Food for Thought ...
Read-Across Approaches – Misconceptions, Promises and Challenges Ahead

Grace Patlewicz [1], Nicholas Ball [2], Richard A. Becker [3], Ewan D. Booth [4], Mark T. D. Cronin [5], Dinant Kroese [6], David Steup [7], Ben van Ravenzwaay [8] and Thomas Hartung [9]

[1]DuPont Haskell Global Centers for Health and Environmental Sciences, Newark, DE, USA; [2]Toxicology & Environmental Research and Consulting (TERC), The Dow Chemical Company, Midland, MI, USA; [3]Regulatory and Technical Affairs Department, American Chemistry Council (ACC), Washington, DC, USA; [4]Syngenta Ltd, Jealott's Hill International Research Centre, Bracknell, Berkshire, UK; [5]Liverpool John Moores University, School of Pharmacy and Biomolecular Sciences, Liverpool, UK; [6]TNO, Zeist, The Netherlands; [7]Shell Oil Company, Shell Health Risk Science Team, Houston, TX, USA; [8]BASF Aktiengesellschaft, Experimental Toxicology and Ecology, Ludwigshafen, Germany; [9]Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, USA and University of Konstanz, CAAT-Europe, Germany

Summary

Read-across is a data gap filling technique used within category and analogue approaches. It has been utilized as an alternative approach to address information requirements under various past and present regulatory programs such as the OECD High Production Volume Programme as well as the EU's Registration, Evaluation, Authorisation and restriction of CHemicals (REACH) regulation. Although read-across raises a number of expectations, many misconceptions still remain around what it truly represents; how to address its associated justification in a robust and scientifically credible manner; what challenges/issues exist in terms of its application and acceptance; and what future efforts are needed to resolve them. In terms of future enhancements, read-across is likely to embrace more biologically-orientated approaches consistent with the Toxicity in the 21st Century vision (Tox-21c). This Food for Thought article, which is notably not a consensus report, aims to discuss a number of these aspects and, in doing so, to raise awareness of the ongoing efforts and activities to enhance read-across. It also intends to set the agenda for a CAAT read-across initiative in 2014-2015 to facilitate the proper use of this technique.

Keywords: analogue approach, category approach, (Q)SARs, adverse outcome pathways (AOP), scientific confidence framework

1 Introduction

Read-across is not a novel concept but in recent years – and certainly in the run up to the REACH regulation (EC, 2006) – there was a concerted effort to leverage existing guidance for its use and establish a consistent set of considerations that could form the basis of the REACH technical guidance. Van Leeuwen et al. (2007) briefly summarized these considerations, highlighting experiences within the UK agencies (Hanway and Evans, 2000), the OECD and US EPA High Production Volume (HPV) programs (OECD, 2002; US EPA, 2004), and the US EPA's New Chemicals Program (US EPA, 2010), among others. The development of technical guidance for REACH was organized in partnership with the OECD to produce a comprehensive guidance document addressing the EU regulatory needs while also being sufficiently general to accommodate regulatory programs in other member countries. One of the first issues addressed was the need to provide clear definitions of terms associated with read-across, which can be summarized briefly as: the analogue approach, which is based on a chemical group with a very limited number of structurally similar substances (usually a target and source substance), and the category approach, which is based on a more extensive number of structurally similar analogues. In contrast, the term read-across was defined as a data gap filling technique within an analogue or category approach. Thus, a read-across represents a qualitative

http://dx.doi.org/10.14573/altex.1410071

or quantitative prediction made within a grouping approach. Other data gap filling techniques include trend analysis as well as external (Q)SARs.

The definitions and the workflows associated with both analogue and category approaches are described in the REACH guidance (ECHA, 2008) and the original OECD guidance (OECD, 2007), as well as the revised OECD guidance (OECD, 2014a). Pertinent definitions from various sources are provided in Box 1, pointing to a further need for terminology harmonization.

Read-across as a data gap filling technique raises many expectations. Many still believe that a "read-across" is a simple inference of endpoint information between 2 structurally similar substances – one with experimental data and one without – and that the justification is based on structural similarity alone. The latter actually describes the process of making a prediction based on a structural alert (SAR) for a specific endpoint, e.g., predicting mutagenicity on the basis of a chemical containing a nitro aromatic moiety. In contrast, while a read-across could conceivably start from a structural alert as a means of identifying structurally related analogues, the overall process of substantiating the similarity both structurally and biologically involves a weight of evidence (WoE) assessment of many different pieces of information, both directly and indirectly related to the endpoint data gap under consideration. As such, the effort needed to construct a robust and credible read-across justification is not trivial.

2 Read-across considerations before and during development

Under regulatory programs such as REACH there are a number of considerations to take into account before even undertaking a read-across approach. The purpose of REACH is to address a specific information requirement, i.e., produce information to characterize an endpoint by way of a study, and therefore a number of additional factors come into play as described by Patlewicz et al. (2013a). The two main areas for consideration are practical and scientific – the number of data gaps and pertinent endpoints. If the number of data gaps is minimal – a heuristic could be less than 3 – then an analogue or category approach may not be warranted. Individual data gaps could be conceivably filled by using either individual (Q)SAR models, *in vitro* assays, or a combination of both. However, this strongly depends on the endpoints being evaluated. A data gap for a "simpler" endpoint, such as skin or eye irritation or *in vitro* mutagenicity, could be readily addressed using a combination of (Q)SAR and/or *in vitro* techniques. However, if the data gap were for an endpoint that would ordinarily merit higher tier testing in animals (such as a 90-day study), then an analogue/category approach to build up a WoE rationale, which would justify consistency in effects across relevant endpoints for the studied analogues, would be more appropriate. Clearly, although the number of data gaps is important, the type of data gaps will drive the practical gap filling strategy. Another practical issue concerns legitimate

Box 1

Known definitions of chemical categories and read-across

1. The European Commission (2006): "Read-across approach: Prediction from data for reference substance(s) within the group or 'category' of substances by interpolation to other substances in the group" (Ferrario et al., 2014).
2. US EPA: "Read Across from Analogs/Categories – 'Read across' is a technique of filling data gaps. To 'read across' is to apply data from a tested chemical for a particular property or effect (cancer, reproductive toxicity, etc.) to a similar untested chemical. The read across technique is often applied within groups of similar chemicals assembled for assessment using either analog approach (grouping based on a very limited number of chemicals) or category approach (grouping based on a larger number of chemicals). In an analog/category approach, not every chemical needs to be tested for every endpoint."
3. ECHA: "Read-across is an approach for filling data gaps, either by using a category or an analogue approach. For the purposes of the REACH Regulation (Article 13(1)), read-across is considered by ECHA to be an alternative method."
4. ECHA: "Category: Group of substances with physicochemical, toxicological and ecotoxicological properties that are likely to be similar or follow a regular pattern as a result of structural similarity."
5. OECD (2014a): "Chemical category: A group of chemicals whose physico-chemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristic)."
6. OECD (2014a): "In the read-across technique, endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered to be 'similar' in some way (usually on the basis of structural similarity) ... Within a group of chemicals, read-across can be performed in the following ways to fill data gaps:

   a) one-to-one (one analogue used to make an estimation for a single chemical)

   b) many-to-one (two or more analogues used to make an estimation for a single chemical)

   c) one-to-many (one analogue used to make estimations for two or more chemicals)

   d) many-to-many (two or more analogues used to make estimations for two or more chemicals)."

access to data; there are important implications depending on whether "read-across" data is available (and at what financial cost) as well as whether the data is of sufficient quality. If there is data for a source analogue but the existing studies are systematically[1] rated as low quality (using a scale such as the Klimisch score (e.g., K3 or K4) (Klimisch et al., 1997)), then it is questionable whether the overall information will be sufficient to build a WoE or whether new data would still be needed.

A key consideration, from a technical standpoint, is whether there is a plausible hypothesis, and what the ease and cost of substantiating the hypothesis might be. Existing biological evidence (e.g., from existing *in vivo* studies or *in vitro* studies) might be able to support a hypothesis. Another consideration is whether the use of a read-across approach will allow accurate and credible assessment of the hazards (e.g., toxicological endpoints) under consideration – whether the outcome is too conservative or not precautionary enough both have implications in practice. One must also consider the consequences and subsequent costs of the read-across approach not being accepted. Consequences may be the loss of, or delay to, market and monetary costs of additional studies needed to support the read-across.

Defining a plausible overarching hypothesis supporting read-across is a key scientific requirement. There are three main facets to this – outlining a hypothesis on why two or more structurally similar substances should be grouped together, determining which endpoints this applies to, and justifying relevance for each endpoint. An alternative is to consult the initial considerations made to evaluate the suitability of a source analogue for inclusion in an analogue/category approach. Indeed, Wu et al. (2010) devised an approach for determining analogue suitability, which was tested in a series of different case studies by Blackburn et al. (2011). The overarching hypothesis will typically factor in structural similarity; this, however, is only one aspect of many. Other general considerations will address commonality in functional groups, the likelihood of common precursors and breakdown products. Furthermore, whenever biological evidence is available to support (or refute) the hypothesis (e.g., *in vitro* studies demonstrating interaction with receptors, enzymes, protein reactivity, etc.), this must also be taken into account.

Beyond such considerations, attention must also be given to toxicokinetic aspects. It is important to emphasize bioavailability when deciding which data are necessary for a read-across. If there is evidence that a material is not absorbed, then there is little need to pursue it further and generate a collection of (inevitably) negative test data. These general considerations will allow an evaluation of the likely bioavailability, metabolism and reactivity. Bioavailability may be crudely estimated by the log of the octanol-water partition coefficient ($LogK_{ow}$) or using an absorption algorithm. Examples of such models for dermal absorption include those by ten Berge (2009) or Dancik et al. (2013). Metabolism in terms of the metabolic pathway might be estimated by *in silico* tools (such as those included in the OECD Toolbox) in the absence of empirical data and relevant reactivity considerations may reference established organic chemistry reaction principles (Aptula and Roberts, 2006). These sorts of considerations are critical because structural similarity (using indices such as Tanimoto, 1957; Willett et al., 1998) may give rise to very "similar" structures in terms of the commonality in "bits" but result in substances that will behave very differently on account of their chemistry. For example, an alcohol and an aldehyde could conceivably be grouped together based on a Tanimoto index, but their chemical behavior would be very different.

The next step is to construct a matrix of available data to explore existing data for each of the analogues initially selected and determine how the outcomes of the different endpoints of concern align. Assuming consistency between analogues for each endpoint, the read-across justification also needs to consider what endpoint-specific similarities (i.e., mechanistic knowledge) can be used to justify the commonality in behavior. This is easier for some endpoints than for others. For simpler endpoints, a wealth of information can be drawn from (Q)SAR or *in vitro* data. Skin sensitization, for example: There is an abundance of mechanistic knowledge about the induction of skin sensitization as summarized in the adverse outcome pathway (AOP) recently published by the OECD (2012). In turn, much of this information has been encoded into software tools such as TIMES (Patlewicz et al., 2014a) or the OECD (Q)SAR Toolbox, as well as into QSARs published in the literature (Roberts et al., 2008; Patlewicz and Worth, 2008).

Difficulties start to become apparent for the more complex multifactorial endpoints. For instance, how should a read-across of a 90-day study be justified? How much information is sufficient to credibly rationalize the similarity in behavior? A typical strategy under REACH has been to favor analogue approaches, as there is a perceived confidence in extrapolating between two substances where there is an extensive body of existing *in vivo* data with few data gaps for the target substance (ECETOC, 2012). Arguably, if there are consistent outcomes for the majority of (if not all) endpoints with both source and target data, a read-across for the required data gap (then using a WoE approach) appears to be a reasonable strategy – especially if coupled with inferred or empirical metabolic information.

Such strategies have been employed by industry registrants under REACH but their acceptability is, as yet, unclear. The outcomes of the 2010 dossiers submitted to the ECHA are only now becoming apparent; additionally only about 5% of these dossiers will be subject to a formal technical compliance evaluation. That said, for cases where read-across has been accepted or denied (see Ball et al., 2014), a number of questions have been raised. Is there a higher burden of proof depending on the presence or absence of adversity? Are some endpoints of greater concern than others? How is uncertainty identified and addressed in the read-across? Important future work will need to define what makes up a WoE approach justifying a read-across and what elements are considered critical – only then can the aspects/elements/steps in a read-across justification be linked to the associated uncertainty.

[1] Notably, there are now systematic tools available to assign Klimisch scores, i.e., the ToxRTool (Schneider et al., 2009) originating out of the Evidence-based Toxicology efforts (see www.ebtox.com, Hoffmann and Hartung, 2006; Hartung, 2009).

3 Uncertainty in read-across

3.1 Sources of uncertainty in read-across

There are several sources of possible uncertainty in a read-across – the presence or absence of an adverse effect, the type of endpoint, and the read-across prediction approach. Clearly, a greater burden of proof is required to support the use of read-across for the absence of an effect (see below). In theory, the same sound guiding principles are relied upon, but in practice the barrier to acceptance appears to be of greater significance when there is no ostensible hazard. The issue with so-called "negative read-across" is that there is a perception of greater uncertainty versus when reading across the presence of toxicity. When doing negative read-across, one could miss a significant effect even though the possibility of missing something exists in both cases. This assumes that the experimental data is of sufficient sensitivity in the first place. This type of uncertainty appeared to be a factor in the case of Ball et al. (2014) where there was no significant adversity for the analogues and endpoints under consideration. The specific endpoint under consideration also appears to be an issue. In Ball et al. (2014), the read-across was accepted for a 90-day study but the same underlying rationale was not accepted for the developmental toxicity endpoint.

There is also uncertainty by default in the read-across approach itself, since the target substance data is not relied upon. Assessment factors may be adopted to address uncertainty in a read-across prediction, and the magnitude of the assessment factor will depend on the nature of the uncertainty for the endpoint data under consideration. For example, extrapolating from a 28-day study to a 90-day study in the same species of rat is reasonable since there is sufficient evidence that the NOAEL decreases approximately threefold. However, very little experience is available to support an appropriate assessment factor for an extrapolation from a 28-day study in rats to a rabbit prenatal toxicity study. Thus, there may be justifiable reasons to use assessment factors in specific cases and for endpoints where a threshold can be defined.

Using a default assessment factor simply on account of the data being "read-across" is of itself not justified but will depend on, e.g., whether data comes from a single source chemical, i.e., this is more uncertain than when data is generated with the target chemical itself. If more source chemicals are available and used, and read-across is done by interpolation, then read-across may indeed be less uncertain as when it is generated with the target chemical itself. Arguably, a WoE based on a read-across approach could, in practice, offer more scientific confidence than relying on a single study for the target substance itself. It also serves to potentially identify effects that may not have been identified by performing the study.

3.2 "Interpolation" vs. "Extrapolation"

Another aspect of read-across is "Interpolation" vs. "Extrapolation," which was touched upon by ECHA in some of their comments on submissions. In some cases, read-across apparently was only acceptable when the substance(s) in question fell between two or more evaluated substances. As a hypothetical example, if one had a series of long-chain alcohols with data on a C_8 and a C_{16} molecule, in theory ECHA would accept interpolation of that data to a C_{12} substance but not necessarily extrapolation to a C_{18} analog. The pros and cons of such an argument were raised in the ECETOC report (ECETOC, 2012) as well as by Patlewicz et al. (2013a), and these were both referenced in the recently revised OECD grouping guidance (OECD, 2014a). More work needs to be performed to demonstrate that the scientific confidence driving the grouping approach is more critical than whether the read-across prediction is a result of an extrapolation or interpolation. This is anticipated to be one of the objectives of the CAAT read-across initiative (see below).

4 Validity of read-across

4.1 The concept of local validity

Ultimately, the key question is that of the validity of read-across. While we feel reasonably comfortable with approaches based on multiple substances defining an applicability domain for *in silico* predictions, most read-across exercises do not come with this associated luxury and they may also lack formal data association by relevant algorithms. This may explain why hazard-only determinations (yes/no decisions) are typically made by read-across. Perhaps, however, we need a change of viewpoint. The desire to develop an individual method to cover the entire chemical universe is probably naïve. Rarely will we find methods that are applicable to all chemicals, either because the properties of test substances are incompatible with our test system or because of the multitude of mechanisms to be reflected (or "cliffs," i.e., sudden changes in properties with changes in structure). However, we can usually rely on the accurate assessment of close neighbors in the chemical universe. So, instead of aiming for a large applicability domain, why not start with a smaller one and demonstrate validity for a similar group of chemicals? The manner in which the similar grouping is characterized will be a critical consideration. This might later be expanded to other groupings to form a region of validity.

The concept of the applicability domain was translated from the field of (Q)SAR to alternative methods in the "Modular Approach to Validation" (Hartung et al., 2004). At this stage, the concept describes the requirement to associate information with a validated method for chemicals whose validity had been shown. To date, this has not been used strategically, i.e., starting with a small applicability domain to validate a method and later expanding it. This approach would, however, require a tracking mechanism hosted by regulatory or validating agencies that would make test applicability information available in a continuously updated form.

On a smaller scale, local validity could be used in a WoE rationale to support read-across, grouping and "test-across" (i.e., using tests for which validity has been shown only for a certain group of chemicals, see below) approaches. Similar to validation, this would require assembly of the information available on standardization, reproducibility, and reliability of the method, with the reliability showing that consistent, relevant data are available for some representatives of this group of chemicals.

4.2 Negative read-across

The issue of absence of adversity is not new or distinct to read-across. Indeed, each and every test has its own limitations, though these are to a large degree recognized and addressed through the adoption of standardized test guidelines. We cannot show with absolute certainty, however, that a substance does not have a certain effect in a biological system. There might be species or inter-individual differences or co-factors not reflected in the reference study. We can also go back to philosophy, where Popper showed that we can only falsify but not verify (as laid out in Hartung et al., 2013b). This thinking can result in endless testing to rule out more and more such possibilities at the detriment of accumulating false-positive results. An alternative approach to this question may be the use of omics data obtained at a cellular level. In the absence of any subcellular response, the (extrapolated) absence of effects becomes plausible. In any case, logic dictates testing needs to stop when the data are very probably correct.

For negative read-across, we need to determine whether we are probably correct and this could be expressed as a probability of hazard. While exact probabilities might be difficult to estimate, we may be able to develop a scoring system or graphical representation, such as an inverse ToxPi (Reif et al., 2010). ToxPi was developed to visualize toxic liabilities as suggested by positive assays in the ToxCast battery at lower concentrations. Each slice of the ToxPi represents an assay or a compilation of assays; the slice grows the lower the concentrations is that triggers the assay. A putatively bad substance (positive in many assays at low concentrations) will thus have a very large pie. For read-across, an opposite scheme could conceivably be devised, where a substance starts with a full pie (we know nothing and thus have to assume the worst). The respective slice of the pie is then reduced with accumulating different evidence. There are many questions about such an approach and its utility, but it could be the start of some work by the CAAT read-across initiative (see below). A scoring system, where the quality of evidence provided is summarized, might be simpler. Although such a system raises many concerns regarding practical implementation, it may be a useful starting point to categorize substances as part of the initial hypothesis.

Another possibility is to consider other means of characterizing biological similarity. Efforts have been made to tackle this using transcriptomics or metabolomics. In ECETOC TR 109 (ECETOC, 2010) a case study illustrating the utility of mass-spectroscopy-based metabolomics in quantifying biological similarity was presented. This demonstrated the limitations of relying on structural similarity alone in read-across prediction. A pioneering example is the use of metabolomics for grouping (Bouhifd et al., 2013). BASF has been developing a database (MetaMap®Tox) in which metabolomics and toxicity data are evaluated in combination (van Ravenzwaay, 2007). The database contains data for more than 500 chemicals (from 28-day rat studies or studies of similar design) to generate metabolomic patterns for different toxicological targets (e.g., liver, kidney, thyroid, testes, blood, nervous and endocrine systems) and to combine these with the observed toxicological outcomes. The general use of metabolomics data in a broad context has been described by Ramirez et al. (2013). The metabolomics approach suggests toxicological modes of action at a relatively early time-point (within 7 days) and can, thereby, facilitate safety decisions and lower costs through a reduced need for animal studies. The use of such omics tools to enhance the quality of read-across has been discussed by van Ravenzwaay et al. (2012) (from QSAR to QBAR (quantitative biological activity relationships). An example of such biology based approaches can be seen by comparing the metabolome of di(2-ethyl-hexyl)phthalate and that of dibutylphthalate. Both compounds induce very similar forms of systemic toxicity (hepatomegaly with peroxisome proliferation, spermatogenesis with seminiferous tubule atrophy and vacuolization in the Sertoli cells), and their metabolome patterns are remarkably similar with more than 30 commonly regulated metabolites (van Ravenzwaay et al., 2010). In contrast, two structurally similar chemicals (2-AAF and 4-AAF; 2- resp. 4-acetyl-aminofluorene) have very significant differences in their toxicity profile (one is a liver carcinogen, the other not) and quite different metabolomic profiles. In a ranking process of similarity of the overall 2-AAF metabolome profiles with all other profiles in the data base, 4-AAF was found at rank 1080 and 798 (out of 1733 comparisons) in males and females respectively. Today, this approach requires generating *in vivo* data, e.g., a 28-day study. However, advances of *in vitro* metabolomics (Balcke et al., 2011) may help to overcome the current limited utility, perhaps in time for the 2018 REACH deadline for phase-in substances.

Other approaches are also being developed. As part of DECO[2] (a Cefic-LRI project under AIMT-3 "Predictive toxicology using 'omics', high-throughput data, and cheminformatics") a web interface known as DIAMONDs was created to facilitate the integration of different sources of data relevant for repeated dose toxicity focusing on the liver as a target organ for read-across. Alternatively, the biological phenotyping in ToxCast by US EPA could serve different purposes, including defining how to characterize a biological profile in terms of a fingerprint or using the activity as a means of setting a threshold beyond which no overt toxicity is observed. The former approach may assist in grouping chemicals *a priori* into mechanistically similar categories, while the latter approach might provide a better anchor for characterizing the absence of effects. To some extent, the current pivotal use of biological phenotyping for prioritization of chemicals, mainly for endocrine disruptor testing in the emerging EDSP21 program[3], is already a grouping exercise based on biological information. With the expansion of the substance pool to 10,000 chemicals within the Tox21 alliance among US agencies[4], this could become a tremendous database

[2] http://www.cefic-lri.org/uploads/Project%20publications/AIMT3%20poster%20DECO%20November%202012.pdf

[3] http://www.epa.gov/endo/pubs/edsp21_work_plan_summary%20_overview_final.pdf

[4] http://www.epa.gov/ncct/Tox21/

for "biological" read-across. Some issues must still be resolved, including: Assay applicability and alignment with apical endpoints; the absence of metabolism in the assays; compatibility of test materials with the assays; and the fact that the list of assays and their reproducibility in different labs is still very much under discussion.

Work to investigate some of these strategies is under discussion in projects such as SEURAT[5] and AIMT-4[6] (Use of non-animal data to supplement and strengthen read-across), an extension of AIMT-3 mentioned above.

4.3 UVCB substances

UVCB substances, i.e., chemical substances of unknown or variable composition, complex reaction products and biological materials, and how they are addressed via categories and read-across present a major challenge. There are inherent read-across issues in dealing with any UVCB, since even if one is fully tested, another sample of the same material will never be exactly the same. A pertinent example is petroleum streams, where extensive use of categories has been accepted in some cases (such as HPV) and proposed in others (particularly REACH). Duplicative testing of similar streams/mixtures would be particularly wasteful, but there is no universal agreement on how they should be grouped or how "similar" can be defined. This also connects to the issue of what is "close enough," since these UVCB can never be totally characterized or replicated.

5 Towards read-across enhancement

The issues surrounding read-across can be categorized as follows – practical considerations before embarking on a read-across and scientific considerations, such as justifying analogues for each endpoint in turn, identifying the uncertainties, and addressing the uncertainties to assure scientific confidence of the read-across for specific purposes. For the first two, much has been described in the following references: ECETOC (2012), Patlewicz et al., (2013a), and the revised OECD guidance (OECD, 2014a). In terms of identifying uncertainties, some of the issues and approaches were discussed during the 2012 ECHA-Cefic LRI workshop where ECHA outlined a read-across assessment framework (RAAF) (Patlewicz et al., 2013c). Subsequently, efforts have been made to develop similar frameworks, with Blackburn and Stuard (2014) outlining a systematic uncertainty framework for developmental and reproductive toxicants (DART). In the meantime, the Cefic LRI organizing committee refocused its efforts into a read-across team and began to formulate its own scientific confidence framework taking into account the principles outlined by Cox et al. (2014) for Tox-21c approaches and their prediction models, as well as building on the work of Blackburn and Stuard (2014). A framework to characterize scientific confidence of read-across is in development (Patlewicz et al., in prep), elements of which were communicated at QSAR 2014 (Aptula et al., 2014) and at the 9th World Congress on Alternatives and Animal Use in the Life Sciences (Ball et al., 2014). Neither the ECHA RAAF or the Blackburn and Stuard (2014) frameworks discuss how to resolve read-across uncertainties (aside from using default assessment factors), and understanding how to practically exploit other tools such as the Tox-21c approaches to address uncertainty is a major priority.

One strategy of exploiting Tox-21c approaches is through making use of the adverse outcome pathway (AOP) concept (Ankley et al., 2010) where an AOP represents "*existing knowledge concerning the linkage between the molecular initiating event (MIE) and an adverse outcome at the individual or population level*." The practical application of an AOP is through integrated approaches to testing and assessment (IATA) and integrated testing strategies (ITS) (Hartung et al., 2013a). The AOP thus summarizes the biological pathway, whereas the IATA is constructed based on the maturity and completeness of the associated AOP and the availability of non-testing or test systems – each characterizing different key events – to determine the applicability of the AOP for a given substance or group of substances. More information and discussion about the maturity and utility of an AOP and its associated IATA for different purposes are to be found in Tollefsen et al., (2014). The strategy of using an AOP-IATA in read-across is exemplified in the recent OECD Toolbox implementation of the AOP for skin sensitization (OECD, 2014b) and also by Patlewicz et al. (2014b) in their skin sensitization IATA pipeline. Both could be likened to a guided WoE approach. Hartung (2007) articulated some of the same principles in a "test-across" construct (see Fig. 1): "*A very interesting approach called 'read-across' is taken in the regulatory field. Here, results of sufficiently similar chemicals, for which animal test data are available, are used to intrapolate for a non-tested substance. In a similar manner, it should be possible to mini-validate an in vitro test for a given substance, i.e., by showing that related compounds are judged correctly, the result for a substance where there are no in vivo data that can be relied on. I would suggest to call this "test-across" (Fig. 1); this clearly represents an advantage over mere structure/relationships, since in addition actual testing in a living system is carried out. This might, at the same time, represent a solution in cases where no formal validation for the respective part of the chemical universe has been done (applicability domain) or where a full validation is not (yet) possible*." Examples of this concept, i.e., showing good correlation of *in vitro* test responses with *in vivo* observations for a group of chemicals are available with the zebrafish embryo test (ZET) (Beker van Woudenberg et al., 2013) and with a test battery for reproductive toxicity including ZET (Kroese et al., 2014).

Whilst AOP and related IATA represent a promising strategy for the future, at present there are only a handful of AOP in development and their application for particular uses through

[5] http://www.seurat-1.eu/; http://chemicalwatch.com/19594/seurat-1-homes-in-on-test-chemicals-for-read-across

[6] http://www.cefic-lri.org/request_for_proposals/1256744568/20/LRI-AIMT4-Use-of-non-animal-data-to-supplement-and-strengthen-read-across/?cntnt01orderby=rfp_date+DESC

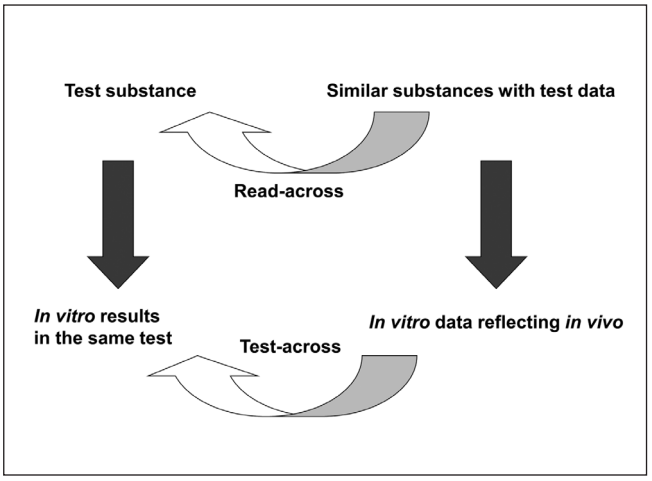

Fig. 1: The *in vitro* test-across concept (reproduced from Hartung, 2007)

IATA has received little attention. Nonetheless, AOP are expected to serve as an important approach for addressing uncertainties in read-across in the future. AOP and their role in read-across were discussed by the OECD in a workshop entitled "Using Mechanistic Information in Forming Chemical Categories" (OECD, 2011). The findings were subsequently incorporated into the OECD revised grouping guidance document (OECD, 2014a). A workshop on the application of AOP in regulatory contexts was organized by NICEATM/PCRM and held September 3-5, 2014, and OECD will hold a similar workshop focusing in on AOP-informed IATA on the November 17-19, 2014.

In addition to reaching a critical mass of useful AOP, addressing "negative read-across" will be a key challenge. AOP, by their nomenclature, appear anchored in adversity and the notion of being delineated by only "known knowledge." This might be a red herring, as in Patlewicz et al. (2013c) it was argued that AOP (in terms of their naming) were a misnomer, as the simplicity of the linear pathway obscured the nuances of dose response relationships between different key event relationships (KER) – hence, being able to use the associated data generated for both absence and presence of adversity within a continuum scale where appropriate. Whether the breadth of AOP covering the types of effects would be necessarily and sufficiently exhaustive to be able to make a decision about the absence of effects with sufficient confidence is a major question.

6 The work of CAAT and the read-across steering group

The final deadline under REACH for phase-in substances is June 2018. This coincides with many testing proposals now being evaluated from earlier phase-in substance deadlines. In contrast to the spirit of the legislation (Hartung, 2009), *in vitro* and *in silico* methods have played only a modest role in REACH submissions (Rovida et al., 2011), while grouping approaches are consistently praised by agencies for relieving the testing burden[7]. The ECHA report (ECHA, 2014) states, that a "read-across" or category approach was used in up to 75% of analyzed dossiers for at least one endpoint: "*When analysing dossiers that were not submitted as a category (i.e., without the use of the IUCLID category template), read-across was still found to be used more often than weight of evidence and calculated results/(Q)SARs (equivalent to 72%, 51% and 22% of the substances, respectively)*."

Read-across is considered by ECHA as an alternative method and it is arguably the only one used to a major extent. *In vitro* approaches are used to some extent for eye and skin irritation, but these contribute little to reducing animal numbers (Rovida and Hartung, 2009). Read-across is thus the single most important tool for the reduction of animal use and will continue to increase in importance. The final submission deadline in 2018 will include many more substances (several tens of thousands), which at the same time come with much less existing test data. In contrast to the earlier deadlines, in 2018 complete Chemical Safety Reports and no prior testing proposals must be submitted (Hartung, 2010), creating enormous pressure for short-term solutions and thus prompting the consideration to use read-across.

CAAT, in conjunction with various stakeholders, is developing a read-across product for release at the end of 2015. Its public presentation and discussion will be accompanied by a stakeholder forum similar to the roadmap for animal-free systemic toxicity testing exercise (Basketter et al., 2012; Leist et al., 2014).

An important activity that can take place in tandem (and would complement this work program) is the compilation of successful and unsuccessful read-across submissions under REACH. Currently, there is no way to determine whether the use of read-across in a REACH registration dossier was assessed and considered acceptable. For dossiers in which the use of read-across was rejected or accepted in part, the outcome and associated discussion is typically communicated to the registrants, captured in the minutes of the member states' committee meetings, and eventually uploaded to the ECHA website. For a third party, however, this information is difficult to collate and distill into useful data for the preparation of subsequent read-across justifications. Therefore, industry and scientific associations such as CEFIC, CONCAWE or ECETOC, should have member companies contribute their experiences (positive and negative) as case studies to incorporate as guidance for successfully utilizing read-across.

Notwithstanding the scientific challenges of read-across, and although this discussion has been biased towards the application and acceptance under REACH, there is a strong motivation to

[7] http://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2014_en.pdf

use read-across in regions outside of the EU, especially in view of global trade. The dialogue commenced with a session at the 9th World Congress on Animal Use and Alternatives (2014), but this dialogue will need to continue under the auspices of the CAAT read-across program.

7 Conclusions

Read-across is a data gap filling technique used within category and analogue approaches. It has been utilized extensively under REACH with mixed success as an alternative approach to address information requirements. There are a number of considerations that come into play when considering read-across, both practical and scientific. The practical challenges concern which data gap filling strategies are feasible and depend on the number and types of data gaps as well as the availability of high quality data. The scientific challenges include outlining an initial rationale for forming an analogue/category approach and justifying it effectively on both a general and endpoint-specific basis. Defining what merits sufficient justification to assure scientific confidence remains elusive with both endpoint and presence/absence of adversity being key uncertainties. Frameworks/constructs to make explicit the assumptions and considerations taken into account for read-across are in development by industry (Blackburn and Stuard, 2014; Patlewicz et al., in prep) and ECHA (Patlewicz et al., 2013b). However, while these begin to explicitly identify the uncertainties, addressing them with more than assessment factors or embracing the Tox-21c tools has not yet been attempted to any great extent. Using the construct of AOP (Ankley et al., 2010) should help build scientific confidence for read-across, particularly for more complex endpoints, by anchoring it to mechanistic knowledge and identifying types of testing and non-testing approaches to utilize when supporting read-across approaches. The AOP approach only goes so far – other Tox-21c approaches may be needed to address the scientific confidence for negative read-across cases. Biological phenotypical profiling, such as that conducted within ToxCast or using metabolomics approaches, offer practical alternatives. Some of these strategies may well be applied with rigor as part of the ongoing AIMT-4 and SEURAT-1 programs.

In an area where one solution does not fit all, as all chemicals come with different available data for themselves and their neighbors in the chemical universe, a fully standardized approach can hardly be developed. A set of principles and guidance, however, can advise the beginners, ease quality assurance and communication with regulators, and increase the overall confidence in the results.

The CAAT read-across program intends to foster a wider dialogue on the utility and promise of such approaches to facilitate and promote their acceptance for REACH 2018. This, however, is only one avenue of engagement. Extending this dialogue beyond REACH and the EU to generate awareness and promote outreach in other regions such as China, Korea and Brazil (among others) is critically needed.

References

Ankley, G. T., Bennett, R. S., Erickson, R. J. et al. (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem 29*, 730-741. http://dx.doi.org/10.1002/etc.34

Aptula, A. O. and Roberts, D. W. (2006). Mechanistic applicability domains for non animal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem Res Toxicol 19*, 1097-1105. http://dx.doi.org/10.1021/tx0601004

Ball, N., Bartels, M. J., Budinsky, R. A. et al. (2014). The challenge of using read-across within the EU REACH regulatory framework; how much uncertainty is too much? Dipropylene Glycol Methyl Ether Acetate, an exemplary case study. *Reg Toxicol Pharmacol 68*, 212-221. http://dx.doi.org/10.1016/j.yrtph.2013.12.007

Balcke, G. U., Kolle, S. N., Kamp, H. et al. (2011). Linking energy metabolism to dysfunctions in mitochondrial respiration – a metabolomics in vitro approach. *Toxicol Lett 203*, 200-209. http://dx.doi.org/10.1016/j.toxlet.2011.03.013

Basketter, D. A., Clewell, H., Kimber, I. et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX 29*, 3-89. http://dx.doi.org/10.14573/altex.2012.1.003

Beker van Woudenberg, A., Wolterbeek, A., Te Brake, L. et al. (2013). A category approach to predicting the developmental (neuro) toxicity of organotin compounds: the value of the zebrafish (Danio rerio) embryotoxicity test (ZET). *Reprod Toxicol 41*, 35-44. http://dx.doi.org/10.1016/j.reprotox.2013.06.067

Blackburn, K., Bjerke, D., Daston, G. et al. (2011). Case studies to test: a framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul Toxicol Pharmacol 60*, 120-135. http://dx.doi.org/10.1016/j.yrtph.2011.03.002

Blackburn, K. and Stuard, S. B. (2014). A framework to facilitate consistent characterization of read across uncertainty. *Regul Toxicol Pharmacol 68*, 353-362. http://dx.doi.org/10.1016/j.yrtph.2014.01.004

Bouhifd, M., Hartung, T., Hogberg, H. T. (2013). Review: toxicometabolomics. *J Appl Toxicol 33*, 1365-83. http://dx.doi.org/10.1002/jat.2874

Cox, L. A., Douglas, D., Marty, S. et al. (2014). Applying a scientific confidence framework to a HTS-derived prediction model for endocrine endpoints: lessons learned from a case study. *Reg Toxicol Pharmacol 69*, 443-450. http://dx.doi.org/10.1016/j.yrtph.2014.05.010

Dancik, Y., Miller, M. A., Jaworska, J. and Kasting, G. B. (2013). Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. *Adv Drug Deliv Rev 65*, 221-236. http://dx.doi.org/10.1016/j.addr.2012.01.006

ECETOC (2010). Technical Report 109 High information content technologies in support of read-across in chemical risk

assessment. http://www.ecetoc.org/technical-reports

ECETOC (2012). Technical Report 116 Category approaches, read-across, (Q)SAR. http://www.ecetoc.org/technical-reports

EC – European Commission (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. *Off J Eur Union L396*, 1-849. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R1907&from=EN

ECHA (2008). Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

ECHA (2014). The Use of Alternatives to Testing on Animals for the REACH Regulation: Second report under Article 117(3) of the REACH Regulation. http://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2014_en.pdf

Ferrario, D., Brustio, R. and Hartung, T. (2014). Glossary of reference terms for alternative test methods and their validation. *ALTEX 31*, 319-335. http://dx.doi.org/10.14573/altex.1403311

Hanway, R. H. and Evans, P. F. (2000). Read-across of toxicological data in the notification of new chemicals. *Toxicol Lett 116, Suppl 1*, 61.

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim 32*, 467-72.

Hartung, T. (2007). Food for thought … on cell culture. *ALTEX 24*, 143-147. http://www.altex.ch/All-issues/Issue.50.html?iid=87&aid=1

Hartung, T. (2009). Food for thought … on evidence-based toxicology. *ALTEX 26*, 75-82. http://www.altex.ch/All-issues/Issue.50.html?iid=105&aid=1

Hartung, T. (2010). Food for thought … on alternative methods for chemical safety testing. *ALTEX 27*, 3-14. http://www.altex.ch/All-issues/Issue.50.html?iid=112&aid=1

Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013a). Integrated testing strategies for safety assessments. *ALTEX 30*, 3-18. http://dx.doi.org/10.14573/altex.2013.1.003

Hartung, T., Stephens, M. and Hoffmann, S. (2013b). Mechanistic validation. *ALTEX 30*, 119-130. http://dx.doi.org/10.14573/altex.2013.2.119

Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human Exp Toxicol 25*, 497-513. http://dx.doi.org/10.1191/0960327106het648oa

Klimisch, H.-J., Andreae, M. and Tillmann, U. (1997). Systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Reg Toxicol Pharmacol 25*, 1-5. http://dx.doi.org/10.1006/rtph.1996.1076

Kroese, E. D., Bosgra, S., Buist, H. E. et al. (2014). Evaluation of an alternative in vitro test battery for detecting reproductive toxicity in a grouping context. *Reprod Toxicol*, in press.

Leist, M., Hasiwa, N., Rovida, C. et al. (2014). Consensus report on the future of animal-free systemic toxicity testing. *ALTEX 31*, 341-156. http://dx.doi.org/10.14573/altex.1406091

OECD (2002). Manual for investigation of HPV Chemicals. Section 3.2. Guidance on the development and use of chemical categories. Organisation for Economic Co-operation and Development, Paris, France.

OECD (2007). Guidance on grouping of chemicals. OECD Series on Testing and Assessment No. 80. Organisation for Economic Co-operation and Development, Paris, France.

OECD (2011). OECD Series on Testing and Assessment Number 138. Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. ENV/JM/MONO(2011) 8. Organisation for Economic Co-operation and Development. Paris, France.

OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 1: Scientific Evidence. Series on Testing and Assessment No. 168 ENV/JM/MONO(2012)10/PART1

OECD (2014a). Guidance on grouping of chemicals. OECD Series on Testing and Assessment No. 194. Organisation for Economic Co-operation and Development, Paris, France.

OECD (2014b). How to use the Toolbox AOP workflow for Skin Sensitization. http://bit.ly/1tlBOtf

Patlewicz, G. and Worth, A. (2008). Review of data sources, QSARs and integrated testing strategies for skin sensitisation. EUR 23225 EN. European Commission, JRC, IHCP, Ispra, Italy.

Patlewicz, G., Ball, N., Booth, E. D. et al. (2013a). Use of category approaches, read-across and (Q)SAR: general considerations. *Reg Toxicol Pharmacol 67*, 1-12. http://dx.doi.org/10.1016/j.yrtph.2013.06.002

Patlewicz, G., Roberts, D. W., Aptula, A. et al. (2013b). Workshop: Use of 'read-across' for chemical safety assessment under REACH. *Reg Toxicol Pharmacol 65*, 226-228. http://dx.doi.org/10.1016/j.yrtph.2012.12.004

Patlewicz, G., Simon, T., Goyak, K. et al. (2013c). Use and validation of HT/HC assays to support 21st century toxicity evaluations. *Reg Toxicol Pharmacol 65*, 259-268. http://dx.doi.org/10.1016/j.yrtph.2012.12.008

Patlewicz, G., Kuseva, C., Mehmed, A. et al. (2014a). TIMES-SS – recent refinements as a result of an industrial skin sensitisation consortium. *SAR QSAR Environ Res 25*, 367-391. http://dx.doi.org/10.1080/1062936X.2014.900520

Patlewicz, G., Kuseva, C., Kesova, A. et al. (2014b). Towards AOP application – implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Reg Toxicol Pharmacol 69*, 529-545. http://dx.doi.org/10.1016/j.yrtph.2014.06.001

Ramirez, T., Daneshian, M., Kamp, H. et al. (2013). Metabolomics in toxicology and preclinical research. *ALTEX 30*,

209-225. http://dx.doi.org/10.14573/altex.2013.2.209

Reif, D., Martin, M. T., Tan, S. et al. (2010). Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ Health Perspect 118*, 2-8. http://dx.doi.org/10.1289/ehp.1002180

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 sensitization potential. *J Appl Toxicol 28*, 443-454. http://dx.doi.org/10.1002/jat.1293

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements. *ALTEX 26*, 187-208. http://www.altex.ch/All-issues/Issue.50.html?iid=107&aid=4

Rovida, C., Longo, F. and Rabbit, R. R. (2011). How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? *ALTEX 28*, 273-294. http://dx.doi.org/10.14573/altex.2011.4.273

Schneider, K., Schwarz, M., Burkholder, I. et al. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicol Lett 189*, 138-144. http://dx.doi.org/10.1016/j.toxlet.2009.05.013

Tanimoto, T. T. (1957). IBM Internal Report 17th Nov.

Ten Berge, W. (2009). A simple dermal absorption model: derivation and application. *Chemosphere 75*, 1440-1445. http://dx.doi.org/10.1016/j.chemosphere.2009.02.043

Tollefsen, K. E., Scholz, S., Cronin, M. T. et al. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Reg Toxicol Pharmacol*, in press. http://dx.doi.org/10.1016/j.yrtph.2014.09.009

US EPA (2004). Status and future directions of the High Production Volume Challenge programme. US Environmental Protection Agency. Office of Pollution Prevention and Toxics, Washington, DC.

US EPA (2010). TSCA New Chemicals Program (NCP). USEPA, Office of Pollution Prevention and Toxics, Washington, DC. http://www.epa.gov/oppt/newchems/pubs/chemcat.htm

van Leeuwen, C. J., Patlewicz, G. Y. and Worth, A. P. (2007). Intelligent Testing Strategies. In C. J. Leeuwen and T. Vermeire (eds.), *Risk Assessment of Chemicals: An Introduction* (467-509). Springer.

van Ravenzwaay, B., Coelho-Palermo Cunha, G., Leibold, E. et al. (2007). The use of metabolomics for the discovery of new biomarkers of effects. *Toxicol Lett 172*, 21-28. http://dx.doi.org/10.1016/j.toxlet.2007.05.021

van Ravenzwaay, B., Coelho-Palermo Cunha, G., Strauss, V. et al. (2010). The individual and combined metabolite profiles (metabolomics) of dibutylphthalate and di(2-ethylhexyl)phthalate following a 28-day dietary exposure in rats. *Toxicol Lett 198*, 159-170. http://dx.doi.org/10.1016/j.toxlet.2010.06.009

van Ravenzwaay, B., Herold, M., Kamp, H. et al. (2012). Metabolomics: a tool for early detection of toxicological effects and an opportunity for biology based grouping of chemicals-from QSAR to QBAR. *Mutat Res 746*, 144-150. http://dx.doi.org/10.1016/j.mrgentox.2012.01.006

Willett, P., Barnard, J. M. and Downs, G. M. (1998). Chemical similarity searching. *J Chem Inf Comput Sci 38*, 983-996. http://dx.doi.org/10.1021/ci9800211

Wu, S., Blackburn, K., Amburgey, J., Jaworska, J. and Federle, T. (2010). A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul Toxicol Pharmacol 56*, 67-81. http://dx.doi.org/10.1016/j.yrtph.2009.09.006

Acknowledgements

The perspectives and discussion in this manuscript represent those from individual members of the Johns Hopkins Center for Alternatives to Animal Testing (CAAT) read-across steering group chaired by Grace Patlewicz of DuPont and Thomas Hartung of CAAT. We thank all members for their insights and contributions. We would like to particularly thank Katy Goyak and Jim Freeman of Exxon Mobil, Volker Strauss of BASF, as well as Karen Blackburn and Jorge Naciff of P&G for their contributions. A number of the authors are employees of chemical companies or of the American Chemical Council (ACC), and received no other compensation but their salary for this work.

Correspondence to

Thomas Hartung, MD, PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
W7032, Baltimore, MD 21205, USA
e-mail: thartung@jhsph.edu