptic terminal, thus preventing glutamate-mediated excitotoxicity, which can induce neuronal damage and death (He and Sun, 2007). Astrocytes also play a most relevant function in maintaining homeostasis of the extracellular environment for the proper functioning of neurons; during depolarization neurons release potassium ions into the intercellular space, and this excess potassium is taken up by astrocytes.

In addition to ion channels and transporters, astrocytes express receptors for neurotransmitters, neuromodulators, hormones, growth factors and cytokines, and they synthesize and secrete a number of growth and trophic factors, cytokines, and components of the extracellular matrix. There are a number of active exchanges between astrocytes and neurons that are essential to neuronal function. For example, glucose uptake from the circulation occurs mostly through GLUT1 transporters present in astrocytes, which also contain high amounts of glycogen. Lactate derived from astrocytic glucose or glycogen is transferred to neurons and used by these cells as source of energy. A fraction of the glutamate taken up by astrocytes is converted to glutamine by glutamine synthetase, which is abundant in astrocytes and absent in neurons. Glutamine is then released by astrocytes and taken up by neurons, where it is converted to glutamate by mitochondrial glutaminase.

The extensive cross-talk between astrocytes and neurons also involves the secretion by astrocytes of important factors that promote neuronal neuritogenesis and synaptogenesis. Important factors for the latter are astrocyte-derived cholesterol and thrombospondins (Pfrieger, 2009). Communications are bi-directional, in that astrocytes can receive signals from neurons through, for example, neurotransmitters. It should be also mentioned that astrocytes communicate with each other through waves of calcium ions, propagating information over large distances (Allen and Barres, 2009).

Because of the multiple and varied roles of astrocytes in neuronal development and function, impairment of astrocytic function is often associated with neurotoxicity, and these cells are increasingly recognized as important targets for neurotoxic chemicals (Aschner and Costa, 2004). In the remainder of this article, examples of astrocyte-neuron interaction leading to modulation of neuronal functioning will be discussed. Specifically, it will be indicated how astrocytes can protect neurons from damage caused by oxidative stress and how astrocytes contribute to neuritogenesis. The latter examples also provide evidence on how certain chemicals, by targeting astrocytes, can disrupt astrocyte-neuron communication, leading to neurotoxicity.

3 Astrocytes protect neurons against neurotoxicant-induced oxidative stress

Oxidative stress refers to the cytotoxic consequences of reactive oxygen species (ROS), which are generated as by-products of normal and aberrant metabolic processes that use molecular oxygen. The tripeptide GSH (□-glutamyl-cysteinylglycine) is one of the most abundant cellular thiols. GSH is a major player in cellular defense against ROS, because it nonenzymatically scavenges both singlet oxygen and hydroxyl radicals and is used by glutathione peroxidases and glutathione transferases to limit the levels of certain reactive aldehydes and peroxides within the cell. When ROS production exceeds the antioxidant defense capacity of the cell, oxidative stress ensues, leading to damage of DNA, proteins, and membrane lipids. In the brain, GSH is found at higher levels in astrocytes than in neurons (Rice and Russo-Menna, 1998), and astrocytes protect neurons against oxidative stress by providing GSH. Cystine is taken up by astrocytes and converted to cysteine, which serves for the synthesis of GSH; GSH is then released from astrocytes and metabolized by γ-glutamyl transpeptidase (GGT) to cysteinylglycine, from which cysteine is released by an endopeptidase located on the neuronal membrane surface. Cysteine is then taken up by neurons through the EAAC1 transporter (Aoyama et al., 2006) and utilized for GSH synthesis. Astrocytes have also been shown to induce transcriptional up-regulation of neuronal GSH through the release of still unidentified factors (Iwata-Ichikawa et al., 1999). Furthermore, GGT has also been shown to transfer the γ-glutamyl moiety of GSH to extracellular cystine to form γ-glutamylcystine, which can be taken up into cells, reduced to γ-glutamylcysteine, and used by GSH synthethase to synthesize GSH, thus bypassing glutamate-cysteine ligase (GCL; Anderson and Meister, 1983; Chinta et al., 2006), the first and rate-limiting enzyme in the synthesis of GSH (Dringen, 2000).

An example of how astrocytes can protect neurons from neurotoxicant-induced, oxidative stress-mediated neurotoxicity is provided by a recent study with the polybrominated diphenyl ether (PBDE) flame retardant mixture DE-71 (Giordano et al., 2009). PBDEs have become ubiquitous environmental contaminants, and human body burden is particularly high in infants and children because of exposure though breast milk and house dust (Costa and Giordano, 2007). Evidence from animal studies suggests that PBDEs are developmental neurotoxicants, as motor and cognitive dysfunctions have been found following pre-and/or post-natal exposures (Costa and Giordano, 2007). Though the exact mechanisms of PBDE developmental neurotoxicity are still elusive, DE-71 has been shown to cause oxidative stress and apoptotic cell death in mouse neurons, and these effects are modulated by intracellular GSH levels (Giordano et al., 2008). DE-71 caused apoptotic cell death of mouse cerebellar granule neurons (CGNs) with an IC_{50} of 7.2 μM. When CGNs were co-cultured with cerebellar astrocytes, the toxicity of DE-71 was decreased by fivefold, resulting in an IC_{50} of 38.7 μM. In contrast, cerebellar astrocytes from *Gclm* (-/-) mice, which lack the modifier subunit of GCL and have much lower GSH levels (4.3 vs 19.3 nmol/mg of protein in the wild-type), were much less effective at protecting CGNs toward DE-71 neurotoxicity; indeed, the IC_{50} of DE-71 increased by less than twofold, to 13.8 μM (Giordano et al., 2009). As indicated above, a mechanism by which astrocytes may protect neurons is by increasing their GSH content. Indeed, wild-type astrocytes, incubated with CGNs, increased the GSH content of neurons from 11.7 to 18.6 nmol/mg of protein. In contrast, *Gclm* (-/-) astrocytes were much less effective in this regard and could only increase neuronal GSH to 14.6 nmol/mg of protein. If GSH levels in CGNs were increased directly, by means of GSH ethyl ester (2.5 mM for 30 min), in the absence of astrocytes, there was a similar protection against DE-71 toxicity (Giordano et al., 2009).

This example indicates how astrocytes can provide a most relevant protection to neurons against compounds that induce oxidative stress-mediated neurotoxicity, by supplying GSH.

4 Astrocytes promote neuronal differentiation

Neurite outgrowth is a fundamental event in brain development as well as in the regeneration of damaged nervous tissue, and it is accomplished by signals from the extracellular space, such as extracellular matrix proteins, cell adhesion molecules, and soluble factors that can promote or inhibit the growth of the neurite (Kiryushko et al., 2004). Neuron-glia interactions play a key role in neurite outgrowth. Astroglial cells express and release molecules such as laminin-1, fibronectin, N-cadherin, and neural cell adhesion molecule (Price and Hynes, 1985; Liesi et al., 1986; Neugebauer et al., 1988), which can promote neurite outgrowth during development and during regeneration after a lesion (Tom et al., 2004). They also can release neurite-inhibiting factors, such as the chondroitin sulfate proteogly-

cans neurocan and brevican (Yamada et al., 1997; Asher et al., 2000). Such inhibitory proteoglycan released by glial cells during development contribute to the establishment of the correct architecture by preventing neurite growth in every direction (Snow et al., 1990) and, after brain lesion, contribute to the nonpermissive environment preventing regeneration of axons (Asher et al., 2000).

By targeting astrocytes, endogenous substances can foster astrocyte-neuron communications. For example, thyroid hormone (T3) has been shown to stimulate the release of epidermal growth factor from astrocytes, leading to neuritogenesis of cerebellar neurons (Martinez and Gomes, 2002), and vasoactive intestinal polypeptide stimulates the release by astrocytes of activity-dependent neurotrophic factor, promoting morphological and functional differentiation of hippocampal neurons (Blondel et al., 2000).

Astrocytes express on their cell surface several receptors for neurotransmitters, including acetylcholine muscarinic receptors (Guizzetti et al., 1996). As such receptors activate a robust signaling involving the generation of lipid messengers and the activation of kinases and transcription factors, it was recently hypothesized that activation of muscarinic receptors in astrocytes may induce changes in the astrocyte extracellular environment leading to neuronal differentiation (Guizzetti et al., 2008). When cortical (or hippocampal) astrocytes were treated with the cholinergic agonist carbachol for 24 h, followed by wash-out, and were then co-cultured with hippocampal neurons, neuronal differentiation was stimulated. Specifically, carbachol-treated astrocytes induced a 2-3-fold increase in the length of the longest neurite (the axon, as determined by Tau-1 staining), and a 2-fold increase in the length of minor neurites. Such effect was due to activation of M3 muscarinic receptors on astrocytes.

A separate study had characterized the astrocyte secretoma by shotgun proteomics (Moore et al., 2009). A total of 302 proteins secreted by astrocytes were identified; of these, 133 were recognized as proteins with extracellular localization, and they were further characterized. Of interest to the present discussion is that proteins involved in neuronal differentiation were included in this list; these include, for example, fibronectin, thrombospondin-1, PAI-1 (plasminogen activator inhibitor-1), and Plau (plasminogen activator urokinase). Cholinergic stimulation of astrocytes modulated the release of a number of proteins, as indicated by semi-quantitative proteomics analysis and by Western blot. In particular, a 24 h stimulation of astrocytes with carbachol increased the expression of fibronectin and of laminin-1 (Guizzetti et al., 2008; Moore et al., 2009). This was verified by confocal miscroscopy, by analysis of fibronectin and laminin-1 protein by Western blot in the astrocyte lysate and in the medium, and by measurements of their mRNAs by Taq-PCR (Guizzetti et al., 2008). Carbachol also increased the levels of PAI-1, which inhibits laminin-1 and fibronectin degradation and can therefore contribute to increasing their extracellular levels in addition to exerting a neuritogenic effect on its own.

Altogether, these results indicate that activation of muscarinic M3 receptors on astrocytes promotes the release of neuritogenic factors that cause differentiation of hippocampal neurons. Fibronectin appears to play a primary role in this process. Indeed, fibronectin function-inhibiting antibodies completely prevented the neuritogenic effect of carbachol-treated astrocytes.

Given the described action of astrocyte muscarinic receptors in promoting neuritogenesis and the known ability of ethanol to inhibit muscarinic receptor signal transduction (Costa and Guizzetti, 1999), it was then hypothesized that, by targeting astrocytes, ethanol may "indirectly" affect neuronal differentiation. Ethanol is a known developmental neurotoxicant, and *in utero* exposure results in a number of CNS abnormalities, which are the hallmark of Fetal Alcohol Syndrome. Results of these experiments indeed indicate that when astrocytes were incubated with carbachol for 24 h in the presence of low alcohol concentrations (25-100 mM), followed by washout, the neuritogenic effect of astrocytes was significantly reduced in a concentration-dependent manner (Guizzetti et al., 2010). For example, 50 mM ethanol decreased average axon length from 157.2 μm to 65.5 μm, average neurite length from 41.3 to 26.3 μm, and number of neurites/cell fom 8.0 to 4.2 (Guizzetti et al., unpublished). This can be ascribed, at least in part, to a reduced expression and release of fibronectin by astrocytes exposed to carbachol in the presence of ethanol. Further experiments have also shown that a probable target for ethanol action appears to be phospholipase D. Ethanol is known to inhibit phospholipase D in astrocytes by substituting for water in a transphosphatidylation reaction (Guizzetti et al., 2004). 1-Butanol, an inhibitor of phospholipase D, had effects similar to ethanol with regard to astrocyte-mediated neurite outgrowth and fibronectin expression, while tert-butanol, which does not inhibit phospholipase D, was devoid of effects, providing further evidence of phospholipase D as a likely target for ethanol's action (Guizzetti et al., unpublished).

Another example of the relevance of astrocyte-neuron interactions in modulating neurotoxicity and developmental neurotoxicity can be found in recent studies on manganese (Giordano et al., 2009). Manganese (Mn) is a neurotoxic metals, and has been shown to contribute to neurological and behavioral abnormalities in children as well (Levy and Nassetta, 2003; Erikson et al., 2007). Mn easily enters the brain and, at the cellular level, it accumulates in astrocytes, which have high capacity transporters for this metal (Ashner et al., 1999). Concentrations of Mn in astrocytes can be 50-60-fold higher than in neurons (Ashner et al., 1999), and brain concentrations of Mn of 200-300 μM have been reported (Erikson et al., 2004). By targeting astrocytes, Mn may indirectly affect neuronal differentiation.

The experiments involving carbachol and ethanol described above were carried out with 24 h incubations of astrocytes (previously treated with carbachol/ethanol or untreated) with hippocampal neurons. By prolonging the time of co-culture with neurons (to 48 or 72 h), untreated astrocytes can promote the differentiation of hippocampal neurons which elongate axon and neurites. When astrocytes are incubated with Mn (50-500 μM), followed by wash-out, before astrocyte and neurons are co-cultured, the ability of astrocytes to promote neurite outgrowth is significantly reduced. For example, at 200 μM, a concentration that is not cytotoxic in astrocytes, axon length is 91 um vs 155 μm in control (Giordano et al., 2009). Given the

previous finding of a primary role of fibronectin in modulating astrocytes' neuritogenic effect, the effect of Mn on fibronectin was then investigated. Mn was found to decrease the expression of fibronectin protein and mRNA in a concentration-dependent fashion.

Mn is known to cause oxidative stress in astrocytes (Erikson et al., 2004; Milatovic et al., 2007; Giordano et al., 2009), and this action may contribute to its inhibition of fibronectin expression and hence, of the neuritogenic action of astrocytes. This hypothesis was confirmed by a series of findings. First, two antioxidants (melatonin and N-t-butyl alpha-phenylnitrone) antagonized the effect of Mn on neuritogenesis and on fibronectin expression. Second, increasing astrocytic glutathione levels by using glutathione ethyl ester, similarly antagonized the effects of Mn. Third, depletion of astrocytic glutathione with buthionine sulfoximine potentiated the effect of Mn. Fourth, two other compounds known to cause oxidative stress (H_2O_2 and 2,3-dimethoxy-1,4-naphtoquinone), when incubated with astrocytes at sub-cytotoxic concentrations, decreased fibronectin expression and inhibited the ability of astrocytes to promote neuronal differentiation (Giordano et al., 2009).

These findings indicate that Mn interferes with an important aspect of glial-neuronal interactions, i.e. the ability of astrocytes to promote neuritogenesis in hippocampal neurons. The final toxic effect of Mn on neurons, was thus "indirect" and mediated by astrocytes. Indeed, direct exposure of neurons to Mn did not alter the length of the axon and of neurites, causing only a slight increase at high concentrations.

5 Conclusions

Cell-cell interactions play a fundamental role in brain development and functioning. In particular, astrocytes influence the development, functioning, and well-being of neurons. The examples discussed in this paper indicate that astrocytes can protect neurons against neurotoxicant-induced oxidative stress and that they can foster neuronal differentiation. By targeting astrocytes, and thus disrupting astrocyte-neuron interactions, neurotoxicants can adversely affect neuronal development and function. *In vitro* co-culturing systems such as those described provide useful tools to investigate the nature of such interactions and the underlying mechanisms.

References

Allen, N. J. and Barres, B. A. (2009). Glia-more than just brain glue. *Nature 457*, 675-677.

Anderson, M. E. and Meister, A. (1983). Transport and direct utilization of γ-glutamylcyst(e)ine for glutathione synthesis. *Proc. Natl. Acad. Sci. 80*, 707-711.

Aoyama, K., Suh, S. W., Hamby, A. M. et al. (2006). Neuronal glutathione deficiency and age-dependent neurodegeneration in the EAAC1 deficient mouse. *Nature Neurosci. 9*, 119-126.

Aschner, M. and Costa, L. G. (eds.). (2004). *The Role of Glia in Neurotoxicity*. 2nd ed. CRC Press, Boca Raton; pp.456.

Aschner, M., Vrana, K. E. and Zhang, W. (1999). Manganese uptake and distribution in the central nervous system (CNS). *Neurotoxicology 20*, 173-180.

Asher, R. A., Morgenstern, D. A., Fidler, P. S. et al. (2000). Neurocan is upregulated in injured brain and in cytokine-treated astrocytes. *J. Neurosci. 20*, 2427-2438.

Bal-Price, A. K., Sunol, C., Weiss, D. G. et al. (2008). Application of *in vitro* neurotoxicity testing for regulatory purposes: Symposium III summary and research needs. *Neurotoxicology 29*, 520-531.

Barres, B. A. (2008). The mistery and magic of glia: a perspective on their roles in health and disease. *Neuron 60*, 430-440.

Blondel, O., Collin, C., McCarran, W. J. et al. (2000). A glia derived signal regulating neural differentiation. *J. Neurosci. 20*, 8012-8020.

Chinta, S. J., Kumar, J. M., Zhang, H. et al. (2006). Up-regulation of γ-glutamyl transpeptidase activity following glutathione depletion has a compensatory rather than inhibitory effect on mitochondrial complex I activity: implications for Parkinson's disease. *Free Radic. Biol. Med. 40*, 1557-1563.

Costa, L. G., Giordano, G. and Guizzetti, M. (2010). Predictive models for neurotoxicity assessment. In: J.J. Xu and L. Urban (eds.) *Predictive Toxicology in Drug* Safety. Cambridge: University Press, in press.

Costa, L. G. and Giordano, G. (2007). Developmental neurotoxicity of polybrominated diphenyl ether (PBDE) flame retardants. *Neurotoxicology 28*, 1047-1067.

Costa, L. G. and Guizzetti, M. (1999). Muscarinic cholinergic receptor signal transduction as a potential target for the developmental neurotoxicity of ethanol. *Biochem. Pharmacol. 57*, 721-726.

Costa, L. G. (1998). Neurotoxicity testing: a discussion of in vitro alternatives. Env. Health Persp. 106, Suppl. 2, 505-510.

Dringen, R. (2000). Metabolism and functions of glutathione in brain. *Progr. Neurobiol. 62*, 649-671.

Erikson, K. M., Thompson, K., Aschner, J. L. and Aschner, M. (2007). Manganese neurotoxicity: a focus on the neonate. *Pharmacol. Ther. 113*, 369-377.

Erikson, K. M., Dobson, A. W., Dorman, D. C. and Aschner, M. (2004). Manganese exposure and induced oxidative stress in the rat brain. *Sci. Total Environ. 334-335*, 409-416.

Giordano, G., Kavanagh, T. J. and Costa, L. G. (2009). Mouse cerebellar astrocytes protect cerebellar granule neurons against toxicity of the polybrominated diphenyl ether (PBDE) mixture DE-71. *Neurotoxicology 30*, 326-329.

Giordano, G., Pizzurro, D., VanDeMark, K. L. et al. (2009). Manganese inhibits the ability of astrocytes to promote neuronal differentiation. *Toxicol. Appl. Pharmacol.240*, 226-235.

Giordano, G., Kavanagh, T. J. and Costa, L. G. (2008). Neurotoxicity of a polybrominated diphenyl ether mixture (DE-71) in mouse neurons and astrocytes is modulated by intracellular glutathione levels. *Toxicol. Appl. Pharmacol. 232*, 161-168.

Grandjean, P. and Landrigan, P. J. (2006). Developmental neurotoxicity of industrial chemicals. *Lancet 368*, 2167-2178.

Guizzetti, M., Moore, N.H., Giordano, G. et al. (2010). Ethanol inhibits neuritogenesis induced by astrocyte muscarinic receptors. *Glia*, in press.

Guizzetti, M., Moore, N. H., Giordano, G. and Costa, L. G. (2008). Modulation of neuritogenesis by astrocyte muscarinic receptors. *J. Biol. Chem. 283*, 31884-31897.

Guizzetti, M., Thompson, B. D., Kim, Y. et al. (2004). Role of phospholipase D signaling in ethanol induced inhibition of carbachol-stimulated DNA synthesis of 1321N1 astrocytoma cells. *J. Neurochem. 90*, 646-653.

Guizzetti, M., Costa, P., Peters, J. and Costa, L. G. (1996) Acetylcholine as a mitogen: muscarinic receptor – mediated proliferation of rat astrocytes and human astrocytoma cells. *Eur. J. Pharmacol. 297*, 265-273.

He, F. and Sun, Y. E. (2007). Glial cells more than support cells? *Int. J. Biochem. Cell Biol. 39*, 661-665.

Iwata-Ichikawa, E., Kondo, Y., Miyazaki, I. et al. (1999). Glial cells protect neurons against oxidative stress via transcriptional up-regulation of glutathione synthesis. *J. Neurochem. 72*, 2334-2344.

Kiryushko, D., Berezin, V. and Bock, E. (2004). Regulators of neurite outgrowth: role of cell adhesion molecules. *Ann. N.Y. Acad. Sci. 1014*, 140-154.

Levy, B. S. and Nassetta, W. J. (2003). Neurologic effects of manganese in humans: a review. *Int. J. Occup. Environ. Health 9*, 153-163.

Liesi, P., Kirkwood, T. and Vaheri, A. (1986). Fibronectin is expressed in astrocytes cultured from embryonic and early postnatal rat brain. *Exp. Cell Res. 163*, 175-185.

Martinez, R. and Gomes, F. C. (2002). Neuritogenesis induced by thyroid hormone-treated astrocytes is mediated by epidermal growth factor/mitogen-activated protein kinase-phosphatidylinositol-3-kinase pathways and involves modulation of the extracellular matrix. *J. Biol. Chem. 277*, 49311-49318.

Milatovic, D., Yin, Z., Gupta, R. C. et al. (2007). Manganese induces oxidative impairment in cultured rat astrocytes. *Toxicol. Sci. 98*, 198-205.

Moore, N. H., Costa, L. G., Shaffer, S. A. et al. (2009). Shotgun proteomics implicates extracellular matrix proteins and protease systems in neuronal development induced by astrocyte cholinergic stimulation. *J. Neurochem. 108*, 891-908.

Neugebauer, K. M., Tomaselli, K. J., Lilien, J. and Reichardt, I. F. (1988). N-cadherin, NCAM, and integrins promote retinal neurite outgrowth on astrocytes in vitro. *J. Cell Biol. 107*, 1177-1187.

Pfrieger, F. W. (2009). Roles of glial cells in synapse development. *Cell. Mol. Life Sci. 66*, 2037-2047.

Price, J. and Hynes, R. O. (1985). Astrocytes in culture synthesize and secrete a variant form of fibronectin. *J. Neurosci. 5*, 2205-2111.

Rice, M. E. and Russo-Menna, I. (1998). Differential compartmentalization of brain ascorbate and glutathione between neurons and glia. *Neuroscience 82*, 1213-1223.

Snow, D. M., Lemmon, V., Carrino, D. A. et al. (1990). Sulfated proteoglycans in astroglial barriers inhibit neurite outgrowth in vitro. *Exp. Neurol. 109*, 111-130.

Tom, V. J., Doller, C. M., Malouf, A. T. and Silver, J. (2004). Astrocyte-associated fibronectin is critical for axonal regeneration in adult white matter. *J. Neurosci. 24*, 9282-9290.

Yamada, H., Fredette, B, Shitara, K. et al. (1997). The brain chondroitin sulfate proteoglycan brevican associates with astrocytes ensheathing cerebellar glomeruli and inhibits neurite outgrowth from granule neurons. *J. Neurosci. 17*, 7784-7795.

Acknowledgements

Studies by the authors were supported in part by grants from NIH (AA08154, ES07033) and DOD (W81XWH-05-1-0239).

Correspondence to

Dr. Lucio G. Costa
University of Washington
Dept. of Environmental and Occupational Health Sciences
4225 Roosevelt #100
Seattle, WA 98105, USA
e-mail: lgcosta@u.washington.edu

Breakout Sessions

EB7: Status report on OSIRIS

In Vitro / In Vivo Relationship in the Light of Toxcast Phase 1

Romualdo Benigni[1], Cecilia Bossa[1], Alessandro Giuliani[1] and Olga Tcheremenskaia[1]

[1]Environment and Health Department, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

Summary

With the aims of shortening times of toxicity testing, protecting animal health and welfare and saving money, research on the "Three Rs" (replacement, reduction and refinement of animal testing) has been going on for years. Recently, new impetus has been given by the ToxCast project. ToxCast adopts a pathway-based screening, consisting in the analysis of perturbations provoked by chemicals to biochemical and biological pathways supposed to be critical to toxicity. Such perturbations are studied in isolated systems in vitro (both cell-free and cell-based) with the use of modern high-throughput screening (HTS) omics techniques. We analysed the results of ToxCast Phase 1, mainly with the perspective of replacing in vivo assays with in vitro ones. It appears that the correlation in vitro / in vivo is extremely low, or absent, depending on the in vivo toxicity endpoint considered. This evidence is in agreement with findings from other fields, ranging from research on drug design with intensive use of omics technologies to more traditional research on alternative tests for regulatory purposes: isolated systems in vitro - when perturbed by chemicals - respond in a way that is radically different from how they respond when they are integrated into whole organisms. From a practical point view this means that we are still quite far from being able to replace the classical animal toxicity assays with stand-alone in vitro alternatives (first of the "Three Rs"), whereas progress seems closer in terms of the other Rs (reduction, refinement).

Key words: alternatives, Three Rs, toxicity testing, reductionism, systems biology

1 Introduction

Toxicology today faces challenges that are both difficult and exciting. It is estimated that around 74,000 chemicals in the USA (Schmidt, 2009) and 30,000 chemicals in Europe (Pedersen et al., 2003)are in use without adequate toxicological information. Since the task of testing them systematically with classical animal assays seems to exceed the present capabilities, there is a strong societal pressure to investigate and identify suitable alternative testing methods. As an example, the new European legislation on chemicals, REACH, explicitly mentions the possibility of using both experimental (*in vitro*) and non testing (structure-activity relationships, read-across, categories) alternative methods (Worth et al., 2007).

With the aims of shortening times of toxicity testing, protecting animal health and welfare and saving money, research on the "Three Rs" (replacement, reduction and refinement of animal testing) has been going on for years (EFSA, 2009). Recently, new impetus to this research has been given by the ambitious Tox21, or Toxcast project (Dix et al., 2007; Schmidt, 2009). This derives from an agreement between a number of US agencies: Environmental Protection Agency (EPA), National Toxicology Program (NTP) and National Chemical Genomics Center (NCGC) (Schmidt, 2009). The paradigm adopted by Toxcast is a pathway-based screening, and is radically different from traditional animal testing. Pathway-based screening consists in the analysis of perturbations provoked by chemicals to biochemical and biological pathways supposed to be critical to toxicity. Such perturbations are studied in isolated systems *in vitro* (both cell-free and cell-based) with the use of modern high-throughput screening (HTS) omics techniques. The strategy of Toxcast is to validate HTS assays by compar-

ing their results with animal toxicity measures, and eventually with human adverse effects, on selected panels of chemicals.

Recently, a first round of Toxcast (Phase 1) was concluded. More than 500 different HTS assays (both cell-free and cell-based) have been applied to 320 compounds (309 unique chemicals), mostly agrochemicals, for which animal toxicity data already existed (rodent carcinogenicity, developmental and multigeneration toxicity). The data have been made publicly available for analysis: a number of studies were presented and discussed at a Toxcast Data Analysis Summit (http://www.epa.gov/NCCT/toxcast/summit.html).

We have exploited the availability of data to investigate the contribution of the Toxcast Phase 1 HTS measures to the most pressing issue among the Three Rs, i.e. replacement of animal testing.

A detailed description of the ToxCast program and of the ToxCast Data Analysis Summit is on: http://www.epa.gov/ncct/toxcast/summit.html

The data analysed in this paper were downloaded through the ToxCast Environmental Science Connector and were contained in the package: ToxCast_20090406.zip. Briefly, the data refer to 320 compounds (309 unique chemicals), mostly agrochemicals. Each chemical has the following toxicological data:

a) 76 *in vivo* bioassays (target organs (chronic toxicity), reproductive, developmental, carcinogenicity);
b) 524 *in vitro* assays (9 *in vitro* assay providers for a total of 285 cell-based and 239 cell-free toxicity measures).

The names of the chemicals and assays can be found on the ToxCast website.

For the present large scale analysis, activity (positive / negative) data were used. The data were analysed with the SAS/

2 Data and Methods

Table 1

The table reports the most representative assays for each of the 6 clusters of *in vitro* assays. The clusters are identified by applying k-means cluster analysis to the profiles of responses of the 524 *in vitro* ToxCast assays to the 309 chemicals. The assays shown are those most representative, i.e. most correlated with own cluster (see own-cluster correlation in last column). Each assay is identified by the code given in the Toxcast Phase 1 exercise, as reported on the official website (http://www.epa.gov/ncct/toxcast/summit.html) The names of the clusters have been attributed by us by inspection of the composition of each cluster: this has been possible because the most representative assays in each cluster appear to share a similar biological significance. The reader can easily check this by referring to the description of the individual assays on the ToxCast website.

Cluster 1: Cell growth, cell adhesion and shape, etc. (immune-related)

BSK_SAg_IL8	0.5161
BSK_SAg_Eselectin	0.5262
BSK_hDFCGF_Proliferation	0.5396
CLM_CellLoss_72	0.5478
BSK_SAg_Proliferation	0.6876

Cluster 2: Signalling (cell cycle, apoptosis…)

NVS_ENZ_hMAPK3	0.8023
NVS_ENZ_hPKBa	0.8023
NVS_ENZ_hPKBb	0.8023
NVS_ENZ_hSRC	0.8023
NVS_ENZ_hSRC_Activator	0.8023
NVS_ENZ_hAurA	0.8115
NVS_ENZ_hMAPKAPK5	0.8115
NVS_ENZ_hPTP1b	0.8115
NVS_ENZ_hCK1D	0.8463
NVS_ENZ_hMet	0.8463
NVS_ENZ_hMsk1	0.8463
NVS_ENZ_hSGK1	0.8463
NVS_ENZ_hPTPSHP1	0.862

Cluster 3: Nervous system factors

NVS_GPCR_hDRD2s	0.5433
NVS_GPCR_hAdrRa2C	0.5499
NVS_GPCR_hM1	0.5538
NVS_GPCR_hM5	0.5572
NVS_GPCR_bNPYNon_Selective	0.6587
NVS_GPCR_hAdrRb1	0.6662
NVS_GPCR_hDRD4_4	0.7027

Cluster 4: Metabolic factors

NVS_ADME_rCYP3A1	0.5095
NVS_ADME_hCYP2B6	0.5526
NVS_ADME_hCYP2C9	0.559
NVS_ADME_rCYP2C11	0.5825
NVS_ADME_rCYP2B1	0.5826
NVS_ADME_hCYP2C18	0.5904

Cluster 5: Cell growth, inflammation

CLM_p53Act_1	0.5162
BSK_3C_SRB	0.5189
CLM_OxidativeStress_1	0.524
BSK_hDFCGF_SRB	0.5432
BSK_SAg_SRB	0.5878
BSK_LPS_SRB	0.6085
BSK_4H_SRB	0.6515

Cluster 6: Transcription factors, gene activation

ATG_EGR_CIS	0.5132
ATG_Oct_MLP_CIS	0.515
ATG_HIF1a_CIS	0.5229
ATG_VDRE_CIS	0.5249
ATG_BRE_CIS	0.5583
ATG_PBREM_CIS	0.5679
ATG_MRE_CIS	0.6021

STAT statistical software (http://www.sas.com/technologies/analytics/statistics/stat/index.html).

3 Results

The availability of large numbers of *in vitro* (n=524) and *in vivo* (n=76) measures permits the exploration of an extremely vast range of *in vitro* / *in vivo* comparisons. However, this can easily give rise to the observation of correlations purely due to chance. As a matter of fact, the analysis of the distribution of correlations has shown that the fraction of *in vitro* / *in vivo* pairs with statistically significant positive agreement is identical to the significant fraction expected by chance (Rathman & Yang, 2009).

Thus, averting chance correlations is a crucial pre-requisite. As a consequence, the first step of our analysis was aimed at "cleaning" the Toxcast *in vitro* data. The 524 *in vitro* measures were subjected to k-means cluster analysis, which pointed to an optimal partition into 6 clusters. The total variance obtained was 0.25, thus indicating the presence of a large proportion of uncorrelated (singular) results. The 6 clusters partition was confirmed by principal component analysis (results not shown). Table 1 reports representative assays in each cluster. The inspection of the composition of the various clusters indicates that the assays that characterise each cluster have a similar mechanistic meaning. Thus it is possible to recognise the biological significance of the clusters and give them a name (see Table 1).

Table 2a
Five most representative *in vitro* assays in each cluster were used to model carcinogenicity in mouse and rat, respectively. The operation was performed separately for each cluster by applying linear discriminant analysis. The fitting of each model is measured by squared canonical correlation (SQCC).

Table 2b
All representative *in vitro* assays (five from each cluster, 30 total) were subjected to linear discriminant analysis to model mouse and rat carcinogenicity. The resultant SQCC is given.

Tab. 2a: Correlation of clusters of *in vitro* assays with rodent carcinogenicity

Assays	Squared Canonical Correlation (SQCC)	
	Mouse	Rat
Cluster 1	0.02	0.03
Cluster 2	-	-
Cluster 3	-	-
Cluster 4	0.04	0.01
Cluster 5	0.01	-
Cluster 6	-	0.03

Tab. 2 b: Global correlation of *in vitro* assays with rodent carcinogenicity

Assays	SQCC	
	Mouse	Rat
30 representatives	0.04	0.09

In vitro assays as predictors of rodent carcinogenicity
As next step, the correlation between *in vitro* assays and rodent carcinogenicity was assessed. Five central (i.e. mostly correlated with own cluster) *in vitro* assays from each cluster were subjected to discriminant analysis. The operation was repeated separately with each cluster. Table 2a shows that the maximum correlation was 0.04 (squared canonical correlation, corresponding to the proportion of variance explained).

The combination of all 30 representative assays from the 6 clusters improved the squared canonical correlation, but only up to 0.04 – 0.09 (Table 2b).

Overall, the Toxcast *in vitro* assays are very poor predictors of rodent carcinogenicity for the 309 Toxcast Phase 1 chemicals.

Other in vitro / in vivo correlations
The availability of a rich data set in Toxcast provides the opportunity to explore the *in vitro* / *in vivo* relationship based on other *in vivo* toxicity endpoints besides rodent carcinogenicity. We considered *Rat Cholinesterase Inhibition* and *Mouse Liver Necrosis* endpoints, since the preliminary analysis of *in vitro* results pointed to the existence of clusters of *in vitro* assays ideally directly related – in mechanistic terms - to the above *in vivo* endpoints.

Rat Cholinesterase Inhibition was modelled with assays in the "Nervous system factors" cluster, and *Mouse Liver Necrosis* was modelled with assays in the "Signalling (cell cycle, apoptosis…)" cluster.

Application of discriminant analysis showed that in neither case a correlation between *in vitro* and *in vivo* measures (even though mechanistically related) was apparent (analytical results not shown).

In vivo versus in vivo
Since the Toxcast chemicals are characterised also in terms of developmental and multigeneration toxicity measures, it was

Table 3
In vivo toxicity measures were used to model mouse carcinogenicity by linear discriminant analysis. The toxicity measures refer to developmental (DEV) and multigeneration (MGR) toxicity, measured in rat and rabbit. The table reports the variables accepted into the model with cumulative SQCC. The codes of the various toxicity measures – though mostly self-explanatory – are given on the ToxCast website.

Tab. 3: Correlation of *in vivo* toxicity measures with mouse carcinogenicity

Variables entered into the model	SQCC (cumulative)
DEV_Rat_Urogenital_Ureteric	0.06
DEV_Rat_Skeletal_Appendicular	0.12
DEV_Rat_PregnancyRelated_Maternal	0.13
MGR_Rat_Ovary	0.15
MGR_Rat_Uterus	0.17
DEV_Rabbit_Cardiovascular_MajorV	0.18
DEV_Rabbit_Urogenital_Renal	0.20

possible to study the degree of correlation between rodent carcinogenicity and these other *in vivo* toxicity data. This was performed by applying discriminant analysis.

Table 3 shows that the combination of a number of *in vivo* toxicity assays correlate with *Mouse Carcinogenicity*, with an overall squared canonical correlation of 0.20. Similarly, a combination of *in vivo* toxicity measures correlates with *Rat Carcinogenicity*, with an overall squared canonical correlation of 0.17 (Table 4).

It should be emphasised that: a) these *in vivo* / *in vivo* correlations are more than double in magnitude compared to *in vitro* / *in vivo* correlations (Table 2); b) no obvious mechanistic link between carcinogenicity endpoint and the toxicity measures in Tables 3 and 4 can be hypothesised.

4 Discussion

The analysis of Toxcast Phase 1 data provided clear-cut indications regarding the relationship between *in vitro* assays and *in vivo* toxicological endpoints: the correlation is extremely weak (*Mouse* or *Rat Carcinogenicity*), or completely absent (*Rat Cholinesterase Inhibition* and *Mouse Liver Necrosis*). This evidence is even more striking in light of the assumed mechanistic links between the above *in vitro* and *in vivo* endpoints. On the contrary, a fair proportion of the rodent carcinogenicity effects (around 20%) can be modelled by combinations of other *in vivo* toxicity effects (Tables 3 and 4), for which no obvious mechanistic link with carcinogenicity can be hypothesised. An immediate explanation is that the pharmacokinetic properties that are typical for the whole animal and influence the various toxicity endpoints are the same. In this sense, the 20% figure can be considered a rough estimate of the contribution of adsorption, distribution, metabolism and excretion (ADME) effects to the final carcinogenicity phenotype.

However interesting, the Toxcast Phase 1 results are only one isolated observation, so it is crucial to put them into a wider perspective and check if they are supported - or contradicted - by other evidence.

A field where there has been an intensive use of genomics and proteomics technologies in recent years is that of drug design. Wide panels of HTS tools are used to identify promising compounds to be studied to a deeper extent in further steps of drug development, as well as to predict undesirable toxic effects early in the process.

It appears that in recent years the number of new drugs entering the US market has declined sharply, while spending by the pharmaceutical industry on research and development has steadily increased (Young, 2009). The two single most important reasons for attrition in clinical development are: a) lack of efficacy and b) clinical safety or toxicology, which each account for 30% of failures (Hopkins, 2008). Failures have been largely ascribed to the lack of correlation between effects observed in isolated receptors *in vitro* and those observed in whole animals and in humans (for insightful discussions see (Hopkins, 2008) and (MacDonald & Robertson, 2009)).

Further evidence on the *in vitro* / *in vivo* relationship comes from a recent position paper written by a committee of experts under the aegis of the European Food Safety Agency (EFSA) (EFSA, 2009). The paper provides a very detailed review of state-of-the-art approaches incorporating replacement, reduction and refinement of animal testing and examines the whole

Table 4

In vivo toxicity measures were used to model rat carcinogenicity by linear discriminant analysis. The toxicity measures refer to developmental (DEV) and multigeneration (MGR) toxicity, measured in rat and rabbit. The table reports the variables accepted into the model, with cumulative SQCC. The codes of the various toxicity measures – though mostly self-explanatory - are given on the ToxCast website.

Tab. 4: Correlation of *in vivo* toxicity measures with rat carcinogenicity

Variables entered into the model	SQCC (cumulative)
DEV_Rabbit_Orofacial_JawHyoid	0.05
DEV_Rat_General_GeneralFetalPath	0.07
MGR_Rat_Mating	0.09
MGR_Rat_GestationalInterval	0.11
MGR_Rat_Spleen	0.13
DEV_Rabbit_Skeletal_Appendicular	0.14
DEV_Rat_Urogenital_Genital	0.15
MGR_Rat_Kidney	0.17

Table 5

The table is an elaboration of the main results of an EFSA report on the state-of-the-art of replacing, reducing and refining animal toxicity assays. It appears that only skin irritation and corrosion have valid replacement by in vitro assays. In addition, in vitro assays contribute to reduce animal testing for four toxicity endpoints. No replacement or reduction is possible today for the other toxicity endpoints.

Tab. 5: Replacement and reduction of *in vivo* toxicological assays by *in vitro* assays: state-of-the-art

In vivo toxicity	
• Toxicokinetic	
• **Acute toxicity**	***Reduction***
• **Skin irritation and corrosion**	*Replacement*
• Skin sensitisation	
• **Eye irritation**	***Reduction***
• Acute systemic and local toxicity	
• **Genotoxicity**	***Reduction***
• **Carcinogenicity**	***Reduction***
• Repeated dose toxicity	
• Reproduction	
• Developmental toxicity	
• Ecotoxicity	

range of toxicological endpoints, from toxicokinetic studies to genotoxicity, reproduction and developmental toxicity. Table 5 shows a summary of the conclusions, indicating which types of animal toxicity assays can be either replaced or substantially reduced with validated *in vitro* approaches.

The EFSA survey concludes that only skin irritation and corrosion testing in animals can confidently be replaced by *in vitro* alternatives. Skin irritation can be replaced by an *in vitro* method using human skin models. The information on skin corrosion can be obtained by a number of *in vitro* tests relying on measurement of for example transcutaneous electrical resistance and/or pH, interaction with a bio-membrane or a human skin model. It should be emphasised that both are local effects with no systemic response involved. On the contrary, skin sensitization testing, where a systemic response is elicited, cannot be replaced or reduced by in vitro assays.

For a number of other endpoints (acute toxicity, eye irritation, genotoxicity, carcinogenicity), *in vitro* assays can help to direct the animal testing, thus contributing to diminishing the number of animals used and the animals' suffering. On the other hand, the development of alternative methods has shown to be more difficult for a range of other toxicological endpoints, such as toxicokinetics, skin sensitization, acute systemic and local toxicity, repeated dose toxicity, reproduction and developmental toxicity studies. The same applies for ecotoxicity endpoints, such as acute and chronic toxicity in fish and birds, and bioconcentration in fish (EFSA, 2009).

All the above evidence, ranging from research on drug design with intensive use of omics technologies to more traditional research on alternative tests for regulatory purposes, converge with the evidence provided by the Toxcast data: isolated systems *in vitro* - when perturbed by chemicals - respond in a way radically different from how they respond when they are integrated into whole organisms. The patterns of interactions among the different systems and organs in the whole organism appear to be the major determinants of the response of the organisms, including toxicity, and – obviously - these networks cannot be reflected by studies relying on *in vitro*, isolated systems and pathways (Hopkins, 2008). It can be hypothesised that the *in vitro* / *in vivo* gap is substantial and that it will not be possible to get past such a conundrum with simplified approaches (e.g. expanding the number of HTS assays used) without major paradigm changes.

From a practical point of view this means that we are still quite far from being able to replace the classical animal toxicity assays with stand-alone *in vitro* alternatives (first of the "Three Rs"), whereas progress seems possible in terms of the other Rs (reduction, refinement).

Regarding Toxcast in particular, a very promising avenue is provided by the fact that the chemicals are characterised - systematically for the first time in this type of exercise - by chemical structure indexing and chemical descriptors. Thus, for example, if groups of chemicals with similar patterns of HTS responses are highlighted, it is possible to investigate to what extent the patterns are typical of well defined chemical classes or - on the contrary - encompass wider chemical groupings. Putting together (correlating) the three edges of Toxcast information (*in vivo* toxicity, *in vitro* HTS patterns, chemical structure) for the identified local classes may provide the ground to establish predictive models based on the integration of *in vitro* and structure-activity relationship approaches (Dix et al., 2007).

These models – obviously valid only within well defined groups of chemicals - will contribute to the reduction of animal testing. The diversity of chemicals in Toxcast (present and future) and the richness of *in vitro* and *in vivo* endpoint data are promising enough to warrant interesting findings.

However, a number of warnings are necessary. As already emphasised, a first warning is that large numbers of measures easily lend themselves to the generation of patterns and correlations owing to mere chance (Rathman & Yang, 2009) (Topliss & Edwards, 1979). In this context it is crucial that e.g. HTS patterns found for a sample of chemicals are carefully scrutinised and validated for their toxicological relevance. This will require a comparison with both human (if existing) and animal data. The identification of HTS patterns per se without a toxicological counterpart is not particularly informative.

The need for a careful scrutiny of the biological relevance of HTS patterns, and in general of patterns of perturbations identified *in vitro*, is emphasised by experiences already made in other scientific fields.

One field is that of pharmaceutical industry, where the systematic use of HTS can already be considered "old", and enough experience has accumulated. As eloquently expressed by (MacDonald & Robertson, 2009): "…What we have experienced is that the use of *in vitro* / molecular approaches can be very powerful in understanding the mode of action of a pharmaceutical chemical once the actual tissue response *in vivo* is known … (whereas) using the same data prospectively however is much more difficult".

Another lesson to be learned comes from the field of mutagenicity Short-Term Tests (STT) for the screening of carcinogenicity. Following the seminal research by the Millers on electrophilic carcinogens (as such, or after metabolic transformation) (Miller & Miller, 1981), Bruce Ames engineered a series of strains of *Salmonella typhimurium* that mutated in response to the classes of carcinogens identified by the Millers. It appeared (and it is still valid evidence) that chemicals positive in the Ames test have a high probability of being (genotoxic) carcinogens (Maron & Ames, 1983) (Zeiger, 1987). Since the Ames test detects only gene mutations (base substitutions and deletions/additions), scientists hypothesised that other genetic events could be at the basis of chemical carcinogenicity, i.e. structural chromosome aberrations (breaks and rearrangements) and numerical chromosome aberrations (loss or gain of chromosomes, defined as aneuploidy). Thus, in order to increase the probability of detecting chemical carcinogens, additional *in vitro* STTs were generated, incorporating also a wider range of cell types (with particular emphasis on cultured mammalian cells).

This idea has been substantially accepted also in regulatory settings: normally three (or in some cases two) *in vitro* tests are required by regulatory authorities, namely a test for induction of gene mutations in bacteria, a test for induction of gene mutations in mammalian cells and a test for induction of chromosomal aberrations in mammalian cells. A corollary is that positive genotoxicity *in vitro* should be confirmed, in a second

phase, in an *in vivo* genotoxicity assay in order to avoid false positive results (see the Technical Guidance Documents of the European Chemicals Agency (EChA): http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08).

Thus, the screening for carcinogens by mutagenicity STTs implements a paradigm similar to that of Toxcast: perturbations of hypothesised toxicity pathways are detected by a panel of *in vitro* tests. However, whereas the mechanistic relation between the Ames test and the carcinogenicity bioassay is supported also by statistics to a remarkable extent, such a statistically significant relationship between the rodent bioassay results and the "additional" genotoxicity assays is not apparent. For example, an NTP exercise with 114 chemicals demonstrated that only the Ames test had a strong statistically significant association with the rodent bioassay. Chromosomal aberrations in CHO cells had a weaker correlation (not complementary to *Salmonella* in a battery approach), whereas mutation in mouse lymphoma cells and sister chromatid exchanges in CHO cells had no correlation at all (Tennant et al., 1987; Zeiger et al., 1990). Whereas the Ames test has a strong positive predictivity for the rodent bioassay (chemicals positive in *Salmonella* have a high probability of also being carcinogenic), the three other *in vitro* genotoxicity assays responded positively to a large number of chemicals that are non-carcinogenic. Other analyses, on these and other data, have led to similar conclusions (Benigni & Giuliani, 1994; Benigni et al., 1995).

This discrepancy between the hypothesised relationship of the additional genotoxicity tests with the carcinogenicity endpoint, and their actual response to carcinogens and non-carcinogens is still to be solved and creates problems in the interpretation of contradictory genotoxicity results (Elespuru et al., 2009; Kirkland et al., 2005).

To avoid such problems in future toxicology it will be important to reflect on this and other lessons and to validate rigorously the *in vitro* assays against animal and human toxicological data.

References

Benigni, R., Andreoli, C. and Giuliani, A. (1995). Relationships among in vitro mutagenicity assays: quantitative versus qualitative test results. *Environ. Mol. Mutagen. 26*, 155-162.

Benigni, R. and Giuliani, A. (1994). Quantitative modeling and biology: the multivariate approach. *Am J Physiol. 266*, R1697-R1704.

Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci. 95*, 5-12.

EFSA (2009). Existing approaches incorporating replacement, reduction and refinement of animal testing: applicability in food and feed risk assessment. *EFSA J. 1052*, 1-57.

Elespuru, R. K., Agarwal, R., Atrakchi, A. H. et al. (2009). Current and future application of genetic toxicity assays: the role and value of in vitro mammalian assays. *Toxicol. Sci. 109*, 172-179.

Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nature Chem Biol. 4*, 682-690.

Kirkland, D., Aardema, M. J., Henderson, L. and Muller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutag. Res. 584*, 1-256.

MacDonald, J. S. and Robertson, R. T. (2009). Toxicity Testing in the 21st Century: a View from the pharmaceutical industry. *Toxicol. Sci. 110*, 40-46.

Maron, D. M. and Ames, B. N. (1983). Revised methods for the Salmonella mutagenicity test. *Mutat. Res. 113*, 173-215.

Miller, E. C. and Miller, J. A. (1981). Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer 47*, 2327-2345.

Pedersen, F., de Brujin, J., Munn, S. J. and Van Leeuwen, K. (2003). Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. *JRC report EUR 20863 EN*. Ispra.

Rathman, J. F. and Yang, C. (2009). Navigating through the domains of biology and chemistry. ToxCast Data Analysis Summit. Slide Presentation. http://www.epa.gov/NCCT/toxcast/files/summit/13%20Yang%20ToxCast%20TDAS.pdf

Schmidt, C. (2009). Testing for carcinogens: shift from animals to automation gathers steam--slowly. *J. Natl. Cancer Inst. 101*, 910-912.

Tennant, R. W., Margolin, B. H., Shelby, M. D. et al. (1987). Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. *Science 236*, 933-941.

Topliss, J. G. and Edwards, R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *J. Med . Chem. 22*, 1238-1244.

Worth, A. P., Bassan, A., de Brujin, J. et al. (2007). The Role of the European Chemicals Bureau in Promoting the Regulatory Use of (Q)SAR Methods. *SAR QSAR Environ. Res. 18*, 111-125.

Young, D. (2009). Drug innovation on the decline. *Am. J. Health Syst. Pharm. 64*, 228-231.

Zeiger, E. (1987). Carcinogenicity of mutagens: predictive capability of the Salmonella mutagenesis assay for rodent carcinogenicity. *Cancer Res. 47*, 1287-1296.

Zeiger, E., Haseman, J. K., Shelby, M. D. et al. (1990). Evaluation of four in vitro genetic toxicity tests for predicting rodent carcinogenicity: confirmation of earlier results with 41 additional chemicals. *Environ. Mol. Mutagen. 16*, 1-14.

Correspondence to

Romualdo Benigni
Dirigente di ricerca
IstitutoSuperiore di Sanità
Viale Regina Elena 299
00161 Roma
Italy
e-mail: romualdo.benigni@iss.it

EB10: Status report on Sens-it-iv

T Cell Priming and Amplification: Exploiting Key Events in Contact Sensitization

Stefan F. Martin[1], Andrea Cavani[2], Enrico Maggi[3], Hermann-Josef Thierse[4], Anne Richter[5] and Federica Sallusto[6]

[1]Allergy Research Group, Department of Dermatology, University Medical Center Freiburg, Germany; [2]IDI-IRCCS, Rome, Italy; [3]University of Florence, Italy; [4]University Medical Center Mannheim, Germany; [5]Miltenyi Biotec GmbH, Bergisch-Gladbach, Germany; [6]Institute for Research in Biomedicine, Bellinzona, Switzerland

Summary

More than 4,000 chemicals are known to cause allergic contact dermatitis, a T cell mediated inflammatory skin disease. One of the hallmarks of chemical contact allergens is their reactivity with proteins. This reactivity causes the activation of skin inflammation that is essential for the disease to occur and a prerequisite for the activation and recruitment to the skin of chemical reactive T cells. Further, the protein reactivity results in the generation of antigenic determinants for recognition by these T cells. This immunogenicity of contact allergens can be exploited for the development of T cell based assays. Within the EU funded project Sens-it-iv we are developing an in vitro T cell priming assay as well as a T cell amplification assay. In the in vitro T cell priming assay naïve human T cells are activated with myeloid dendritic cells either directly modified with putative contact allergens or fed with contact allergen modified human serum albumin. Contact allergens can be identified by their ability to induce specific T cell proliferation or cytokine production. In the T cell amplification assay, rare allergen-specific precursor T cells are expanded by polyclonal stimulation in replicate cultures. T cell lines can be developed whereby each naïve T cell is expanded several hundred-fold into clonally activated T cell blasts. The second step of antigen-specific stimulation allows the identification of cultures containing specific T cells as well as the isolation of antigen-specific T cell clones. This assay has been used to measure frequency and assess avidity of naïve T cells specific for different types of antigens including protein allergens (e.g. Der p I from house dust mite) and is now being tested with chemical-modified self-proteins.

These assays may be a valuable tool to avoid some of the shortcomings of the innate immune cell-based assays and could be useful as a second or third line test for substances that cannot be safely classified.

Keywords: alternative methods, contact sensitizer, respiratory sensitizer, allergen, T cell, dendritic cell

1 Identification of contact allergens using *in vitro* assays

Consumer products contain many chemicals that can induce contact dermatitis or respiratory allergies. In order to ensure product safety several methods are used that should allow identifying the skin and respiratory sensitizing potential of chemicals and proteins used for a consumer product or encountered in the work place. So far animals have been used for such tests. The local lymph node assay (LLNA) in mice for example has been established as the gold standard for contact allergen identification (Basketter, 2008). Due to EU legislation and worldwide efforts to reduce animal use for such tests and for ethical reasons it is necessary to develop suitable *in vitro* alternatives to reduce and eventually replace animal testing (dos Santos et al., 2009). The strategies that are pursued worldwide aim at the reliable *in vitro* recapitulation of crucial steps in the sensitization to contact and respiratory allergens (and drugs) in easy-to-perform, high throughput assays. Moreover, one of the most important features of these allergens is their ability to induce a T cell response

following successful sensitization (Martin, 2004). Therefore, it is also useful to establish T cell based in vitro assays that allow a conclusion on whether a successful sensitization is likely to result in a productive allergic response.

Sensitization to contact and respiratory allergens involves the activation of tissue and innate immune cells and the triggering of innate immune and stress responses (Edele et al., 2007; Freudenberg et al., 2009; Marin 2010). This leads to an inflammatory response in the affected tissue, the skin or the lung, and activates dendritic cells to migrate into the draining lymph nodes, present the allergen to naïve T cells and activate them. The activated effector and memory T cells then find their way into the tissue and cause the typical symptoms of the allergic response to skin and respiratory sensitizers, such as allergic contact dermatitis (ACD) and chemical-induced allergic respiratory disease (CARD) as well as asthmatic responses to protein allergens. *In vitro* assays that cover sensitization aim at using single cell systems such as keratinocytes or dendritic cells, ideally cell lines. Typical readout parameters in such assays are the induction of surface markers or of cytokine and chemokine production of the cells in response to the test compounds.

2 T cell-based assays for the identification of chemical- and protein-allergens

As mentioned above one of the crucial steps in the allergic response to contact and protein allergens is the induction of a T cell response. The development of *in vitro* assays based on the T cell response has been difficult due to the necessity to use primary cells from the peripheral blood of donors and autologous antigen presenting cells (APC). Moreover, sufficient knowledge about the chemistry of the test compound is crucial with regards to choosing a suitable inert solvent and the right conditions that allow the chemical to bind to proteins or peptides in order to produce MHC-presented T cell epitopes. In the case of protein allergens this is not necessary, since suitable APC will most likely take up and process the protein.

Within and outside of the EU project Sens-it-iv (www.sens-it-iv.eu), such efforts are made. An improvement in the methods to purify T cells and APC and to generate dendritic cells *in vitro* and progress in understanding the immunology of these allergic responses, especially the important role of regulatory T cells (Cavani, 2008), has led to refined T cell assay protocols.

2.1 The *in vitro* T cell priming assay

In our Sens-it-iv consortium we use the *in vitro T cell priming assay* for chemicals. In this assay purified naïve T cells depleted of putative CD25+ Treg cells are stimulated with autologous CD14+ monocyte-derived DC for a period of 7-10 days. The test chemicals are either added directly or the DC are pretreated with the chemical and then washed to remove unbound material that may become toxic eventually. Alternatively, proteins such as human serum albumin (HSA) are modified with the test chemical and added to the assay (Dietz, L. et al., 2010; Martin, S. F. et al., 2010). The readout is done by measuring chemical-specific proliferation and by detection of cytokine production, mostly IFN-γ by ELISA, detection of secreted cytokines immobilised by antibody capture on the producing cell or by intracellular cytokine staining for flow cytometry following restimulation. The current status of this assay demonstrates a reliable detection of strong contact allergens such as 2,4-dinitrochlorobenzene (DNCB) and some other, weaker allergens such as eugenol, cinnamic aldehyde but also beta-lactams. Problems encountered are donor to donor variability and sometimes a negative result upon repeated testing of the same donor. Moreover, for some chemicals it is unclear whether they can react with proteins or cells under the assay conditions. In order to reduce these problems one always has to test more than one donor in order to exclude individual differences in T cell reactivity or T cell receptor repertoires. Also, the quality of the APC has to be assured by suitable methods (cell viability, phenotypic quality control by flow cytometry of surface markers).

The T cell priming assay thus seems to work especially for strong contact allergens. Given its limitations, e.g. the need for primary cells and more than one donor, it may not be suitable as a first line, high throughput assay. However if we can use test compounds for which we can control their protein reactivity, e.g. by a mass spectrometry based assay using coupling to model peptides and detection of the resulting mass shift (Dietz, L. et al., 2010; Martin, S. F. et al., 2010), this assay may be very useful as a second line assay to identify strong contact allergens and to sort out doubtful candidates that give no clear score in a panel of single cell based assays.

2.2 The T cell library assay

A novel method, the *T cell library assay* was designed to detect very rare antigen-specific naïve T cells from the blood (Geiger et al., 2009). In a first step "T cell amplification" of very rare antigen-specific T cells is achieved by polyclonal stimulation of small numbers of naïve T cells in replicate cultures leading to the development of a "library" of T cell lines. This allows expansion of each naïve T cell several hundred-fold into clonal T cell blasts. The second step comprises antigen-specific stimulation for the identification of cultures containing specific T cells. This enables us to isolate antigen-specific T cell clones. The advantages of this method are the increased number of responding effector T cells as well as their lower activation threshold in the second step. The assay seems especially suited to measure T cell responses to protein allergens. It can be used to measure frequencies and assess avidities of naïve T cells specific for different types of protein antigens including protein allergens (e.g. Der p I from house dust mite) and it is now being tested with chemical-modified self-proteins.

3 Conclusion

The rationale for T cell based assays, despite their somewhat complicated nature when compared with single cell assays that detect aspects of the sensitization process, is the crucial importance of the T cell response in the allergic response to contact and respiratory allergens. As one step within a test panel lined up in a tiered strategy that scores the results of

each assay to give a final score based on all steps (Jowsey et al., 2006), T cell based assays may be very useful. The technical problems will most likely be overcome by incorporating knowledge of the chemistry and the immunology of the allergens.

References

Basketter, D. A. (2008). Skin sensitization: strategies for the assessment and management of risk. *Br. J. Dermatol. 159*, 267-273.

Cavani, A. (2008). T regulatory cells in contact hypersensitivity. *Curr. Opin. Allergy Clin. Immunol. 8,* 294-298.

Dietz, L., Esser. P. R., Schmucker, S. et al. (2010). Tracking human contact allergens: From mass spectrometric validated chemical-peptide interactions to chemical-specifiv naïve human T cell priming. *Toxicol. Sci.*, accepted for publication.

Dos Santos, G. G., Reinders, J., Ouwehand, K. et al. (2009). Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound. *Toxicol. Appl. Pharmacol. 236*, 372-382.

Edele, F., Esser, P. R., Lass, C. et al. (2007). Innate and adaptive immune responses in allergic contact dermatitis and autoimmune skin diseases. *Inflamm. Allergy Drug Targets 6*, 236-244.

Freudenberg, M. A., Esser, P. R., Jakob, T. et al. (2009). Innate and adaptive immune responses in contact dermatitis: analogy with infections. *G. Ital. Dermatol. Venereol. 144*, 173-185.

Geiger, R., Duhen, D., Lanzavecchia, A. and Sallusto, F. (2009). Human naive and memory CD4+ T cell repertoires specific for naturally processed antigens analyzed using libraries of amplified T cells. *J. Exp. Med. 206*, 1525-1534.

Jowsey, I. R., Basketter, D. A., Westmoreland, C. and Kimber, I. (2006). *J. Appl. Toxicol. 26*, 341-350.

Martin, S. F. (2004). T lymphocyte-mediated immune responses to chemical haptens and metal ions: implications for allergic and autoimmune disease. *Int. Arch. Allergy Immunol. 134*, 186-198.

Martin, S. F., Esser. P. R., Schmucker, S. et al. (2010). T cell recognition of chemicals, protein allergens and drug: toward the development of in vitro assays. *Cell. Mol. Life Sci.*, accepted for publication.

Martin, S. F. (2010). The role of the innate immune system in allergie contact dermatitis. *Allergologie 33*, 66-70

Acknowledgments

The work summarised here was supported by a grant from the European Commission as part of the project "Novel Testing Strategies for In Vitro Assessment of Allergens (Sens-it-iv)", LSHB-CT-2005-018681.

Correspondence to

Stefan F. Martin, PhD
Allergy Research Group
Department of Dermatology
University Medical Center Freiburg
Hauptstrasse 7
79104 Freiburg
Germany
e-mail: stefan.martin@uniklinik-freiburg.de

Satellite Meeting

Symposium

From Tissue Engineering to Alternatives: Research, Discovery and Development

A satellite symposium to the 7th World Congress on Alternatives and Animal Use in the Life Sciences, organised under the auspices of CELLTOX (Associazione Italiana di Tossicologia in vitro) and ESTIV (European Society of Toxicology in Vitro)

Francesca Caloni[1], Isabella De Angelis[2] and Jan van der Valk[3]

[1]Università degli Studi di Milano, Department of Veterinary Sciences and Technologies for Food Safety Faculty of Veterinary Medicine, Milan, Italy; [2]Department of Environmental and Primary Prevention, Istituto Superiore di Sanità, Rome, Italy; [3]Netherlands Centre Alternatives to Animal Use, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

Introduction

Research animals still serve as a model for man in biomedical research. There is a strong drive to limit animal experimentation for ethical, scientific, practical and economic reasons. *In vitro* models are widely developed and used to (partly) replace animal studies. Most *in vitro* models are based on single cell type 2D cultures, which lack the 3D structures of tissues under investigation and the interactions of the different cell types normally present *in vivo*. Tissue engineering promises to overcome some of these limitations.

Tissue engineering is an emerging multidisciplinary technique that involves the use of living, preferably human, cells to create tissues and organs *de novo*, either for therapeutic or diagnostic applications. Tissue engineering research includes different areas of application: biomaterials (e.g. scaffold), biomolecules (e.g. bone morphogenic proteins), engineering design aspects (e.g. 3D structure) and biomechanical aspects (e.g. safety and efficacy of engineered tissues).

3D culture systems, moving from "flat biology" of cell monolayers to a more realistic culture environment that maintains some aspects of the original tissue/organ characteristics and better preserves the differentiated cells' phenotype, offer new opportunities for biomedical studies, becoming a bridge between *in vitro* and *in vivo* models.

3D reconstructed human skin equivalent (HSE) can be used for drug testing and for fundamental research purposes to better understand skin disease development in order to find appropriate therapies. Full thickness skin equivalents are therefore attractive for the study of cell-cell, cell-matrix and dermal-epidermal interactions and to mimic diseased skin disorders *in vitro* in order to test therapeutics. Studies with HSEs can therefore contribute to our knowledge of basic biochemical mechanisms underlying irritant reactions and can be used to understand the structural features of molecules that may be responsible for eliciting an irritant reaction. Reconstructed epidermal models used for screening of potential skin irritants have recently been accepted by the European Union (EU) as alternative methods (EPISKIN™ , SkinEthic RTE, EpiDerm™), while three-dimensional reconstructed tissue models of human cornea (EpiOcular™ , SkinEthic) are regarded with great interest for the study of eye irritation.

Stem cell research is a promising approach for tissue therapy of neurodegenerative diseases, brain injury and cardiovascular disorders. Adult stem and progenitor cells, such as hNT neurons (derived from an embryonal human teratocarcinoma cell line), bone marrow and human umbilical cord blood (hUCB) derived cells are alternative cell sources compared to embryonic stem cells, the use of which is restricted in certain countries for ethical reasons. hUCB cells represent a non-controversial source of stem and progenitor cells appearing to have a multipotent capacity to differentiate into endothelial cells, neurons, glia and other cell types. They are easily obtained, provide repair through trophic and immunological mechanisms as shown in *in vitro* and *in vivo* studies and are associated with a low incidence of graft-versus-host (GVHD) diseases.

The development, use and opportunities of tissue engineering in different areas of application were discussed during a satellite symposium at the 7th World Congress on Alternatives and Animal Use in the Life Sciences organised under the auspices of CELLTOX and ESTIV on August 30, 2009 in Rome, Italy. The subsequent papers by A. E. Ghalbzouri, P. R. Sandberg and G. Mazzoleni follow this introduction.

Correspondence to:
Dott.ssa Francesca Caloni
Università degli Studi di Milano
Department of Veterinary Sciences and
Technologies for Food Safety
Faculty of Veterinary Medicine,
Milan, Italy
e-mail: francesca.caloni@unimi.it

In Vitro Disease Models: Human Skin Equivalents for Research Purposes

Abdoelwaheb El Ghalbzouri and Suzan Commandeur

Department of Dermatology, Leiden University Medical Center, The Netherlands

Summary

Reconstructed human skin equivalents (HSEs) are representative models of human skin and widely used for safety testing, tissue engineering, clinical applications and research purposes. Since in vitro skin models recapitulate most of the in vivo characteristics of human skin, they contribute to the replacement of animal experimentation. Epidermal HSEs are mostly used to predict the safety of ingredients used in various industries, while full thickness HSEs are more often applied for research and tissue engineering purposes. A highly active research area is emerging from the combination of tissue engineering principles and knowledge from developmental biology to establish complex three-dimensional in vitro disease model systems representing various skin diseases, including recessive epidermolysis bullosa simplex and cutaneous squamous cell carcinoma. These diseased skin models help us to gain more insight into the mechanisms of various pathological conditions and to screen for novel therapeutics.

Keywords: human skin equivalent, skin model, epidermolysis bullosa simplex, squamous cell carcinoma

1 Introduction and questions

Human skin equivalents

Since culture of human epidermal keratinocytes on feeder layers was developed, research to produce human skin equivalents (HSEs) has been motivated by the challenge of treating large burns and chronic wounds and by European regulations which both require proof of the innocuousness and the effectiveness of cosmetic products, and which forbid animal testing (Rheinwald and Green, 1975). Engineered skin that is constructed from living human skin cells *in vitro* has become an increasingly functional alternative to animal experimentation. Since the start of its development in the early 1980's, applications of HSEs have expanded throughout the fields of safety testing, tissue engineering and basic research, thereby greatly adding to the reduction and replacement of animal use in research (Ponec et al., 1988). As HSEs are based on human components, they represent a unique tool specific for humans, in contrast to animal models whose extrapolation to human situations in terms of functional biology may be highly questioned. Epidermal HSEs consist of a fully functional and differentiated epidermis with basal, spinous, granular and cornified layers of keratinocytes (Bell et al., 1981; El Ghalbzouri et al., 2002; Parenteau et al., 1991; Ponec et al., 1997; Prunieras et al., 1983). These models first greatly increased insight into normal human skin biology by allowing functional cell-cell interaction studies. Currently, epidermal HSEs are commercially available and successfully applied in and validated for irritation and corrosion testing, thereby replacing the use of animals for these purposes. The irritant potential can be evaluated using various endpoints, including the induction of tissue damage, the release of the pro-inflammatory mediator interleukin-1α and changes in protein and mRNA expression profiles (Boxman et al., 2002). Studies with epidermal HSEs therefore contribute to our knowledge on the basic biochemical mechanisms underlying irritant reactions and can be used to understand the structural features of molecules potentially responsible for eliciting irritant reactions. The most recent challenges lie in the application of epidermal HSEs in food, chemical, cosmetic and pharmaceutical industries and in validation of skin sensitisation *in vitro*. In addition, epidermal HSEs are successfully applied in skin grafting and basic research on human skin wounding and scar formation (El Ghalbzouri et al., 2004).

Full thickness HSEs, containing both a reconstructed human fibroblast-populated dermis and an epidermal layer, allow for numerous applications requiring fully functional dermal-epidermal and cell-matrix interactions. The presence of dermal human fibroblasts in full thickness HSEs facilitates the formation of the basement membrane, which is required for the maintenance of skin integrity (El Ghalbzouri et al., 2002; Grassel et al., 1999). Although the use of full thickness HSEs in validated safety testing is still limited due to their highly complex structure, these models offer immense possibilities in the fields of tissue engineering and basic research. In particular, full-thickness HSEs form a functional tool for studying complex skin abnormalities involving the interaction between dermis and epidermis.

In vitro diseased skin models

Tissue engineering concerns both the generation of healthy tissues and diseased tissues in order to better understand various pathological conditions. The establishment of human cell-based *in vitro* engineered disease model systems could repre-

sent a paradigm shift from inadequate conventional monolayer cell cultures, or moderately successful animal models, towards more physiologically tissue-relevant, patient-specific approaches. This concept of *in vitro* diseased skin models is illustrated by the models for recessive epidermolysis bullosa simplex (REBS) and cutaneous squamous cell carcinoma (SCC) that were developed in our laboratory. REBS is characterised by severe intra-epidermal blister formation, resulting from fragility of the basal keratinocytes that lack keratin tonofilaments due to a homozygotic null mutation in the keratin 14 gene (Bonifas et al., 1991; Chan et al., 1994). *In vivo* EBS models, including murine and canine models, are limited and do not represent the human microenvironment (Arin and Roop, 2001; Palazzi et al., 2000). Earlier *in vitro* studies have indicated an essential role for fibroblasts in the REBS phenotype (Ehrlich et al., 1983; Oakley and Priestley, 1985). However, the limited availability of human REBS skin samples hampered further *in vitro* research on dermal-epidermal interactions in REBS. To overcome this, an explant approach was used in which REBS skin biopsies were placed on a dermal equivalent in which REBS-associated fibroblasts were seeded (El Ghalbzouri et al., 2003). A similar approach was used to construct an *in vitro* model for human SCC, which is a malignant tumour of epidermal keratinocytes characterised by invasive growth into the dermis. After basal cell carcinoma, SCC is the most common malignancy in white populations with epidemic incidence rates (de Vries et al., 2005; Staples et al., 2006). Traditional *in vivo* SCC models rely on the use of chemical, genetic or mechanical induction or propagation of carcinogenesis in mice (Atillasoy et al., 1997; Brown and Balmain, 1995). Current *in vitro* approaches are limited to the use of cell lines, which often lack a true representation of primary skin cancer (Popp et al., 2000; Proby et al., 2000). The development of a representative *in vitro* skin carcinoma model based on primary SCC biopsies allows for a better understanding of fundamental carcinogenesis mechanisms and may serve as a validated pre-screening platform for candidate drugs, thereby eradicating the need for animals for these purposes.

2 Materials and methods

Human material

Healthy human skin samples originated from mamma reduction surplus skin, which was obtained with informed consent of the donors according to the Dutch law on medical treatment agreement. Primary affected and apparently healthy skin biopsies from several REBS (Fig. 1A) and immunocompromised

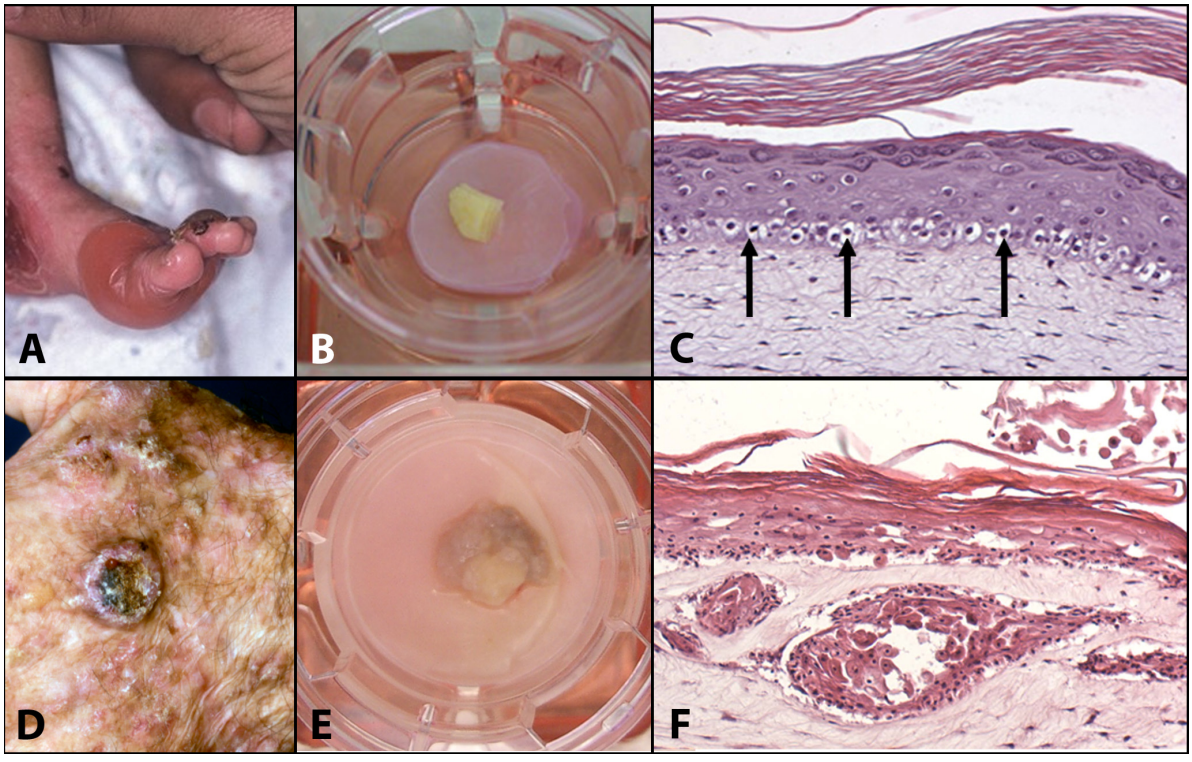


Fig. 1: *In vitro* disease models represent their primary counterparts.
A skin biopsy of recessive epidermolysis bullosa simplex (REBS, A) was placed on a dermal equivalent (B), ultimately leading to an *in vitro* REBS model with extensive vacuolisation, thereby representing a characteristic REBS morphology (C). In analogy, a skin biopsy of cutaneous squamous cell carcinoma (SCC, D) was placed on a dermal equivalent (E), ultimately leading to an invasive *in vitro* SCC model reflecting characteristics of its primary counterpart (F).

SCC patients (Fig. 1D) were freshly collected with written informed consent according to the Declaration of Helsinki after approval of the medical ethical committee of our institution. Of these biopsies, one part was processed for diagnosis while the remainder was used for explant culture.

Culture of diseased in vitro human skin explant models
Both the REBS and SCC models are based on a standardised explant method described earlier (Boxman et al., 2000). In short, dermal equivalents were generated by seeding either healthy or disease-associated fibroblasts in a surplus rat-tail derived collagen suspension. After submerged incubation for either one day or one week, fresh biopsies of REBS or SCC were placed onto the fibroblast-seeded dermal collagen matrices. The diseased explants were then cultured serum-free and air-exposed for three to four weeks, allowing affected keratinocytes to expand over the associated dermal equivalents. Culture medium was used as described earlier and refreshed twice a week (El Ghalbzouri et al., 2003; Commandeur et al., 2009).

Histology and immunohistochemistry
Upon fixation, diseased explant models were subject to histological and immunohistochemical analyses. For histology, 5 μm paraffin sections were stained with haematoxylin and eosin (HE). Immunohistochemical analyses were performed on either 5 μm paraffin- or cryosections. Antibodies used are listed in literature and included markers specific for keratinocyte activation, proliferation, differentiation and basement membrane composition (Commandeur et al., 2009; El Ghalbzouri et al., 2003). All sections were counterstained with haematoxylin.

3 Results

In vitro human recessive epidermolysis bullosa simplex model
After four weeks of culture, *in vitro* human REBS models were terminated (Fig. 1B). Histological analysis of the *in vitro* human REBS model revealed a fully differentiated epidermis with extensive basal cell vacuolisation (Fig. 1C). In addition, K14 ablation could also be reproduced in human REBS cultures established with REBS biopsies (El Ghalbzouri et al., 2003).

In vitro human cutaneous squamous cell carcinoma model
After three weeks of culture, *in vitro* human SCC models were terminated (Fig. 1E). Histological analysis of the *in vitro* human SCC model showed lateral expansion of the biopsy-associated keratinocytes over the dermal equivalent (Fig. 1F). In addition, invasion of a subpopulation of expanded keratinocytes was observed (Fig. 1F). In contrast to healthy human keratinocytes in human skin models, these keratinocytes were able to form invasive nests in the dermal equivalent, both in the direct vicinity of and more distant from the original carcinoma biopsy. Furthermore, immunohistochemical analyses revealed persistence of putative SCC-specific markers K4 and K13 in addition to increased epidermal activation, disturbed differentiation and an overall normal basement membrane (Commandeur et al., 2009).

4 Discussion

Full thickness HSEs are a functional tool for studying complex skin abnormalities. Their functional dermal-epidermal interactions reflecting normal skin homeostasis can be modulated towards diseased dermal-epidermal interactions and even tumour-stroma interactions. Using primary human biopsies of either REBS or SCC, we were able to create three-dimensional *in vitro* diseased skin models harbouring both disease-specific keratinocytes and fibroblasts in a well-controlled microenvironment. Both models are highly representative for the complex skin diseases they originate from. The possibility to generate HSEs with skin biopsies of patients suffering from complex genetic disorders and cancer offers an attractive approach for *in vitro* studies focusing on the mechanisms of various pathological conditions. However, despite major resemblance of both healthy and diseased *in vitro* human skin models to their *in vivo* counterparts, these models still have limitations, warranting further development of three-dimensional *in vitro* culturing. Given the essential interaction of keratinocytes and fibroblasts in the epidermal and dermal compartments, respectively, modulation of the dermal microenvironment often results in functional improvements. For example, *in vitro* healthy human skin models used to be restrained by limited culture durations of approximately eight weeks. Recent developments led to the replacement of rat tail-derived collagen matrix with human fibroblast-derived dermal matrix, resulting in skin models fully composed of human material that can be successfully cultured for up to 20 weeks (El Ghalbzouri et al., 2008). As the REBS model still lacks extensive cornification, which is another feature associated with the REBS phenotype, it may be further improved by optimising the culture duration and mechanical shear forces. Furthermore, as the *in vitro* SCC model initially presented with limited invasion, replacement of healthy human fibroblasts with carcinoma-associated fibroblasts in the dermal compartment of these models is highly likely to improve their resemblance to native SCCs on the virtue of invasion.

In conclusion, both epidermal and full thickness HSEs serve a great purpose in validated safety testing, tissue engineering and basic research. In particular, full thickness HSEs harbouring fully functional dermal-epidermal interactions can be applied to replace diseased animal models used for studying pathogenesis and for screening of therapeutics for a wide variety of skin diseases including basal cell carcinoma, atopic dermatitis, psoriasis and fibrotic disorders.

References

Arin, M. J. and Roop, D. R. (2001). Disease model: heritable skin blistering. *Trends Mol.Med. 7*, 422-424.

Atillasoy, E. S., Elenitsas, R., Sauter, E. R. et al. (1997). UVB induction of epithelial tumors in human skin using a RAG-1 mouse xenograft model. *J.Invest Dermatol. 109*, 704-709.

Bell, E., Ehrlich, H. P., Buttle, D. J. and Nakatsuji, T. (1981). Living tissue formed in vitro and accepted as skin-equivalent tissue of full thickness. *Science 211*, 1052-1054.

Bonifas, J. M., Rothman, A. L. and Epstein, E. H. (1991). Epidermolysis bullosa simplex: evidence in two families for keratin gene abnormalities. *Science 254*, 1202-1205.

Boxman, I. L., Hensbergen, P. J., Van Der Schors, R. C. et al. (2002). Proteomic analysis of skin irritation reveals the induction of HSP27 by sodium lauryl sulphate in human skin. *Br. J. Dermatol. 146*, 777-785.

Boxman, I. L., Mulder, L. H., Vermeer, B. J. et al. (2000). HPV-DNA is not detectable in outgrowing cells from explant cultures of skin lesions established at the air-liquid-interface. *J. Med. Virol. 61*, 281-288.

Brown,K. and Balmain, A. (1995). Transgenic mice and squamous multistage skin carcinogenesis. *Cancer Metastasis Rev. 14*, 113-124.

Chan, Y., Anton-Lamprecht, I., Yu, Q. C. et al. (1994). A human keratin 14 "knockout": the absence of K14 leads to severe epidermolysis bullosa simplex and a function for an intermediate filament protein. *Genes Dev. 8*, 2574-2587.

Commandeur, S., de Gruijl, F. R., Willemze, R. et al. (2009). An in vitro three-dimensional model of primary human cutaneous squamous cell carcinoma. *Exp.Dermatol. 18*, 849-856.

de Vries, E., van de Poll-Franse, L. V., Louwman, W. J. et al. (2005). Predictions of skin cancer incidence in the Netherlands up to 2015. *Br. J. Dermatol. 152*, 481-488.

Ehrlich, H. P., Buttle, D. J., Trelstad, R. L. and Hayashi, K. (1983). Epidermolysis bullosa dystrophica recessive fibroblasts altered behavior within a collagen matrix. *J. Invest Dermatol. 80*, 56-60.

El Ghalbzouri, A., Commandeur, S., Rietveld, M. H. et al. (2008). Replacement of animal-derived collagen matrix by human fibroblast-derived dermal matrix for human skin equivalent products. *Biomaterials*.

El Ghalbzouri, A., Hensbergen, P., Gibbs, S. et al. (2004). Fibroblasts facilitate re-epithelialization in wounded human skin equivalents. *Lab. Invest 84*, 102-112.

El Ghalbzouri, A., Jonkman, M., Kempenaar, J. and Ponec, M. (2003). Recessive epidermolysis bullosa simplex phenotype reproduced in vitro: ablation of keratin 14 is partially compensated by keratin 17. *Am. J. Pathol. 163*, 1771-1779.

El Ghalbzouri, A., Gibbs, S., Lamme, E. et al. (2002). Effect of fibroblasts on epidermal regeneration. *Br. J. Dermatol. 147*, 230-243.

Grassel, S., Unsold, C., Schacke, H. et al. (1999). Collagen XVI is expressed by human dermal fibroblasts and keratinocytes and is associated with the microfibrillar apparatus in the upper papillary dermis. *Matrix Biol. 18*, 309-317.

Oakley, C. A., Priestley, G. C. (1985). Collagen synthesis and degradation by epidermolysis bullosa fibroblasts. *Acta Derm. Venereol. 65*, 277-281.

Palazzi, X., Marchal, T., Chabanne, L. et al. (2000). Inherited dystrophic epidermolysis bullosa in inbred dogs: A spontaneous animal model for somatic gene therapy. *J. Invest Dermatol. 115*, 135-137.

Parenteau, N. L., Nolte, C. M., Bilbo, P. et al. (1991). Epidermis generated in vitro: practical considerations and applications. *J. Cell Biochem. 45*, 245-251.

Ponec, M., Weerheim, A., Kempenaar, J. et al. (1997). The formation of competent barrier lipids in reconstructed human epidermis requires the presence of vitamin C. *J. Invest Dermatol. 109*, 348-355.

Ponec, M., Weerheim, A., Kempenaar, J. et al. (1988). Lipid composition of cultured human keratinocytes in relation to their differentiation. *J. Lipid Res. 29*, 949-961.

Popp, S., Waltering, S., Holtgreve-Grez, H. et al. (2000). Genetic characterization of a human skin carcinoma progression model: from primary tumor to metastasis. *J. Invest Dermatol. 115*, 1095-1103.

Proby, C. M., Purdie, K. J., Sexton, C. J. et al. (2000). Spontaneous keratinocyte cell lines representing early and advanced stages of malignant transformation of the epidermis. *Exp. Dermatol. 9*, 104-117.

Prunieras, M., Regnier, M. and Woodley, D. (1983). Methods for cultivation of keratinocytes with an air-liquid interface. J. *Invest Dermatol. 81*, 28s-33s.

Rheinwald, J. G. and Green, H. (1975). Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell 6*, 331-343.

Staples, M. P., Elwood, M., Burton, R. C. et al. (2006). Non-melanoma skin cancer in Australia: the 2002 national survey and trends since 1985. *Med.J.Aust. 184*, 6-10.

Acknowledgements

The work on the SCC-model was supported by a grant from the Netherlands Organization for Health Research and Development (ZonMw project number 11.400.0079).

Correspondence to

Abdoelwaheb El Ghalbzouri
Leiden University Medical Center
Department of Dermatology
Postal Zone S2-P, P.O. Box 9600
2300 RC Leiden, The Netherlands
e-mail: a.ghalbzouri@lumc.nl

3D Culture in Microgravity: A Realistic Alternative to Experimental Animal Use

Giovanna Mazzoleni[1] and Nathalie Steimberg[2]

[1]General Pathology & Immunology Unit, Department of Biological Sciences and Biotechnologies, [2]School of Medicine, University of Brescia, Brescia, Italy

Summary

The acceptance of the new paradigm of three-dimensional (3D) cell culture will enhance in the next years the possibility to develop reliable and physiologically relevant in vitro alternatives to the use of in vivo animal-based models. The recent progress in tissue engineering approaches and the rapid advances in emerging technologies have greatly contributed to the design and set up of a large number of relevant bioreactor-based 3D tissue analogues with various characteristic properties.

This article addresses the potential of the microgravity-based RCCS™ technology to provide the best culture conditions for the generation and long-term maintenance of functional 3D cell constructs and tissue explants for future applications in basic research, pharmaco-toxicology and clinical fields.

Keywords: alternative methods, microenvironment, microgravity, 3D models, Rotary Cell Culture System (RCCS™), bioreactors

During the last two decades, the development and implementation of reliable *in vitro* methods as alternatives to traditional animal-based studies increasingly became a priority in basic research and in the more complex fields of safety and risk assessment (Lilienblum et al., 2008). This is clearly demonstrated, for example, by the declared policy of the European Union (e.g. Directive 86/609/EEC on the protection of experimental animals; 7th Amending Directive 2003/15/EC to Council Directive 76/768/EEC on Cosmetics; REACH Regulation on Chemicals and their safe use - EC 1907/2006, and Council Regulation n.440/2008 on dangerous substances). The EU also devoted significant economic efforts in order to support research projects on the 3Rs principle of Russell and Burch (Russell and Burch, 1959) within specific RTD Framework Programmes (i.e. FP6 and FP7, respectively), whose encouraging results have been recently summarised (EC Project Report, 2009). Similarly, this objective has been partaken by the U.S. [see, for example, the report issued by the U.S. National Research Council in 2007 (NRC, 2007), the Multi-Agency Strategic Plan for Federal Government investments in tissue science and engineering of 2007 (MATES IWG, 2007), and, lastly, the U.S. Environmental Protection Agency Plan for Evaluating the Toxicity of Chemicals of 2009 (U.S. EPA, 2009)]. And also other industrialised regions have embraced the 3Rs principle, such as Japan and Canada that, together with EU and US representatives, recently signed an international agreement for advancing alternatives to animal testing (ICATM, 2009). Apart from obvious ethical and economic considerations, animal models are progressively showing their limitations in mirroring human pathophysiology and responses to exogenous factors, which raise also questions about their relevance and reliability for human-specific safety and risk assessment procedures (Cohen, 2004; Bremer et al., 2007).

Due to the impressive scientific and technological progress of these last ten years, *in vitro* cell- and tissue-based methods can now provide substitutes for some traditional *in vivo* animal-based studies. To be reliable and physiologically relevant, experimental *in vitro* models should correctly reproduce the phenotype and function of the specific target system (Bhadriraju and Chen, 2002). Over two decades of research have demonstrated that, with respect to traditional two-dimensional (2D) cell culture systems (that still represent the most popular models for *in vitro* studies) three-dimensional (3D) models have the potential to improve the physiological relevance of cell-based assays and to advance the quantitative modelling of biological systems from cells to living organisms (Pampaloni et al., 2007). *In vivo*, cells develop and grow within complex and dynamic tissue-specific 3D micro-environments that, beside dimensionality and temporally-regulated physical and biochemical properties, also comprise multiple cell populations that, reciprocally, influence each other and their surroundings, thus leading to unique cell behaviour and responses. Based on these considerations, the importance of the specific 3D micro-environment in designing physiologically relevant *in vitro* models of living tissues has led to the development of progressively more complex and differentiated 3D systems that, due to the rapid advances in culturing techniques emerging from the multidisciplinary field of tissue engineering, have considerably increased in number during the last decade (see Mazzoleni et al., 2009 and references therein). For High-Throughput (HT) assays development, for example, new emerging technologies (e.g. improved micro- and nano-fabrication techniques, micro-fluidics and micro-electronics) coupled with systems biology and innovative analytical techniques, generated highly sophisticated platforms and integrated systems of tissue analogues ("*cells*- and *tissue-on-a-chip*" approaches)

(El-Ali et al., 2006; Khademhosseini et al., 2006; Meyvantsson and Beebe, 2008). Even if promising, these micro-culture methods should be further improved and may, at the moment, be useful only for a limited number of short-term experimental applications.

Since it is well known that the metabolic requirements of 3D cell constructs are substantially higher than those of flat, 2D cell monolayers grown in static environments in liquid media, specific devices, known as dynamic bioreactors, were primarily developed to modulate mass transfer. This is a crucial element for guaranteeing long-term gas and nutrient supply and waste elimination, which are essential factors for maintaining cell viability within large 3D cell or tissue masses. Differences between tissues also require specific culture characteristics to be taken into account. A specific bioreactor configuration (design and operational conditions) should then be based on the precise evaluation of all these parameters (Martin and Vermette, 2005). Maintaining tissue explants or producing complex mammalian 3D tissue analogues *in vitro* thus requires appropriate bioreactors that simulate tissue-specific physiological micro-environments. Taking advantage of the new technologies and of computational fluid dynamics, a wide array of dynamic bioreactors has been devised (for detailed reviews see Martin and Vermette, 2005; Meuwly et al., 2007; Catapano and Gerlach, 2007); however, none of these devices (from the simplest stirred- or suspension-culture systems, to the more complex membrane-based reactors, or to their most sophisticated versions, that include load- and perfusion-controlled systems) is, at present, able to provide optimal conditions for the long-term culture of large tissue-like masses. The current generation of bioreactors was developed for yielding large masses of cells or cell products for industrial or clinical applications, and not for supporting the survival or the self-assembly of multiple cell types into complex 3D structures (Hutmacher and Singh, 2008). Even if hydrodynamic forces effectively increase mass transfer, in dynamic bioreactors for 3D culturing this effect should be achieved by considering (and balancing) the detrimental effect of turbulence and shear stress on cell survival and function. Low-shear environment and optimal mass transfer have been attained only with the introduction of the Rotary Cell Culture System (RCCS^{TM}, Synthecon, Inc.) bioreactors. Fruit of N.A.S.A.'s Johnson Space Center technological research, and successfully used in ground- as well as in space-based studies on a wide variety of cell types and tissues (a vast literature is available at http://www.synthecon.com), RCCS^{TM} bioreactors present several advantages with respect to other available 3D culture systems. Horizontally rotating, transparent clinostats, RCCS^{TM} devices efficiently create a unique, highly controlled micro environment that, by reproducing some aspects of microgravity (simulated microgravity) (Klaus, 2001; Ayyaswamy and Mukundakrishnan, 2007), guarantee the most favourable conditions for cell and tissue culturing (Schwarz et al., 1992), and provide potentially powerful tools to reproduce specific 3D tissue morphogenesis (Mazzoleni et al., 2009). Complex tissue-like 3D constructs, different cell types from various origins and various intact tissue explants have been demonstrated, also by our group, to be kept efficiently in culture by these bioreactors, even for long periods of time (Unsworth and Lelkes, 1998; Hammond and Hammond, 2001; Vunjak-

Tab. 1: Principal currently employed 3D culture systems

	Physiological relevance			*Technical complexity*			*Cost*			*Need of specific dynamic bioreactors* (●)	*Long-term cell survival/ maintenance of differentiated functions (microgravity)* (◊)
Cell suspension											
Gel- /matrix-based cultures											
Organotypic cultures *(homo-/ heterotypic multicellular spheroids)*										●	◊
Scaffold- /microcarrier-based cultures										●	◊
Tissue slices / explants										●	◊
Whole perfused organ										●	

: low **: intermediate** **: high**

Novakovic et al., 2002; Nickerson et al., 2007; Cosmi et al., 2009; Steimberg et al., 2009). Table 1 illustrates the principal characteristics of the most commonly used static and dynamic 3D culture systems.

In summary, providing the best life-like conditions for reproducing *in vitro* physiologically relevant tissue substitutes, 3D culture methods based on the use of RCCS^{TM} microgravity technology may represent a realistic alternative to animal-based model systems. Moreover, the possibility to use either normal or pathological cells or tissues of human origin, even obtained from poorly invasive surgical techniques (e.g. needle biopsies), the rapid development and optimisation of specific culture protocols and new analysis methods (with their subsequent standardisation and validation), will also offer to these models concrete prospects of filling the gap between animal- and human-based studies, opening also new opportunities for their application in basic research, pharmaco-toxicology and clinical fields.

References

Ayyaswamy, P. S. and Mukundakrishnan, K. (2007). Optimal conditions for simulating microgravity employing NASA designed Rotating Wall Vessels. *Acta Astronaut. 60*, 397-405.

Bhadriraju, K. and Chen, C. S. (2002). Engineering cellular microenvironments to improve cell-based drug testing. *Drug Discov. Today 7,* 612-620.

Bremer, S., Pellizzer, C., Hoffmann, S. et al. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programmes. *Curr. Pharm. Des. 13*, 3047-3058.

Catapano, G. and Gerlach, J. C. (2007). Bioreactors for liver tissue engineering. In N. Ashammakhi, R. Reis, and E. Chiellini (eds.). *Topics in Tissue Engineering 3*, 1-42. (http://www.oulu.fi/spareparts/ebook_topics_in_t_e_vol3/published_chapters.html).

Cohen, S. M. (2004). Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay. *Toxicol. Sci. 80*, 225-229.

Cosmi, F., Steimberg, N., Dreossi, D. and Mazzoleni, G. (2009). Structural analysis of rat bone explants kept in vitro in simulated microgravity conditions. *J. Mech. Behav. Biomed. Mater. 2*, 164-172.

EC Project Report (2009). *Alternative Testing Strategies – Progress Report 2009 – Replacing, reducing and refining use of animals in research (Genomics & Biotechnology for Health),* European Commission – Directorate General for Research, Life Sciences, Genomics and Biotechnology for Health., E. M. Nogueiro (ed.), European Communities, Brussells (http://cordis.europa.eu/fp7/health/library_en.html).

El-Ali, J., Sorger, P. K. and Jensen, K. F. (2006). Cells on chips. *Nature 442*, 403-411.

Hammond, T. G. and Hammond, J. M. (2001). Optimized suspension culture: the rotating-wall vessel. *Am. J. Physiol.-Renal Physiol. 281*, 12-25.

Hutmacher, D. W. and Singh, H. (2008). Computational fluid dynamics for improved bioreactor design and 3D culture. *Trends Biotechnol. 26*, 166-172.

ICATM (2009). International Cooperation on Alternative Test Methods (ICATM), April 2009 (http://iccvam.niehs.nih.gov/about/icatm.htm).

Khademhosseini, A., Langer, R., Borenstein, J. and Vacanti, J. P. (2006). Microscale technologies for tissue engineering and biology. *Proc. Natl. Acad. Sci. USA 103*, 2480-2487.

Klaus, D. M. (2001). Clinostats and bioreactors. *Gravit. Space Biol. Bull. 14*, 55-64.

Lilienblum, W., Dekant, W., Foth H. et al. (2008). Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH). *Arch. Toxicol. 82*, 211-236.

Martin, Y. and Vermette, P. (2005). Bioreactors for tissue mass culture: design, characterization, and recent advances. *Biomaterials 26*, 7481-7503.

MATES IWG (2007). *Advancing Tissue Science and Engineering: a Foundation for the Future. A Multi-Agency Strategic Plan*. The Multi-Agency Tissue Engineering Science (MATES) Interagency Working Group (IWG), Washington, D.C. (http://www.tissueengineering.gov/).

Mazzoleni, G., Di Lorenzo, D. and Steimberg, N. (2009). Modelling tissues in 3D: the next future of pharmaco-toxicology and food research? *Genes Nutr. 4*, 13-22.

Meuwly, F., Ruffieux, P. A., Kadouri, A. et al. (2007). Packed-bed bioreactors for mammalian cell culture: bioprocess and biomedical applications. *Biotechnol. Adv. 25*, 45-56.

Meyvantsson, I. and Beebe, D. J. (2008). Cell culture models in microfluidic systems. *Annu. Rev. Anal. Chem. 1*, 14.1-14.27.

Nickerson, C. A., Richter, E. G. and Ott, C. M. (2007). Studying host-pathogen interactions in 3-D: organotypic models for infectious disease and drug development. *J. Neuroimmune Pharmacol. 2*, 26-31.

NRC (2007). *Toxicity Testing in the 21st century. A vision and a strategy. Report of the Committee on Toxicity Testing and Assessment on Environmental Agents*. National Research Council of the National Academies. Washington, D.C.: The National Academies Press.

Pampaloni, F., Reynaud, E. G. and Stelzer, E. H. (2007). The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol. 8*, 839-845.

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*, London: Methuen & Co. Ldt.

Schwarz, R. P., Goodwin, T. J. and Wolf, D. A. (1992). Cell culture for three-dimensional modeling in rotating-wall vessels: an application of simulated microgravity. *J. Tissue Cult. Methods. 14*, 51-57.

Steimberg, N., Zarattini, G., Morandini, E. et al. (2009). 3D culture of articular cartilage explants: new perspectives for future possible clinical applications. *J. Bone Joint Surg. Br. 91-B (II)*, 269.

U.S. EPA (2009). *The U.S. Environmental Protection Agency's Strategic Plan for Evaluating the Toxicity of Chemicals*. Washington, DC.: U.S. Environmental Protection Agency (http://www.epa.gov/spc/toxicitytesting/index.htm).

Unsworth, B. R. and Lelkes, P. I. (1998). Growing tissues in microgravity. *Nat. Med.* 4, 901-907.

Vunjak-Novakovic, G., Searby, N., De Luis, J. and Freed, L. E. (2002). Microgravity studies of cells and tissues. *Ann. N. Y. Acad. Sci. 974*, 504-517.

Acknowledgements

This work has been partly supported by European Union grant LSHB-CT-2006-037168 (EXERA).

Correspondence to

Prof. Giovanna Mazzoleni, MD
General Pathology & Immunology Unit
Department of Biological Sciences and Biotechnologies,
School of Medicine
University of Bresciaviale Viale Europa, 11
25123 Brescia, Italy
e-mail: giovanna.mazzoleni@med.unibs.its

The Emerging Field of Human Umbilical Cord Blood Cell Transplantation

Dong-Hyuk Park[1,2], Alison E. Willing[1], Svitlana Garbuzova-Davis[1], Jun Tan[3], Cesario V. Borlongan[1], Nicole Kuzmin-Nichols[4], L. Eduardo Cruz[5] and Paul R. Sanberg[1,6]

[1]Center of Excellence for Aging and Brain Repair, Department of Neurosurgery and Brain Repair, University of South Florida College of Medicine, Tampa, FL, USA; [2]Department of Neurosurgery, Korea University Medical Center, Korea University College of Medicine, Seoul, Korea; [3]Rashid Laboratory for Developmental Neurobiology, Silver Child Development Center, University of South Florida, Tampa, FL, USA; [4]Saneron CCEL Therapeutics, Inc., Tampa, FL, USA; [5]Cryopraxis and Silvestre Laboratory, Cryopraxis, BioRio, Pólo de Biotecnologia do Rio de Janeiro, Brazil; [6]Office of Research and Innovation, University of South Florida, Tampa, FL, USA

Summary

Over the past several years, scientists have demonstrated the therapeutic potential of stem cells derived from human umbilical cord blood (hUCB) using in vitro and in vivo models for stroke and other ischemic diseases. They have revealed hUCB cells' pluripotency to differentiate into all cell types, including neural lineage, as well as the cells' ability to provide a number of neurotrophic and angiogenetic factors and to modulate inflammatory reaction. Unique immune naïveté of hUCB cells also makes them more flexible to allogeneic transplantation compared to other stem cell candidates. These multipotential abilities of hUCB stem cells, which were demonstrated by the in vivo studies, could be almost predicted by the findings observed in the in vitro models. Therefore, numerous in vitro modeling studies will be necessary to answer undetermined properties and to demonstrate the therapeutic effectiveness of hUCB cells.

Keywords: human umbilical cord blood cells, stroke, transplantation, in vitro models

The discovery of stem cells, including cells of both embryonic and adult origins, has brought the field of cell-based therapies into the forefront of promising treatments for intractable disorders such as ischemic diseases, including stroke and myocardial infarction (Park et al., 2009). Nonetheless, clinical use of certain adult stem cells, including neural stem cells (NSCs), or embryonic and fetal originated cells, is still hindered by safety and efficacy problems as well as bioethical and technical issues (Sathananthan and Trounson, 2005). By contrast, human umbilical cord blood (hUCB), which offers several advantages (Tab. 1), has already been applied clinically to repopulate bone marrow (BM) and blood cell lineages in various pediatric hematological malignant and nonmalignant diseases (Sirchia and Rebulla, 1999).

hUCB is a reservoir of hematopoietic, endothelial cell, mesenchymal, and multi/pluripotent stem cells (Kim et al., 2004; Berger et al., 2006). In particular, hUCB-derived hematopoietic stem cells have a number of the most primitive cells and are enough to reconstitute blood lineages of a recipient, even from a single donor (Broxmeyer et al., 1992; Nayar et al., 2002). Interestingly, two different subpopulations in mononuclear fractions of hUCB have been identified, adherent and floating (Chen et al., 2005). While a considerable number of stem cell antigens are expressed on cells in the floating population, the adherent cell population mainly includes lymphocytes expressing hematopoietic antigens. The relative immunological immaturity of hUCB is also noteworthy. hUCB has a high ratio of immature T-cells but low levels of mature memory (Harris et

Tab. 1: Advantages of umbilical cord blood stem cells compared to adult origin for cell based therapies

- Immature immunoregulatory and pro-inflammatory function
- Possible anti-inflammatory effect
- Multiple secretory functions including factors
- Low incidence of GvHD and infection transmission
- Relatively easy procurement and storage
- Immediate feasibility
- No burden to donors

al., 1992; D'Arena et al., 1998) and cytotoxic (D'Arena et al., 1998) T-cells. Lymphocytes in hUCB express pro-inflammatory cytokines and their receptors to a lesser degree than those in adult blood cells (Gluckman and Rocha, 2005), whereas they produce more interleukin (IL)-10, an anti-inflammatory cytokine, than those in adult blood cells (Rainsford and Reen, 2002; Gluckman and Rocha, 2005). This immaturity of immunological properties leads to a low incidence of graft versus host disease and viral transmission. Also, such unique cellular composition allows for relatively flexible donor-recipient matching conditions, potentially decreasing the length of time required to identify and procure the needed cells for transplantation (Newcomb et al., 2007).

Several *in vitro* and *in vivo* investigations have revealed that hUCB cells can transdifferentiate to express various neural markers for early neural precursors (musashi-1, nestin, TuJ1), mature neurons (NeuN, MAP-2), and astrocytes (glial fibrillary acidic protein), as well as oligodendrocytes (galactosylceramide) when grown in specific conditioned media containing the neuromorphogens, retinoic acid or nerve growth factor (Sanchez-Ramos et al., 2001; Bicknese et al., 2002; Buzanska et al., 2002). Recently, NSCs were derived from hUCB cells and differentiated toward mature neurons in three-dimensional environments (Jurga et al., 2009). Moreover, cultured hUCB cells in non-inducing standard growth media also showed expression of neural antigens (Garbuzova-Davis et al., 2003). Chen et al. (Chen et al., 2001) demonstrated the role of intravenous administration of mononuclear hUCB cells in a middle cerebral artery occlusion (MCAO) induced stroke rat model. Implanted hUCB cells significantly improved neurological functions and were found mainly in the ischemic boundary zone. Some grafted cells were immunoreactive for the endothelial cell marker as well as for neural markers, including glial fibrillary acidic protein, MAP-2, and NeuN. The findings of this *in vivo* study were similar to the results of *in vitro* studies, thus supporting the claim that *in vivo* investigations may be predicted by *in vitro* experiments.

Various parameters of hUCB cell transplantation for stroke have been studied by our groups. First, we demonstrated that intravenous injection of mononuclear hUCB cells was more effective than direct striatal implantation in producing long-term functional benefits to the stroke animal (Willing et al., 2003). Next, we established that there is a dose effect of mononuclear hUCB cells on the behavioral recovery and infarct volume in MCAO rats (Vendrame et al., 2004). With respect to the therapeutic window of cell transplantation for stroke, augmented migration in mononuclear hUCB cells towards the hippocampal and striatal extracts harvested at 24-72 hours after stroke has been demonstrated (Newman et al., 2005). On the basis of this *in vitro* investigation, it was predicted that, although the current therapeutic window for the treatment of stroke victims using tissue plasminogen activator is three hours, the therapeutic window may be extended to 24-72 hours after a stroke event by using mononuclear hUCB cell therapy. In a subsequent *in vivo* study, when hUCB cells were systemically administered at times ranging from three hours to 30 days post MCAO, treatment at 48 hours showed maximal improvements (Newcomb et al., 2006), thus supporting the *in vitro* prediction.

Furthermore, there has been increasing evidence of the multifaceted, bystander effects of hUCB on the injured brain, including neurotrophic, angiogenic, and anti-inflammatory actions in the MCAO stroke models. The hUCB treatment decreases microglia/macrophages and the pro-inflammatory cytokines such as tumor necrosis factor-α and IL-1β in the brain after stroke (Vendrame et al., 2005). Additionally, transplantation of hUCB cells induces vascular growth factors and, in turn, neovascularization, leading to functional improvement in the animal ischemic disease model (Taguchi et al., 2004; Das et al., 2009; Yerebakan et al., 2009). These effects likely result from secretion of trophic and other factors from the hUCB cells (Borlongan et al., 2004). Taken together, the preponderance of evidence suggests that hUCB cells will be a good first candidate for cell-based therapies, providing multiple therapeutic effects in a single transfusion that no other pharmacological agent could mimic, assuming that the present issues, such as safety concerns associated with using hUCB for non-hematopoietic disorders (Alvarez-Mercado et al., 2009), hUCB expansion capabilities (Koestenbauer et al., 2009) after conventional storage techniques (Lu et al., 2008), and limited graft survival (Vendrame et al., 2004) can be adequately addressed in the near future. Numerous *in vitro* modeling studies will be necessary to answer many of these questions and to demonstrate the therapeutic effectiveness of hUCB.

Disclosure Statement

AEW, SGD, JT, CVB and PRS are consultants of Saneron. PRS is also its Founder and a cunsultant of Cryopraxis. Research discussed here was supported, in part, by the NIH.

References

Alvarez-Mercado, A. I., Garcia-Mediavilla, M. V., Sanchez-Campos, S. et al. (2009). Deleterious Effect of Human Umbilical Cord Blood Mononuclear Cell Transplantation on Thioacetamide-Induced Chronic Liver Damage in Rats. *Cell Transplant. 18*, 1069-1079.

Berger, M. J., Adams, S. D., Tigges, B. M. et al. (2006). Differentiation of umbilical cord blood-derived multilineage progenitor cells into respiratory epithelial cells. *Cytotherapy 8*, 480-487.

Bicknese, A. R., Goodwin, H. S., Quinn, C. O. et al. (2002). Human umbilical cord blood cells can be induced to express markers for neurons and glia. *Cell Transplant. 11*, 261-264.

Borlongan, C. V., Hadman, M., Sanberg, C. D. and Sanberg P. R. (2004). Central nervous system entry of peripherally injected umbilical cord blood cells is not required for neuroprotection in stroke. *Stroke 35*, 2385-2389.

Broxmeyer, H. E., Hangoc, G., Cooper, S. et al. (1992). Growth characteristics and expansion of human umbilical cord blood and estimation of its potential for transplantation in adults. *Proc. Natl. Acad. Sci. U. S. A. 89*, 4109-4113.

Buzanska, L., Machaj, E. K., Zablocka, B. et al. (2002). Human cord blood-derived cells attain neuronal and glial features in vitro. *J. Cell Sci. 115*, 2131-2138.

Chen, J., Sanberg, P. R., Li, Y. et al. (2001). Intravenous administration of human umbilical cord blood reduces behavioral deficits after stroke in rats. *Stroke 32*, 2682-2688.

Chen, N., Hudson, J. E., Walczak, P. et al. (2005). Human umbilical cord blood progenitors: the potential of these hematopoietic cells to become neural. *Stem Cells 23*, 1560-1570.

D'Arena, G., Musto, P., Cascavilla, N. et al. (1998). Flow cytometric characterization of human umbilical cord blood lymphocytes: immunophenotypic features. *Haematologica 83*, 197-203.

Das, H., Abdulhameed, N., Joseph, M. et al. (2009). Ex vivo nanofiber expansion and genetic modification of human cord blood-derived progenitor/stem cells enhances vasculogenesis. *Cell Transplant. 18*, 305-318.

Garbuzova-Davis, S., Willing, A. E., Zigova, T. et al. (2003). Intravenous administration of human umbilical cord blood cells in a mouse model of amyotrophic lateral sclerosis: distribution, migration, and differentiation. *J. Hematother. Stem Cell Res. 12*, 255-270.

Gluckman, E. and Rocha, V. (2005). History of the clinical use of umbilical cord blood hematopoietic cells. *Cytotherapy 7*, 219-227.

Harris, D. T., Schumacher, M. J., Locascio, J. et al. (1992). Phenotypic and functional immaturity of human umbilical cord blood T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A. 89*, 10006-10010.

Jurga, M., Lipkowski, A. W., Lukomska, B. et al. (2009). Generation of functional neural artificial tissue from human umbilical cord blood stem cells. *Tissue Eng. Part C Methods 15*, 365-372.

Kim, J. W., Kim, S. Y., Park, S. Y. et al. (2004). Mesenchymal progenitor cells in the human umbilical cord. *Ann. Hematol. 83*, 733-738.

Koestenbauer, S., Zisch, A., Dohr, G. and Zech, N. H. (2009). Protocols for hematopoietic stem cell expansion from umbilical cord blood. *Cell Transplant. 18*, 1059-1068.

Lu, X., Proctor, S. J. and Dickinson, A. M. (2008). The effect of cryopreservation on umbilical cord blood endothelial progenitor cell differentiation. *Cell Transplant. 17*, 1423-1428.

Nayar, B., Raju, G. M. and Deka, D. (2002). Hematopoietic stem/progenitor cell harvesting from umbilical cord blood. *Int. J. Gynaecol. Obstet. 79*, 31-32.

Newcomb, J. D., Sanberg, P. R., Klasko, S. K. and Willing, A. E. (2007). Umbilical cord blood research: current and future perspectives. *Cell Transplant. 16*, 151-158.

Newcomb, J. D., Ajmo, C. T., Jr., Sanberg, C. D. et al. (2006). Timing of cord blood treatment after experimental stroke determines therapeutic efficacy. *Cell Transplant. 15*, 213-223.

Newman, M. B., Willing, A. E., Manresa, J. J. et al. (2005). Stroke-induced migration of human umbilical cord blood cells: time course and cytokines. *Stem Cells Dev. 14*, 576-586.

Park, D. H., Borlongan, C. V., Willing, A. E. et al. (2009). Human Umbilical Cord Blood Cell Grafts for Stroke. *Cell Transplant. 18*, 985-998.

Rainsford, E. and Reen, D. J. (2002). Interleukin 10, produced in abundance by human newborn T cells, may be the regulator of increased tolerance associated with cord blood stem cell transplantation. *Br. J. Haematol. 116*, 702-709.

Sanchez-Ramos, J. R., Song, S., Kamath, S. G. et al. (2001). Expression of neural markers in human umbilical cord blood. *Exp. Neurol. 171*, 109-115.

Sathananthan, A. H. and Trounson, A. (2005). Human embryonic stem cells and their spontaneous differentiation. *Ital. J. Anat. Embryol. 110*, 151-157.

Sirchia, G. and Rebulla, P. (1999). Placental/umbilical cord blood transplantation. *Haematologica 84*, 738-747.

Taguchi, A., Soma, T. and Tanaka, H. et al. (2004). Administration of CD34+ cells after stroke enhances neurogenesis via angiogenesis in a mouse model. *J. Clin. Invest. 114*, 330-338.

Vendrame, M., Gemma, C., de Mesquita, D. et al. (2005). Anti-inflammatory effects of human cord blood cells in a rat model of stroke. *Stem Cells Dev. 14*, 595-604.

Vendrame, M., Cassady, J., Newcomb, J. et al. (2004). Infusion of human umbilical cord blood cells in a rat model of stroke dose-dependently rescues behavioral deficits and reduces infarct volume. *Stroke 35*, 2390-2395.

Willing, A. E., Lixian, J., Milliken, M. et al. (2003). Intravenous versus intrastriatal cord blood administration in a rodent model of stroke. *J. Neurosci. Res. 73*, 296-307.

Yerebakan, C., Sandica, E., Prietz, S. et al. (2009). Autologous Umbilical Cord Blood Mononuclear Cell Transplantation Preserves Right Ventricular Function in a Novel Model of Chronic Right Ventricular Volume Overload. *Cell Transplant. 18*, 855-868.

Correspondence to

Paul R. Sanberg
Center of Excellence for Aging and Brain Repair
Department of Neurosurgery and Brain Repair
University of South Florida College of Medicine
MDC 78, 12901 Bruce B. Downs Boulevard Tampa
FL 33612, USA
e-mail: psanberg@health.usf.edu

Authors

ALTEX

A journal for new paths in biomedical science

Vol. 27 (Spec. Issue 1/2010)

www.altex.ch
http://altweb.jhsph.edu/altex/

Official organ of: CAAT, Center for Alternatives to Animal Testing, Johns Hopkins University, Baltimore, USA; CAAT-EU, University of Konstanz, Germany; EUSAAT, European Society for Alternatives to Animal Testing, Vienna, Austria; t4 – transatlantic think tank for toxicology, Baltimore, USA, Konstanz, Germany, Utrecht, The Netherlands; Doerenkamp-Zbinden Chairs in Germany, India, The Netherlands, Switzerland and USA

Address:
Society ALTEX Edition
Weinmanngasse 86,
8700 Kuesnacht ZH, Switzerland
Phone: +41 44 3800830
Fax: +41 44 3800832
e-mail: editor@altex.ch

ALTEX Editorial Office USA:
Thomas Hartung, Carol Howard,
Michael Hughes, Baltimore, MD, USA

Address:
CAAT
615 N Wolfe St., W7032 Baltimore,
MD 21205, USA
Phone: +1 410 614 4990, Fax: +1 410 614 2871
e-mail: thartung@jhsph.edu

ALTEX Online:
www.altex-edition.org
Altweb: http://altweb.jhsph.edu
ALTEX is indexed in MEDLINE, Current Contents®, SciSearch® and ISI Document Solution® and eMBASe.

Layout: H. P. Hoesli, Zurich, Switzerland

Print:
Stuertz GmbH, Wuerzburg, Germany

Distribution:
ALTEX, Weinmanngasse 86
8700 Kuesnacht ZH, Switzerland
Phone: +41 44 380 0830; Fax: +41 44 380 0832
Fax: +49 7531 25833
e-mail: subs@altex.ch

Publisher:
Spektrum Akademischer Verlag.
Springer Science + Business Media.

Edition:
1,500 DVD's

Circulation:
5 times per annum (four English and one German issues) plus supplements

Subscription rates:
Full Subscription (four English issues plus ALTEXethik):
Normal: 90 €
Library: 170 € (companies, institutes, libraries)
Student: 52 € (please fax proof of student status)
Animal Protection and Scientific Societies: 52 €

Engl.-Subscription (four English issues):
Normal: 80 €
Library: 160 € (companies, institutes, libraries)
Student: 42 € (please fax proof of student status)
Animal Protection and Scientific Societies: 42 €

NEW: ALTEXethik subscription
ALTEXethik: 15 € (German issue only)

All prices include postage and handling for all European countries and the USA.
Additional costs for shipment outside of Europe and USA: 20 € for 4 issues, 25 € for 5 issues.

Please ask for special subscription rates for India

Aims and Scope – Protecting More Than Animals

ALTEX, edited by the Swiss Society ALTEX Edition, is the official journal of CAAT, the Center for Alternatives to Animal Testing at the Johns Hopkins University in Baltimore, USA, EUSAAT, the European Society for Alternatives to Animal Testing (former MEGAT), t4, the transatlantic think tank for toxicology (Baltimore, Utrecht, Konstanz) and the Doerenkamp chairs in Germany, India, The Netherlands, Switzerland and USA. ALTEX is devoted to the publication of research on the development and promotion of alternatives to animal experiments according to the 3R concept of Russell and Burch: Replace, Reduce and Refine.

ALTEX publishes original articles, short communications, and reviews, as well as news and comments, meeting reports and book reviews. Animal experiments are defined by the editors as all experimental procedures involving the use of animals in testing, research and education or to obtain tissues, organs, and other animal derived products that may cause pain or distress to animals.

In addition to covering the biomedical aspects of animal experimentation, ALTEX is devoted to the bioethics of the complex relationship between humans and animals. Articles published in ALTEX should convey a basic respect for the dignity of living creatures and a recognition of animals as our partners. Articles devoted to the social and ethical aspects of this topic will be judged according to stringent scientific standards.

Manuscripts submitted to ALTEX are evaluated by two reviewers. The evaluation takes into account the scientific merit of a manuscript and the contribution to animal welfare and the 3R-principle.