Die Entwicklung von Prädiktionsmodellen für drei *in vitro* Embryotoxizitätstests im Rahmen einer ECVAM Validierungsstudie

Elke Genschow[1], Gabriele Scholz[2], Nigel A. Brown[3], Aldert H. Piersma[4], Madeleine Brady[5], Nicole Clemann[6], Hannele Huuskonen[7], Françoise Paillard[8], Susanne Bremer[9] aund Horst Spielmann[1]
[1]BgVV, D-Berlin, [2]ZEBET im BgVV, D-Berlin, [3]St. Georges Hospital Medical School, GB-London, [4]RIVM, NL-Bilthoven, [5]ZENECA, GB-Macclesfield, [6]NOVARTIS, CH-Basel, [7]KTL, FIN-Kuopio, [8]Synthélabo Recherche, F-Gargenville, [9]ECVAM, I-Ispra

Zusammenfassung
Seit 1997 koordiniert ZEBET im Auftrag des EU Validierungszentrums ECVAM eine Prävalidierungs- und Validierungsstudie von drei in vitro Embryotoxizitätstests: WEC-Test (Test mit ganzen Rattenembryonen), Micromass- oder MM-Test (Test mit primären Zellen der Extremitätenknospen von Rattenembryonen) und Embryonaler Stammzelltest, EST (Test mit einer embryonalen Stammzell-Linie der Maus). In der zur Zeit laufenden Validierungsstudie wird jeder der drei Tests unter blinden Bedingungen in vier Laboratorien geprüft. Bisher wurden in der Validierung sechs von 20 Prüfsubstanzen mit unterschiedlichen embryotoxischen Eigenschaften in vivo (stark, schwach und nicht embryotoxisch) getestet. Die Ergebnisse wurden zur Verbesserung der Prädiktionsmodelle (PM) zur Vorhersage embryotoxischer Eigenschaften von Prüfsubstanzen im WEC-Test und im MM-Test verwendet, sowie zur Überprüfung des Prädiktionsmodelles für den EST. Bei der Entwicklung des PM für den EST wurden 94% der Experimente der Lernstichprobe richtig klassifiziert (Scholz et al.,1999), und in dieser Validierungsstudie wurden 80% der Experimente entsprechend der in vivo Embryotoxizität der Testsubstanzen korrekt eingestuft. Mit Hilfe der während der Validierung entwickelten PM für den MM-Test und für den WEC-Test wurden jeweils 79% korrekte Klassifizierungen erreicht. Da im WEC-Test nur Parameter des Wachstums und der Entwicklung von Embryonen bestimmt werden, nicht aber Zytotoxizitätsdaten, wurden in einem zweiten PM für den WEC-Test (PM2) Zytotoxizitätsdaten der 3T3 Maus-Fibroblasten Zell-Linie aus dem EST berücksichtigt. Dadurch wurde die korrekte Klassifizierung im WEC-Test auf 96% verbessert.

Summary: Development of prediction models for three in vitro embryotoxicity tests which are evaluated in an ECVAM validation study
In 1997 ZEBET started the co-ordination of a study funded by the European Centre for the Validation of Alternative Methods (ECVAM) with the aim of prevalidation and validation of three in vitro embryotoxicity tests. These tests employ the cultivation of postimplantation whole rat embryos (WEC test), cultures of primary limb bud cells of rat embryos (micromass test, MM-Test), and cultures of a pluripotent mouse embryonic stem cell line (embryonic stem cell test, EST). In the current Validation Study each of the tests is evaluated in four laboratories under blind conditions. In an initial phase of the validation study six out of 20 test chemicals comprising different embryotoxic potential (non, weak and strong embryotoxic) were tested. The results were used to improve the prediction models (PM) for the WEC test and the MM-Test in order to identify the embryotoxic potential of test chemicals. In addition, the existing PM for the EST was evaluated using the results from testing of the initial six chemicals. The PM for the EST was developed using the results of a previous prevalidation study (Scholz et al.,1999), in which 94% of the learning sample were classified correctly. Applying this PM to the results of the initial phase of the current validation study, 80% of the experiments were classified correctly according to the embryotoxic potential of the tested chemicals in vivo. Applying the PM for the MM-Test and the WEC test that were developed during the current validation study in both tests, 79% correct classifications were achieved. Since the PM of the WEC-Test took into account only parameters of growth and development, but not cytotoxicity data, a second PM (PM2) for the WEC was developed that was improved by incorporating cytotoxicity data of the differentiated mouse fibroblast cell line 3T3 derived from the EST. This approach, which has previously never been used as an adjunct to the WEC test, resulted in an increase of correct classifications to 96%.

Keywords: in vitro *embryotoxicity, embryotoxic potential, embryonic stem cell test, rat whole embryo culture, micromass culture, differentiation, cytotoxicity, prevalidation study, validation study, prediction model, discriminant analysis*

1 Einleitung

Für Industriechemikalien, deren Produktion unter einer Tonne pro Jahr liegt bzw. unter fünf Tonnen insgesamt, müssen in der EU aufgrund der Chemikaliengesetzgebung keine Prüfungen zur Embryo- bzw. Reproduktionstoxikologie durchgeführt werden. Deshalb besteht bei den Behörden in Europa zusätzlicher Bedarf an Informationen zur Reproduktionstoxikologie von Industriechemikalien. Um diese für den Verbraucherschutz wichtige Informationslücke zu schließen, sind die Behörden in Europa in zunehmendem Maße bereit, die Ergebnisse von *in vitro* Embryotoxizitätstests für die sicherheitstoxikologische Risikoabschätzung heranzuziehen. Bisher sind derartige Tests jedoch noch nicht ausreichend experimentell validiert. Unabhängig von dieser Entwicklung setzt die Arzneimittelindustrie bei der Entwicklung neuer Arzneistoffe beim sogenannten „high throughput screening" (HTS) fast ausschließlich *in vitro* Tests ein, und zwar sowohl zur Identifizierung neuer Wirkstoffe mit erwünschten Wirkungen als auch um Stoffe mit unerwünschten, d.h. toxischen, Eigenschaften frühzeitig und kostensparend zu identifizieren. Beim HTS wird der Identifikation von Stoffen mit embryotoxischen Eigenschaften eine hohe Priorität eingeräumt, um diese bereits vor der Prüfung *in vivo,* im belastenden und teuren Tierversuch, beurteilen zu können.

Für das HTS, für das sich in erster Linie Zellkulturverfahren (z.B. Zytotoxizitätstests mit permanenten Zellinien) eignen, konnte bisher jedoch noch kein *in vitro* Embryotoxizitätstest hinreichend standardisiert werden. Falls ein solcher *in vitro* Test entwickelt wird, könnten beispielsweise bei der Arzneimittelentwicklung Substanzen mit hohem embryotoxischem Potential frühzeitig identifiziert und aussortiert werden. Darüber hinaus kann man innerhalb einer Wirkstoffgruppe, wie z.B. den Retinoiden, die embryotoxischen Eigenschaften in einem *in vitro* Test (*ranking*) vergleichen, um Prioritäten bei der Auswahl der Stoffe für die klinische Weiterentwicklung zu setzen. Dies wurde z.B. bei der Entwicklung neuer Retinoide in der Arzneimittelindustrie mit dem MM-Test über lange Zeit praktiziert (Kistler, 1987; Kistler et al.,1990). Jedoch konnte bisher nicht ausreichend evaluiert werden, ob der MM-Test in ähnlicher Weise auch zur Bewertung der embryotoxischen Eigenschaften anderer Wirkstoffgruppen geeignet ist (Spielmann, 1998).

Zwar wurden bereits einige *in vitro* Embryotoxizitätstests, wie z.B. Tests mit Hühnerembryonen, Rattenembryonen und im MM-Test in Ringversuchen in mehreren Laboratorien mit einer begrenzten Zahl von Prüfsubstanzen unter blinden Bedingungen geprüft (Kucera et al.,1993; Piersma et al.,1996), jedoch wurden diese Studien nicht entsprechend den heute gültigen strengen internationalen Empfehlungen für Validierungsstudien durchgeführt (Balls et al.,1990, 1995; OECD, 1996). So wurde in diesen Validierungsstudien nur teilweise nach einheitlichen Standardprotokollen gearbeitet, und darüber hinaus fehlten in diesen Arbeitsvorschriften Positiv- und Negativkontrollsubstanzen sowie biometrisch fundierte Prädiktionsmodelle (Curren et al.,1995). Das führte in den Validierungsstudien zu einer mangelnden Vergleichbarkeit der Reproduzierbarkeit der Ergebnisse, die in den einzelnen Laboratorien erarbeitet wurden. Dies ist sicherlich der Hauptgrund dafür, daß es, ähnlich wie bei der Validierung von Ersatzmethoden zum Draize-Test am Kaninchenauge, bis heute nicht gelungen ist, eine wissenschaftliche oder behördliche Anerkennung auch nur eines einzigen validierten *in vitro* Embryotoxizitätstests zu erreichen (Spielmann, 1998).

Aus den genannten Gründen wurde vom EU-Validierungszentrum ECVAM 1994 ein Workshop über *in vitro* Tests in der Reproduktionstoxikologie zum Screening von Chemikalien durchgeführt (Brown et al.,1995). Bei diesem Workshop wurde dringend empfohlen, *in vitro* Tests auf dem Gebiet der Reproduktionstoxikologie entsprechend den formalen Empfehlungen von ECVAM zu validieren (Balls et al.,1995; Curren et al., 1995). Die Teilnehmer des ECVAM Workshops empfahlen nachdrücklich, bereits existierende *in vitro* Embryotoxizitätstests zu validieren oder neue zu entwikkeln. Da die Bewertung der vorliegenden Daten aus *in vitro* Embryotoxizitätstests, die mit niederen Wirbeltierspezies, wie z.B. Fröschen und Vögeln, nicht positiv ausfiel, haben die Teilnehmer des ECVAM Workshops die Durchführung einer Validierungsstudie mit drei vielversprechenden *in vitro* Tests empfohlen, in denen Zellen und Gewebe von Säugetierembryonen eingesetzt werden. Bei den drei *in vitro* Embryotoxizitätstests handelt es sich um den WEC-Test (*rat whole embryo culture test*), den MM-Test (micromass test) und den EST (Embryonaler Stammzelltest). Entsprechend dieser Empfehlung werden die drei Tests in jeweils 4 Laboratorien in Europa experimentell validiert. Die Validierungsstudie wird von ECVAM finanziert und von ZEBET koordiniert. In der Studie werden unter blinden Bedingungen 20 Testchemikalien geprüft, für die sowohl *in vivo* Daten aus Tierversuchen als auch vom Menschen von sehr hoher Qualität vorhanden sind.

Es ist das generelle Ziel dieser Validierungsstudie, zu prüfen, ob die drei *in vitro* Tests in der Lage sind, das embryotoxische Potential eines Stoffes für den Menschen vorherzusagen. Entsprechend der Ausschreibung darf die Validierungsstudie erst nach erfolgreicher Prävalidierung der drei *in vitro* Embryotoxizitätstests durchgeführt werden. In der Phase der Prävalidierung muß entsprechend den Empfehlungen von ECVAM (Curren et al.,1995) ein Standardprotokoll für jeden der drei *in vitro* Tests erarbeitet werden, das ein biostatistisch fundiertes Prädiktionsmodell (PM) enthält. Für den EST wurde bereits ein PM erarbeitet (Spielmann et al.,1997; Scholz et al.,1999). Dagegen wurden bisher sowohl für den MM-Test als auch für den WEC noch keine PM erarbeitet, obwohl beide Tests in einer größeren Zahl von Laboratorien schon seit langer Zeit etabliert sind (Brown et al.,1995; Spielmann, 1998). Aus den genannten Gründen ist es das wesentliche Ziel der Prävalidierungsphase, PMs experimentell und mit biometrischer Unterstützung zu etablieren. In der formalen Validierungstudie unter blinden Bedingungen sind die experimentelle Überprüfung der PM in unterschiedlichen Laboratorien und die Prüfung der Reproduzierbarkeit der Ergebnisse der Tests die beiden wesentlichen Ziele der Validierungsstudie, über deren erste Phase wir an dieser Stelle berichten möchten.

2 Material und Methoden

2.1 Embryonaler Stammzelltest (EST)

Embryonale Stammzellen (ES-Zellen) der Maus können in Zellkultur mit Hilfe des *Leukemia Inhibitory Factor* (LIF) im undifferenzierten Stadium kultiviert werden (Williams et al.,1988). Der bei ZEBET entwickelte EST (Spielmann et al.,1997) nutzt die Fähigkeit dieser ES-Zellen, *in vitro* unter definierten Kulturbedingungen ohne die Einwirkung von LIF in spontan kontrahierende Herzmuskelzellen zu differenzieren. Der Einfluß von Testsubstanzen auf die Differenzierung von Herzmuskelzellen, die sich aus ES-Zellen der Linie D3 entwickeln (Kontraktionen, die bei mikroskopischer Auswertung registriert werden), wird verglichen mit dem im MTT-Test ermittelten zytotoxischen Effekt auf ES-Zellen und auf differenzierte Fibroblasten-Zellen der Mauszell-Linie 3T3. Der MTT Test beruht auf der Fähigkeit mitochondrialer Dehydrogenasen in lebenden Zellen, die gelb gefärbte Substanz 3-(4,5-dimethylthiazol-2yl)-2,5-diphenyl tetrazolium bromide (MTT) in ein dunkelblaues Formazan umzuwandeln, das quantitativ mit dem ELISA Reader gemessen wird. Die Zytotoxizität wird durch die Halbhemmkonzentration im Vergleich zur Kontrolle bestimmt. Die IC_{50} Konzentrationen für die Zytotoxizität und die Differenzierung werden anhand von Konzentrations-Wirkungs-Kurven bestimmt. Dabei dienen die IC_{50} Konzentrationen für den zytotoxischen Effekt auf 3T3-Zellen und ES-Zellen (IC_{50}D3, IC_{50}3T3) und die IC_{50} Konzentrationen für die Inhibition der Differenzierung (ID_{50}) von ES-Zellen zur Klassifizierung der Testsubstanzen in die drei Embryotoxizitätsklassen *nicht*, *schwach* oder *stark embryotoxisch* mit Hilfe eines PM, das durch Diskriminanzanalyse erstellt wurde (Spielmann et al.,1997; Scholz et al.,1999).

2.2 Micromass (MM)-Test

Am Tag 14 der Schwangerschaft werden Ratten die Embryonen entnommen, deren Extremitätenknospen unter dem Mikroskop isoliert wurden. Nach Dissoziation der Zellen werden diese in hoher Zelldichte (= Micromass) unter definierten Bedingungen kultiviert, die eine Differenzierung in Knorpelzellen erlauben. Nach 6tägiger Kultur in Gegenwart ansteigender Konzentrationen von Testsubstanzen wird geprüft, ob die embryonalen MM-Zellen eine normale Differenzierung durchlaufen haben. Dies geschieht durch eine für differenzierte Knorpelzellen spezifische Färbemethode mit Alcian Blau (Amacher et al.,1996; Flint und Orton, 1984; Kistler, 1987; Uphill et al.,1990). Der Einfluß der Behandlung mit Testsubstanzen auf die Differenzierung der Knorpelzellen wird verglichen mit dem rein zytotoxischen Effekt derselben Testsubstanzen auf die MM-Zellen der Extremitätenknospen, der durch Färbung mit Neutralrot bestimmt wird. Die IC_{50} Konzentrationen für die Zytotoxizität und die Differenzierung werden anhand von Konzentrations-Wirkungs-Kurven bestimmt.

2.3 Kultur ganzer Rattenembryonen (WEC-Test)

Rattenweibchen werden am Tag 10 der Schwangerschaft die Embryonen entnommen und unter definierten Bedingungen bei ansteigenden Konzentrationen von Testsubstanzen kultiviert. Eine große Zahl morphologischer und entwicklungsspezifischer Parameter wird zu Beginn und nach 48stündiger Entwicklung der Embryonen *in vitro* dokumentiert. Den Beobachtungen werden je nach ihrer Ausprägung Zahlenwerte auf einer Skala von 0-6 (*morphological score*) zugeordnet, die eine quantitative Auswertung der gewonnenen Daten in Konzentrations-Wirkungs-Kurven ermöglichen. Die Summe der Einzelwerte für die verschiedenen *morphological scores* werden als *total morphological score* (TMS) erfaßt und dienen als Maß für das Stadium der normalen Entwicklung zum Zeitpunkt der Auswertung nach 48 Stunden. Spezifität und Häufigkeit von Mißbildungen bestimmter Organe und Organanlagen werden gesondert dokumentiert (Amacher et al.,1996; Bechter und Schön, 1988; Brown und Fabro, 1981; Piersma et al.,1995). Dadurch kann man zwischen spezifischen Effekten der Testchemikalien auf die Entwicklung bestimmter Organe (z. B. des Neuralrohrs) und einer generellen Wachstumsverzögerung (z.B. Kopf-Rumpf Länge) unterscheiden.

2.4 Testchemikalien

In der Validierungsstudie werden 20 Testsubstanzen und die positive Kontrollsubstanz 5-Fluorouracil untersucht. Die Testsubstanzen sind in Tabelle 1 unter Angabe ihres embryotoxischen Potentials *in vivo* aufgelistet. Für die Auswahl der Testchemikalien wurden folgende Kriterien angewendet: Die Anzahl der Testchemikalien pro Klasse: nicht, schwach und stark embryotoxisch sollte in etwa gleich sein und die Qualität der *in vivo* Daten sowohl im Tierversuch als auch im Menschen hoch sein.

Die unabhängige Kodierung und der Versand der Prüfsubstanzen wurde von der Fa. *RCC Cytotest Cell Research* (D-Roßdorf) vorgenommen. Die Testchemikalien wurden von den Firmen Merck (D-Darmstadt) und Sigma (D-Deisenhofen) bezogen mit folgenden Ausnahmen: all-trans-Retinsäure wurde von Dr. Heinrich Bürgin (Hoffmann La Roche, CH-Basel) zur Verfügung gestellt, Isobutyl-ethyl-VPA und Pentyl-4-yn-VPA wurde von Prof. Heinz Nau (Zentrumsabteilung für Lebensmitteltoxikologie der Tieräztlichen Hochschule, D-Hannover) bereitgestellt.

Wie in Validierungsstudien üblich, müssen die teilnehmenden Labors davon ausgehen, daß jede Substanz temperatur- und lichtempfindlich ist und der höchsten getesteten Toxizität entspricht. Die Lagerung und Verarbeitung der Testchemikalien muß entsprechend erfolgen.

2.5 Biostatistik

Die lineare Diskriminanzanalyse wurde verwendet, um diejenigen Variablen zu identifizieren, die die beste Trennung zwischen den drei embryotoxischen Klassen ergab (schrittweise Diskriminanzanalyse mit SPSS, Norušis, 1994; Backhaus et al.,1996; Bortz, 1993). Die schrittweise Diskriminanzanalyse wurde für alle drei *in vitro* Tests angewendet.

Als höchste zu testende Konzentration sind 1000µg/ml festgelegt worden. Wird bei dieser Konzentration kein IC_{50} bzw. ID_{50} Wert erreicht, wird per Definition der IC_{50} bzw. ID_{50} Wert auf 1000µg/ml gesetzt. Die Konsequenz ist, daß künftig die höchste zu testende Konzentration 1000µg/ml ist.

3 Ergebnisse

Die der formalen Validierung vorausgegangene Prävalidierungsstudie, die nach den Vorgaben von ECVAM (Curren et al.,1995) durchgeführt wurde, hatte ge-

zeigt, daß die Standard Protokolle (SOP's) der drei *in vitro* Embryotoxizitätstests erfolgreich in ein zweites Labor übertragen und reproduzierbare Ergebnisse erzielt werden können. Mit den drei Methoden konnten die in der Prävalidierungsphase getesteten Testsubstanzen entsprechend ihres embryotoxischen Potentials eingestuft werden (Scholz et al.,1998).

Die initiale Phase der formalen Validierungsstudie, in der die ersten sechs von 20 Testsubstanzen unter blinden Bedingungen getestet wurden, wurde 1998 abgeschlossen. Jeder der drei *in vitro* Tests wurde in jeweils vier europäischen Laboratorien geprüft. In einer Validierungsstudie müssen die Testsubstanzen bis zum Abschluß der Studie kodiert bleiben. Um einerseits das PM des MM-Tests und des WEC-Tests biostatistisch abzusichern, das PM für den EST zu evaluieren und andererseits den blinden Charakter der Studie zu erhalten, wurde für die Analyse der PM in der vorliegenden Arbeit lediglich das embryotoxische Potential der ersten sechs Testsubstanzen von *RCC Cytotest Cell Research* dekodiert.

Die Ergebnisse der verbleibenden 14 Testsubstanzen werden nach Abschluß der Validierungsstudie im Januar 2000 zur Evaluierung der PM des WEC-Tests und des MM-Tests verwendet, während für die Evaluierung des PM für den EST alle 20 Testsubstanzen zur Verfügung stehen. Nachfolgend werden die Prädiktionsmodelle für den WEC-Test und den MM-Test und die Klassifizierungsergebnisse des PM für den EST vorgestellt.

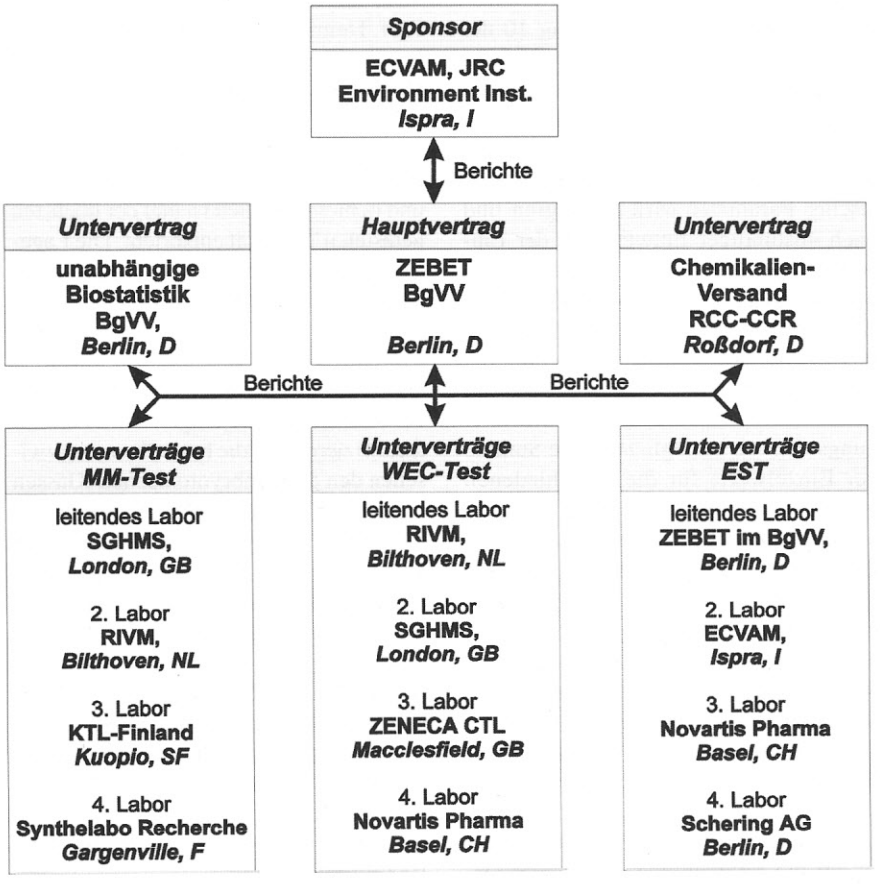

Abbildung 1: Struktur der Validierungsstudie. Dargestellt sind die in die Validierungsstudie involvierten Vertrags- und Untervertragsnehmer. Die Untervertragsnehmer berichten dem Hauptvertragsnehmer ZEBET über ihre Ergebnisse, die von ZEBET zusammengefaßt und an den Sponsor ECVAM berichtet werden. Die Beziehungen der Institutionen untereinander sind durch Pfeile gekennzeichnet (WEC, *whole embryo culture*; MM, *micromass*; EST, Embryonaler Stammzelltest).

3.1 Struktur der Validierungsstudie

ZEBET (Berlin, D) führt als leitendes Labor den EST durch, die weiteren Labors sind ECVAM am JRC (I-Ispra), Novartis (CH-Basel) und Schering (D-Berlin). Für den MM-Test ist die SGHMS (GB-London) das leitende Labor, die drei weiteren Labors sind das RIVM, (NL-Bilthoven), KTL-Finland (FIN-Kuopio) und Synthélabo Recherche (F-Gargenville). Der WEC-Test wird vom RIVM (NL-Bilthoven) als leitendem Labor getestet, wobei die drei weiteren Labors die SGHMS (GB-London), Novartis (CH-Basel) und ZENECA CTL (GB-Macclesfield) sind (Abb. 1).

In der initialen Phase der Validierung wurden sechs Testsubstanzen der drei *in vivo* Klassen *stark*, *schwach* und *nicht embryotoxisch* sowie eine positive Kontrollsubstanz (5-Fluorouracil, 5-FU) getestet, wobei die Dekodierung der Substanzen erst am Ende der Validierungsstudie erfolgen wird. Die Testsubstanzen wurden durch *RCC Cytotest Cell Research* willkürlich aus der Liste der 20 Testchemikalien (Tab. 1) ausgewählt. Auch die Zuordnung zu den Laboratorien wurde kodiert: die Laboratorien für den WEC-Test wurden mit den Buchstaben A bis D gekennzeichnet, jene für den MM-Test mit den Buchstaben E bis H und die Labors für den EST mit I bis L.

3.2 Entwicklung von Prädiktionsmodellen am Beispiel des EST

In einer Validierungsstudie müssen *in vitro* Tests beweisen, daß die Ergebnisse reproduzierbar sind (Balls et al.,1995). Dies bedeutet, daß andere Laboratorien in der Lage sein müssen, die Ergebnisse auch unter blinden Bedingungen in unabhängigen Experimenten zu wiederholen. Allerdings ist die Reproduzierbarkeit nur die notwendige Voraussetzung für eine weitere Anforderung: die Relevanz des *in vitro* Tests für ein bestimmtes Teilgebiet der Toxikologie. Unter Verwendung der Untersuchungsergebnisse muß mit biostatistischen Methoden ein PM errechnet werden, mit dem in der hier vorgestellten Studie eine korrekte Einteilung von Testchemikalien in *stark*, *schwach* und *nicht embryotoxisch* ermöglicht wird.

Mit Hilfe eines PM können die Ergebnisse eines *in vitro* Toxizitätstests zur Vor-

Tabelle 1: Liste der Testchemikalien der Validierungsstudie (Smith et al., 1983; Shepard, 1989; Spielmann et al., 1998)

Testsubstanz	CAS Nr.
in vivo stark embryotoxisch	
5-Bromo-2'-desoxyuridin	59-14-3
Methylquecksilber-Chlorid	115-09-3
Hydroxyharnstoff	127-07-1
Methotrexat	59-05-2
all-trans-Retinsäure	302-79-4
6-Aminonicotinamid	329-89-5
in vivo schwach embryotoxisch	
Borsäure	10043-35-3
Pentyl-4-yn-Valproinsäure	Bojic *et al.*, 1998
Valproinsäure	99-66-1
Lithium-Chlorid	7447-41-8
Dimethadion	695-53-4
Methoxyessigsäure	625-45-6
Salicylsäure-Natriumsalz	54-21-7
in vivo nicht embryotoxisch	
Acrylamid	79-06-1
Isobutylethyl-Valproinsäure	Bojic *et al.*, 1996
D-(+)-Campher	464-49-3
Dimethylphthalat	131-11-3
Diphenhydramine Hydrochlorid	147-24-0
Penicillin G Natriumsalz	69-57-8
Saccharin Natriumhydrat	82385-42-0

hersage der *in vivo* Toxizität verwendet werden.

Als mathematisches Modell für das PM wurde die Diskriminanzanalyse benutzt. Durch sie können Klassenunterschiede zwischen den drei embryotoxischen Klassen (nicht, schwach und stark embryotoxisch) untersucht werden. Dabei wird die Unterschiedlichkeit der drei Klassen hinsichtlich mehrerer Variablen untersucht, um diejenigen Variablen herauszufiltern, die am besten zur Unterscheidung der drei Klassen geeignet sind. (Backhaus et al.,1996).

Als erster Schritt zur Entwicklung von PM müssen für mehrere Testchemikalien Konzentrations-Wirkungs-Kurven erstellt werden, um typische Kurvenverläufe zu unterscheiden und potentielle Endpunkte und Variablen (s.u.) zu identifizieren. Ein typisches Beispiel sind die Konzentrations-Wirkungs-Kurven für die positive Kontrolle 5-FU im EST (Abb. 2). Sie umfassen Konzentrations-Wirkungs-Kurven für die Zytotoxizität mit embryonalen Stammzellen der Zell-Linie D3 und mit 3T3-Fibroblasten, sowie eine Kurve für den Effekt auf die Differenzierung (s. auch Material und Methoden). Aus diesen Kurven werden dann die Endpunkte, die Halbhemmkonzentrationen für die Zytotoxizität ($IC_{50}D3$, $IC_{50}3T3$) und für die Differenzierung (ID_{50}) abgelesen. Die Endpunkte (IC_{50}, ID_{50}) und Variablen (relative Distanz zwischen $IC_{50}3T3$ und ID_{50}), die für das PM verwendet wurden, sind mit Pfeilen markiert. Als biometrische Methode wurde die lineare Diskriminanzanalyse verwendet. Die drei Endpunkte des EST ($IC_{50}3T3$, $IC_{50}D3$, ID_{50}) können nicht direkt in der Diskriminanzanalyse verwendet werden und müssen durch Logarithmieren linearisiert werden. Werte, die direkt aus der Konzentrations-Wirkungs-Kurve abgelesen werden können, wie z.B. eine 50%ige Hemmung der Differenzierung (ID_{50}), werden als „Endpunkt„ bezeichnet. Sobald eine mathematische Umformung der Werte erfolgt (Logarithmus, relative Distanz), wird im folgenden der Begriff „Variable" verwendet.

Da vor Beginn der biometrischen Analyse nicht bekannt ist, welche der möglichen Variablen die höchste Trennkraft zur Unterscheidung von Stoffen mit unterschiedlichem embryotoxischen Potential besitzt, muß die Diskriminanzanalyse schrittweise erfolgen. Im einfachsten Fall reicht eine Variable für die Trennung in die drei Embryotoxizitätsklassen. Um eine möglichst vollständige Trennung zu erzielen, kann es jedoch notwendig sein, mehrere Variablen gleichzeitig in die linearen Funktionen der Diskriminanzanalyse aufzunehmen. Bei der schrittweisen Analyse wird zuerst die Variable gewählt, die allein die höchste signifikante Trennung zwischen den drei Klassen erreicht. Die Analyse wird solange fortgesetzt und neue Variablen werden integriert, bis eine verbesserte Trennung der Klassen erreicht werden kann. Diese hier beschriebene Vorgehensweise wurde bei allen drei *in vitro* Tests angewendet.

3.3 EST: Verbessertes Prädiktionsmodell (iPM)

Die Abbildung 2 zeigt die Konzentrations-Wirkungs-Kurven für die positive Kontrolle 5-FU für zwei valide Einzelexperimente. Um die oben beschriebene Vorgehensweise anzuwenden, müssen zuerst potentielle Variablen identifiziert werden. Insgesamt wurden 5 potentielle Variablen bestimmt: $lg(IC_{50}3T3)$, $lg(IC_{50}D3)$, $lg(ID_{50})$, relativer Abstand zwischen $IC_{50}3T3$ und ID_{50} und der Quotient von $IC_{50}3T3$ und ID_{50}. Pluripotente ES-Zellen sind generell gegenüber embryotoxischen oder teratogenen Substanzen empfindlicher als bereits ausdifferenzierte Zellen wie die Fibroblasten der Zell-Linie 3T3. So kann mit dem EST das zytotoxische und embryotoxische Potential einer Prüfsubstanz bestimmt werden.

Mit den Variablen $lg(IC_{50}3T3)$ und $lg(IC_{50}D3)$ werden Informationen über die Zytotoxizität auf ES-Zellen und auf 3T3-Zellen in die Funktion integriert. Darüber hinaus wurde zusätzlich der relative Abstand gebildet zwischen der $IC_{50}3T3$ (Zytotoxizität auf differenzierte 3T3 Zellen) und der ID_{50} für die Differenzierung, bezogen auf die IC_{50} 3T3. Diese Variable ist unabhängig von der absoluten Konzentration, denn diese Information ist bereits in den beiden anderen Variablen enthalten.

Es wurde eine schrittweise Diskriminanzanalyse durchgeführt, bei der drei Variablen zur optimalen Trennung der drei Klassen notwendig waren. Diese sind im Folgenden dargestellt:

Die linearen Funktionen, die das Ergebnis der Diskriminanzanalyse sind, sind in Tabelle 2 dargestellt. Die praktische Anwendung der linearen Diskriminanzfunktionen, die zur Klassifizierung von Testchemikalien in *stark*, *schwach* und *nicht embryotoxisch* führt, wird durch folgende Verfahrensweise erreicht:

Die $IC_{50}3T3$, $IC_{50}D3$ und ID_{50}-Werte werden aus der Konzentrations-Wirkungs-Kurve bestimmt, die drei Variablen berechnet und in die drei linearen Diskriminanzfunktionen eingesetzt. Übersteigt das Ergebnis der Funktion I das der Funktionen II und III, wird die Chemikalie als *nicht embryotoxisch* eingestuft. Übersteigt das Ergebnis der Funktion II das der Funktionen I und III, wird die Substanz als *schwach embryotoxisch* klassifiziert. Ist hingegen das Ergebnis der Funktion III maximal, wird die Chemikalie als *stark embryotoxisch* eingestuft.

Die Klassifizierung der Lernstichprobe (weitere Informationen sind Scholz et al., 1999, zu entnehmen), mit der das hier vorgestellte PM erarbeitet wurde, ergab zu 94% korrekte Klassifizierungen der Einzelexperimente (Tab. 3).

Die grau schattierten Felder geben jeweils die korrekte Klassifizierung in jeder Klasse an. Die Summe aller richtig klassifizierten Experimente, bezogen auf die Gesamtzahl, ergibt die richtige

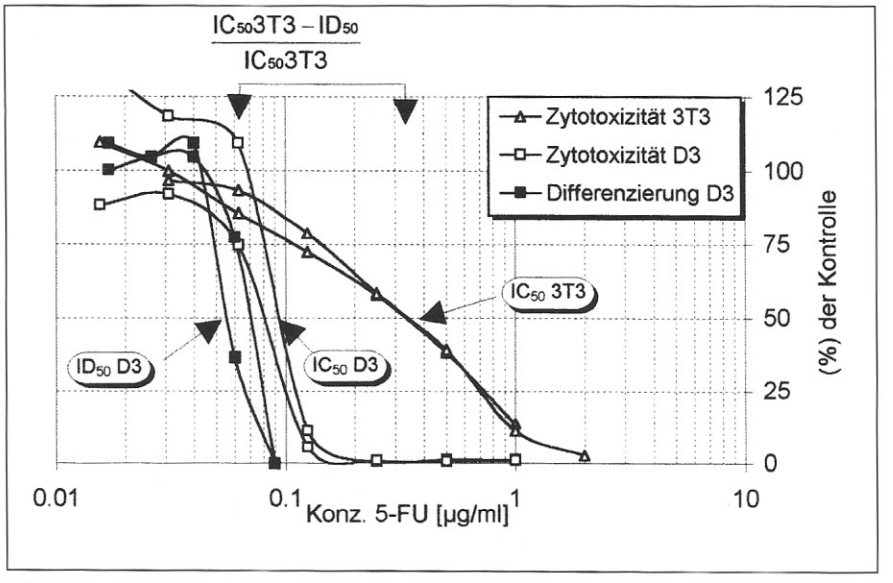

Abbildung 2: Konzentrations-Wirkungs-Kurven für den EST am Beispiel der positiven Kontrolle 5-Fluorouracil (5-FU). Gezeigt sind die Kurven für zwei valide Experimente für die drei Endpunkte Zytotoxizität der D3 und 3T3 Zellen und Hemmung der Differenzierung der D3 Zellen. Die Endpunkte und Variablen sind durch Pfeile gekennzeichnet.

Tabelle 2: Lineare Diskriminanzfunktionen I, II und III für den EST.

Funktion I	$5{,}92\ \lg(IC_{50}\,3T3) + 3{,}50\ \lg(IC_{50}D3) - 5{,}31\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 15{,}7$
Funktion II	$3{,}65\ \lg(IC_{50}3T3) + 2{,}39\ \lg(IC_{50}D3) - 2{,}03\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 6{,}85$
Funktion III	$-0{,}125\ \lg(IC_{50}3T3) - 1{,}92\ \lg(IC_{50}D3) + 1{,}50\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 2{,}67$

Tabelle 3: Klassifizierungsergebnisse für den EST (s. auch Scholz et al., 1999) mit dem verbesserten PM:
a) 94% der Einzelexperimente der Lernstichprobe und
b) 80% der Einzelexperimente in der Validierung wurden korrekt zugeordnet.

	Embryotoxizität *in vivo*	Anzahl der Chemikalien	Anzahl der Experimente	Vorhergesagte Klasse *in vitro*		
				1	2	3
a) Lernstichprobe	*nicht* embryotoxisch[1)]	3	9	9	0	0
	schwach embryotoxisch[2)]	3	9	1	8	0
	stark embryotoxisch[3)]	4	12	0	1	11
b) Evaluation mit 6 Testchemikalien	*nicht* embryotoxisch	2	8	5	3	0
	schwach embryotoxisch	2	10	1	9	0
	stark embryotoxisch	2	12	0	2	10

[1)]Klasse1; [2)]Klasse 2; [3)]Klasse 3

Klassifizierung aller Einzelexperimente.

Bei der Bewertung der korrekt klassifizierten Experimente muß außerdem der Vergleich mit den per Zufall richtigen Klassifizierungen erfolgen. Bei drei Klassen, die die gleiche Anzahl an Fällen aufweisen, läge die zufällig richtige Klassifizierung bei 33%. Darüber hinaus ist zu beachten, daß das PM auf den Daten der Lernstichprobe basiert und die Schätzung der Rate der korrekten Zuordnung des embryotoxischen Potentials überhöht ist und damit für neue Testsubstanzen niedriger ausfallen wird ist. Aus diesem Grund muß jedes neu entwickelte PM mit einem neuen Datensatz evaluiert werden.

3.3.1 Evaluierung des verbesserten PM (iPM) mit den ersten sechs Chemikalien der Validierungsstudie

Die Laboratorien I und J reichten jeweils vollständige Datensätze für die ersten sechs Testchemikalien ein, während Labor K jeweils zwei valide Experimente für drei Testchemikalien übergab. Labor L konnte für keine Testchemikalie vollständige Datensätze einreichen, da Schwierigkeiten bei der Wahl des fötalen Kälberserums für die Zellkultur auftraten. Aus diesem Grund wurden nur Daten von drei Labors evaluiert.

Insgesamt wurden 30 Einzelexperimente in die hier vorgestellte Analyse einbezogen. Das PM trennte die Testchemikalien zufriedenstellend in *stark*, *schwach* und *nicht embryotoxisch*, obwohl die Lernstichprobe lediglich 10 Testchemikalien mit 30 Einzelexperimenten umfaßte. Insgesamt konnten 80% aller Einzelexperimente korrekt klassifiziert werden (Tab. 3).

3.4 MM-Test: Entwicklung des PM

Die Laboratorien E und H übergaben jeweils vollständige, valide Datensätze. Von Labor G wurde bei drei Chemikalien jeweils eins anstelle von zwei validen Experimenten eingereicht. Beim MM-Test wird zur Qualitätskontrolle für jedes Experiment jeweils ein positives Kontrollexperiment durchgeführt. Die eingereichten Kontrollexperimente des Labors F waren nicht valide, da die Konzentrations-Wirkungs-Kurven den relevanten Bereich zwischen 100% bis deutlich unter 50% Färbung der Zellen nicht abdeckten. Deshalb konnte keines der eingereichten Experimente des Labors F in die Auswertung mit einbezogen werden. Insgesamt wurden 62 Einzelexperimente (29 Experimente aus der Prävalidierung (vgl. Scholz et al.,1998) und 33 Experimente aus der initialen Phase der Validierung) in die Diskriminanzanalyse zur Erstellung eines Prädiktionsmodelles einbezogen.

Entsprechend den Empfehlungen von ECVAM für die Prävalidierung von *in vitro* Tests in der Toxikologie (Curren et al.,1995) sollten Prädiktionsmodelle bereits im Rahmen einer Prävalidierung vorgeschlagen werden. In der Prävalidierungsphase dieser Studie wurden jedoch lediglich 3 Testchemikalien und zwei Kontrollsubstanzen untersucht. Allein eine Anzahl von 5 Testchemikalien ist als zu gering einzuschätzen, um ein PM zu berechnen. Außerdem zeigten diese Testchemikalien ein ähnliches Potential bei der Zytotoxizität auf die Zellen der Micromass-Kultur der Extremitätenknospen und bei deren Differenzierung in Knorpelzellen. Da ein PM auch für Testsubstanzen anwendbar sein soll, bei denen das zytotoxische Potential und das Potential der Wirkung auf die Differenzierung (embryotoxisches Potential) nicht identisch sind, müssen solche Substanzen auch in der Lernstichprobe enthalten sein. Um sowohl die Anzahl der Testsubstanzen in der Lernstichprobe zu erhöhen, als auch um unterschiedliche zytotoxische und embryotoxische Potentiale in das PM mit aufzunehmen, wurden für die Lernstichprobe Testchemikalien aus der Prävalidierung und aus der initialen Phase der Validierung verwendet.

Das Vorgehen bei der Erstellung eines PM wurde bereits oben für den EST beschrieben. Bei der MM-Kultur werden ebenfalls Konzentrations-Wirkungs-Kurven erstellt (s. Material und Methoden), aus denen dann die Endpunkte, die Halbhemmkonzentrationen für die Zytotoxizität (IC_{50}, ermittelt im Neutralrot-Test) und für die Differenzierung (ID_{50}, ermittelt durch knorpelspezifische Färbung mit Alcian Blau) abgelesen werden. Bei der schrittweisen Diskriminanzanalyse konnten vier verschiedene Variablen ($lg(IC_{50})$; $lg(ID_{50})$; relativer Abstand zwischen IC_{50} und ID_{50} bezogen auf IC_{50} und der Quotient von IC_{50} und ID_{50}) mit einbezogen werden. Die Endpunkte sind in Abbildung 3 dargestellt.

Aus den Konzentrations-Wirkungs-Kurven war abzulesen, daß zwischen dem zytotoxischen Effekt und dem Effekt auf die Differenzierung kein systematischer Unterschied besteht. Dennoch wurde die schrittweise Diskriminanzanalyse durchgeführt. Tatsächlich wurde dabei nur eine Variable ($lg(ID_{50})$) zur Trennung der drei Embryotoxizitätsklassen identifiziert. Die linearen Diskriminanzfunktionen für den MM-Test sind in Tabelle 4 dargestellt. Die praktische Anwendung der Diskriminanzfunktionen zur Klassifizierung wurde bereits oben beschrieben. Das Klassifizierungsergebnis der Lernstichprobe durch das PM lag bei 79% korrekter Einstufungen (Tab. 5). Aber auch hier gilt, daß die Klassifizierung der Lernstichprobe zu einer überhöhten Schätzung führt, da PM generell besser auf die Lernstichprobe passen als auf neue Testchemikalien. Aus diesem Grund muß das PM mit den verbleibenden 14 Testchemikalien der formalen Validierung geprüft bzw. bestätigt werden, um eine realistischere Schätzung der Güte des PM zu erhalten.

Tabelle 4: Lineare Diskriminanzfunktionen I, II und III für den MM-Test.

Funktion I	$6{,}65 * lg(ID_{50}) - 9{,}49$
Funktion II	$6{,}16 * lg(ID_{50}) - 8{,}29$
Funktion III	$-1{,}31 * lg(ID_{50}) - 1{,}42$

3.5 WEC-Test: Entwicklung des PM

Der WEC-Test ist im Vergleich zu den beiden anderen *in vitro* Embryotoxizitätstests sehr zeitaufwendig, da eine große Zahl von Parametern pro Rattenembryo erfaßt wird; infolgedessen wird für jede Testsubstanz nur ein valides Experiment ausgewertet. Die Laboratorien A und C übersandten Ergebnisse aller sechs Testsubstanzen. Labor B reichte nur die Ergebnisse von fünf Testsubstanzen und Labor D von zwei Testsubstanzen ein. Insgesamt wurden 24 Experimente einschließlich der positiven und negativen Kontrollen ausgewertet. Da für den WEC-Test aus der Prävalidierung nur eine sehr begrenze Anzahl von Versuchen von den zwei teilnehmenden Laboratorien vorlag, mußte eine Verbesserung des PM in der initialen Phase der Validierung erfolgen. Der Typ und die Häufigkeit des Auftretens individueller Mißbildungen (die in jedem Fall genau dokumentiert werden) konnten aufgrund der derzeitig vorhandenen, begrenzten Datenmenge bisher nicht ausgewertet werden.

Es wurden zwei Prädiktionsmodelle, PM1 und PM2, mit der schrittweisen linearen Diskriminanzanalyse entwickelt. PM1 beinhaltet zwei Variablen, deren Werte sich wiederum aus den ermittelten Konzentrations-Wirkungs-Kurven ableiten (s. auch Material und Methoden): (i) der Logarithmus derjenigen Konzentration einer Testsubstanz, bei der 50% der getesteten Embryonen mißgebildet sind ($lg(IC_{50})$ für Mißbildung; Abb. 4) und (ii) der Logarithmus der NOEC des *total morphological score* ($IC_{NOEC}TMS$, Abb. 5). Die NOEC (*no effect concentration*) ist dabei definiert als die höchste von der Kurve abgelesene Konzentration einer Testsubstanz, bei der gerade noch ein TMS von 100% der unbehandelten Kontrollen erreicht wird.

In das zweite Prädiktionsmodell (PM2) wurde zusätzlich als Parameter der Zytotoxizität der Logarithmus des $IC_{50}3T3$ Wertes aus dem EST mit in die Diskriminanzanalyse des WEC-Tests integriert.

Die Variablen des PM2 sind: (i) der Logarithmus der niedrigsten Konzentration, bei der die maximale Mißbildungs-

Tabelle 5: Klassifizierungsergebnisse der Lernstichprobe für den MM-Test (62 Einzelexperimente: 29 Experimente aus der Prävalidierung und 33 Experimente mit den ersten sechs Testchemikalien der Validierung): 79% der Einzelexperimente wurden korrekt zugeordnet.

	Embryotoxizität *in vivo*	Anzahl der Chemikalien	Anzahl der Experimente	Vorhergesagte Klasse *in vitro*		
				1	2	3
Lernstichprobe	*nicht* embryotoxisch[1)]	4	24	17	7	0
	schwach embryotoxisch[2)]	3	17	6	11	0
	stark embryotoxisch[3)]	5	21	0	0	21

[1)]Klasse1; [2)]Klasse 2; [3)]Klasse 3

Tabelle 6: Lineare Diskriminanzfunktionen I, II und III des PM1 für den WEC-Test.

Funktion I:	$14{,}06 * \lg(IC_{50}\,Mal) - 5{,}43 * \lg(IC_{NOEC}\,TMS) - 13{,}14$
Funktion II:	$17{,}94 * \lg(IC_{50}\,Mal) - 9{,}37 * \lg(IC_{NOEC}\,TMS) - 15{,}44$
Funktion III:	$3{,}04 * \lg(IC_{50}\,Mal) - 2{,}14 * \lg(IC_{NOEC}\,TMS) - 1{,}56$

Tabelle 7: Klassifizierungsergebnisse der Lernstichprobe für den WEC-Test mit den ersten sechs Testchemikalien der Validierung:
a) mit PM1 wurden 79% und
b) mit PM2 wurden 96% der Einzelexperimente korrekt zugeordnet.

	Embryotoxizität *in vivo*	Anzahl der Chemikalien	Anzahl der Experimente	Vorhergesagte Klasse *in vitro*		
				1	2	3
a) PM1	*nicht* embryotoxisch[1)]	2	9	8	1	0
Lernstichprobe	*schwach* embryotoxisch[2)]	2	7	4	3	0
	stark embryotoxisch[3)]	2	8	0	0	8
b) PM2	*nicht* embryotoxisch	2	9	8	1	0
Lernstichprobe	*schwach* embryotoxisch	2	7	0	7	0
	stark embryotoxisch	2	8	0	0	8

[1)]Klasse1; [2)]Klasse 2; [3)]Klasse 3

rate der Embryonen erreicht ist ($\lg(IC_{max})$ für Mißbildung, Abb. 4) und (ii) der relative Abstand zwischen der IC_{50}3T3 und der IC_{NOEC}TMS (Abb. 5) bezogen auf die IC_{50}3T3 (Abb. 2).

3.5.1 Prädiktionsmodell 1 (PM1)

Insgesamt konnten beim WEC-Test sieben unterschiedliche Endpunkte (IC_{NOEC} der Mißbildung, Maximum der Mißbildung (%), IC_{50} der Mißbildung, IC_{NOEC}TMS, IC_{max} der Mißbildung, IC_{50}TMS und IC_{75}TMS) auf ihre prädiktive Wirkung untersucht werden. Wie oben beschrieben, wurde eine schrittweise Diskriminanzanalyse durchgeführt. Zwei Variablen wurden in das Prädiktionsmodell aufgenommen. Tabelle 6 zeigt die Diskriminanzfunktionen für das PM1. Die Lernstichprobe basiert auf Daten der ersten sechs Testsubstanzen der formalen Validierung und der positiven und negativen Kontrolle. Die Anwendung der Diskriminanzfunktionen, mit denen berechnet wird, ob ein Stoff als *nicht*, *schwach* oder *stark embryotoxisch* eingestuft wird, wurde bereits für den EST ausführlich beschrieben.

3.5.2 Klassifizierung der Lernstichprobe (PM1)

Mit PM1 wurden für 79% der Experimente das richtige embryotoxische Potential vorhergesagt (Tab. 7). Für die stark embryotoxischen Testchemikalien wurde wie im MM-Test zu 100% eine korrekte Klassifizierung erreicht. Auch hier gilt wie für den MM-Test, daß in der Lernstichprobe die korrekte Klassifizierung überhöht ist und erst die Evaluierung mit neuen Substanzen eine realistische Schätzung der Güte des PM1 erlaubt.

3.5.3 Prädiktionsmodell 2 (PM2)

Im Gegensatz zum EST, in dem sowohl die zytotoxischen Eigenschaften einer Substanz auf differenzierte Zellen (zytotoxisches Potential) als auch die Wirkung auf die Differenzierung von ES-Zellen in kontrahierende Herzmuskelzellen berücksichtigt werden, wird im WEC-Test kein zytotoxischer Endpunkt bestimmt. Da offensichtlich eine Stärke des EST in dieser zusätzlichen Information liegt, wurde ein zweites PM (PM2) entwickelt, bei dem neben den Variablen aus dem WEC-Test zusätzlich die IC_{50}3T3 des EST in die schrittweise Diskriminanzanalyse integriert wurde. Allerdings wurden die Werte nicht direkt als Variable verwendet, sondern wie im EST wurde ein relativer Abstand gebildet zwischen der IC_{50}3T3 und der NOEC des TMS (IC_{NOEC}TMS) des WEC-Tests bezogen auf die IC_{50}3T3.

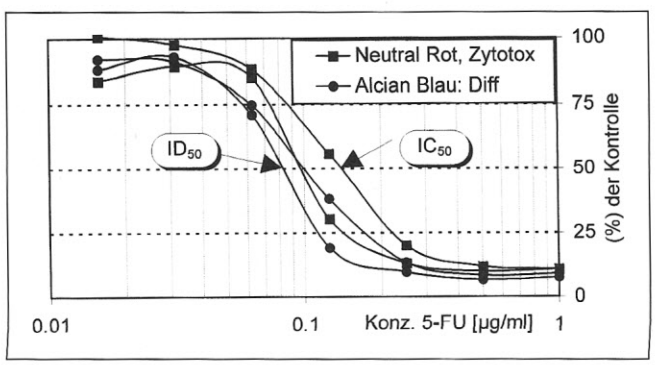

Abbildung 3: Konzentrations-Wirkungs-Kurven für den MM-Test am Beispiel der positiven Kontrolle 5-Fluorouracil (5-FU). Gezeigt sind die Kurven für die Zytotoxizität (Färbung mit Neutral Rot; IC_{50}) und Hemmung der Differenzierung (Färbung mit Alcian Blau; ID_{50}). Die Endpunkte sind durch Pfeile gekennzeichnet.

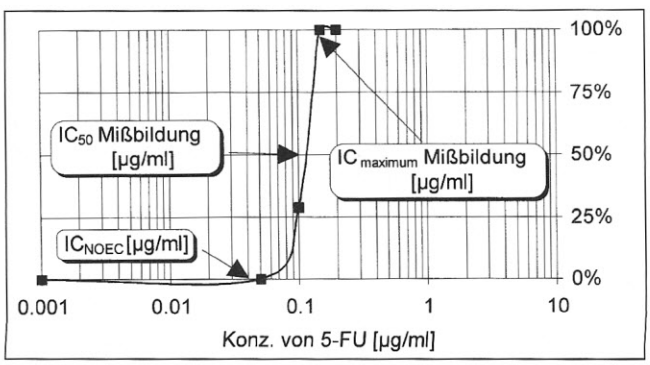

Abbildung 4: Konzentrations-Wirkungs-Kurve für den WEC-Test am Beispiel der positiven Kontrolle 5-Fluorouracil (5-FU). Die Abbildung zeigt die Kurve für den prozentualen Anteil aller Mißbildungen der getesteten Embryonen. Die Endpunkte für PM1 und PM2 sind mit Pfeilen markiert.

Beim relativen Abstand werden besonders die unterschiedlichen zytotoxischen Effekte und Effekte auf die Differenzierung von Prüfsubstanzen berücksichtigt, ohne daß die absolute Konzentration eine Rolle spielt. Die Konzentration, bei der eine Wirkung auf die Mißbildung festgestellt wird, ist bereits in der zweiten Variablen enthalten: der niedrigsten Konzentration, bei der die maximale Mißbildung der Embryonen erreicht ist ($lg(IC_{max})$ der Mißbildung). In Tabelle 8 sind die Diskriminanzfunktionen für das PM2 beschrieben, die Klassifizierung erfolgt.

Tabelle 8: Lineare Diskriminanzfunktionen I, II und III des PM2 für den WEC-Test.

Funktion I:	$0{,}14 * \frac{IC_{50}3T3 - IC_{NOEC}TMS}{IC_{50}3T3} \cdot 100 + 11{,}89 * lg(IC_{max})^{1)} - 17{,}42$
Funktion II:	$0{,}21 * \frac{IC_{50}3T3 - IC_{NOEC}TMS}{IC_{50}3T3} \cdot 100 + 14{,}57 * lg(IC_{max}) - 27{,}49$
Funktion III:	$0{,}074 * \frac{IC_{50}3T3 - IC_{NOEC}TMS}{IC_{50}3T3} \cdot 100 + 3{,}38 * lg(IC_{max}) - 3{,}53$

3.5.4 Klassifizierung der Lernstichprobe (PM2)

Tatsächlich konnte eine deutlich verbesserte Klassifizierung mit dem PM2 erreicht werden; sie lag bei 96% der Experimente (Tab. 7). Lediglich in einem einzigen Experiment wurde falsch klassifiziert, in dem die Testsubstanz *schwach* statt *nicht embryotoxisch* eingestuft wurde. Dieses sehr vielversprechende Ergebnis des PM2 muß entsprechend den Klassifizierungen der Lernstichproben des PM1 und des PM des MM-Tests mit den übrigen 14 Testchemikalien der Validierungsstudie evaluiert werden, um die tatsächliche Güte der prädiktiven Trennschärfe zu ermitteln.

4 Diskussion und Ausblick

Nachdem wir vor einem Jahr die Ergebnisse der Prävalidierungsphase der derzeit laufenden ECVAM Validierungsstudie von drei *in vitro* Embryotoxizitätstests vorgestellt haben (Scholz et al.,1998), berichten wir in der vorliegenden Studie über erste Ergebnisse der formalen Validierungsstudie unter blinden Bedingungen, und zwar für die ersten sechs von 20 Prüfsubstanzen. Die formale Validierungsphase der ECVAM-Studie wird unter den von ECVAM und der OECD definierten Bedingungen für die Validierung von toxikologischen Prüfmethoden, die für regulatorische Zwecke eingesetzt werden sollen, durchgeführt (Balls et al.,1995; OECD, 1996). Wie einleitend dargestellt, wurden bisher *in vitro* Embryotoxizitätstests noch nicht unter vergleichbar strikten Bedingungen experimentell validiert, da in vorangegangenen Studien auf wesentliche Elemente der Prävalidierung verzichtet wurde, die heute Grundelemente jeder Validierungsstudie sind, und ohne die eine formale Validierungsstudie nicht begonnen werden sollte. Dazu gehören z.B. die Erarbeitung eines für alle Teilnehmer der Validierungsstudie verbindlichen Standardprotokolls (SOP) und eines biometrisch abgesicherten Prädiktionsmodells (PM). Die hier vorgestellten vorläufigen Ergebnisse sollen zeigen, daß insbesondere die Entwicklung von PM zu einer Verbesserung der wissenschaftlichen Parameter der *in vitro* Tests führt, da nur streng definierte und reproduzierbar bestimmbare Parameter bzw. toxikologische Endpunkte in einem PM Berücksichtigung finden (Curren et al.,1995; OECD, 1996).

Aufgrund dieser im Vertrag festgelegten Ziele der Studie wurde nach Standardisierung der drei *in vitro* Embryotoxizitätstests in der im Frühjahr 1998 abgeschlossenen Prävalidierungsphase (Scholz et al.,1998) zu Beginn der formalen Validierungsstudie vor allem Wert auf das Erarbeiten von PM auch für den MM- und den WEC-Test gelegt, da die in der Prävalidierungsphase bestimmten Endpunkte dieser beiden lange etablierten Tests nicht ausreichend reproduzierbar waren. Für den EST wurde das PM bereits in einer vorangehenden Prävalidierungsstudie entwickelt und in diesem Rahmen einer ersten Prüfung unterzogen (Scholz et al.,1999). Im Gegensatz dazu sind der MM- und der WEC-Test zwar bereits lange wissenschaftlich etabliert, ohne daß jedoch bisher PM für diese beiden Tests erarbeitet wurden.

Mit Hilfe der ersten Zwischenergebnisse der laufenden Validierungsstudie konnten mit biostatistischen Methoden erstmals PM für den WEC- und den MM-Test definiert, sowie das PM für den EST einer experimentellen Evaluierung unterzogen werden. Es muß jedoch betont werden, daß die Klassifizierung der ersten sechs Prüfsubstanzen mit den PM für den MM- und den WEC-Test auf einer Lernstichprobe beruht, die daher wahrscheinlich in einer zu guten Einschätzung der Zuordnung zu den drei Embryotoxizitätsklassen *nicht, schwach und stark embryo-*

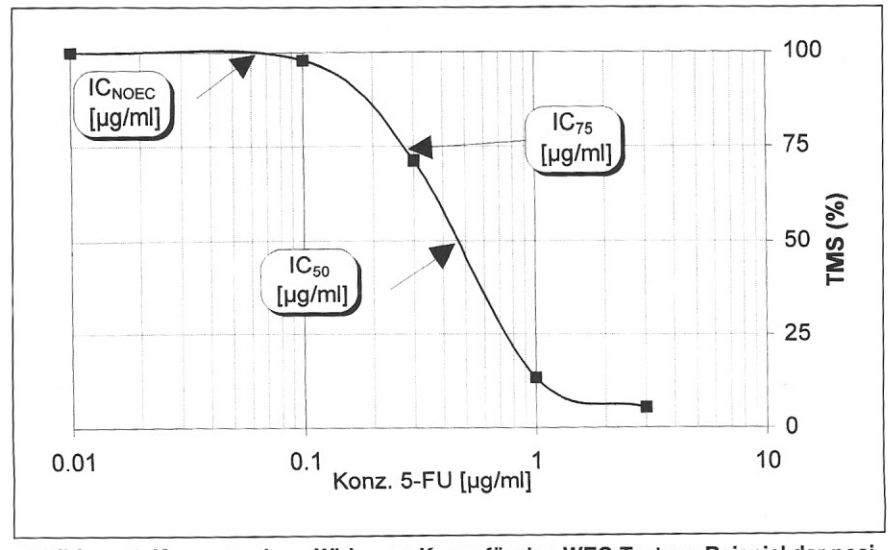

Abbildung 5: Konzentrations-Wirkungs-Kurve für den WEC-Test am Beispiel der positiven Kontrolle 5-Fluorouracil (5-FU). Dargestellt ist eine Kurve für den *total morphological score* (TMS), Endpunkte für PM1 und PM2 sind mit Pfeilen markiert.

toxisch resultiert. Bei Verwendung neuer Prüfsubstanzen ist eine erheblich schlechtere Einschätzung zu erwarten. Dies zeigte sich bereits bei der ersten Evaluierung des PM für den EST. Hierbei wurde bei der Entwicklung des PM mit der Lernstichprobe eine korrekte Klassifizierung von 94% erreicht, die sich bei der Evaluierung mit den ersten sechs Substanzen der Validierung auf 80% verminderte.

Besonders hervorzuheben ist beim WEC-Test, daß eine große Anzahl entwicklungsbiologischer Parameter (u.a. 16 Parameter der generellen, morphologischen Entwicklung und 30 Typen von organspezifischen Mißbildungen) erhoben wird. Dabei waren bisher schon die Summe der generellen, morphologischen Entwicklungsparameter (*total morphological score*, TMS) und der prozentuale Anteil der mißgebildeten Embryonen, unabhängig von der Art der Mißbildungen, ausreichend, um das embryotoxische Potential der getesteten Chemikalien vorherzusagen. Im weiteren Verlauf der Validierungsstudie sollen jedoch auch die Häufigkeit und das Auftreten spezifischer Mißbildungen ausgewertet werden.

Beim MM-Test hatte sich bereits in der Prävalidierung herausgestellt, daß die Testung von Prüfsubstanzen an Micromass-Kulturen des embryonalen Mittelhirns nur schwer reproduzierbar war und keine verwertbaren Informationen lieferte, die die Ergebnisse der Micromass Kultur der Extremitätenknospen ergänzen konnte. Aus diesem Grund wurde in der Validierungsphase im MM-Test nur mit Zellen aus Extremitätenknospen gearbeitet. Die Ergebnisse der vorliegenden Studie bestätigen außerdem die Beobachtung aus der Prävalidierungsphase, daß die Hemmung der Differenzierung (Alcian Blau Färbung) und die Beeinflussung der Vitalität (Zytotoxizität, Färbung mit Neutralrot) kaum Unterschiede zeigten. Im ursprünglichen MM-Test beruhte dagegen die Prädiktion der embryotoxischen Eigenschaften von Prüfsubstanzen auf einer unterschiedlichen Beeinträchtigung der Differenzierung im Vergleich zur Vitalität (Flint und Orton, 1984), die sich als nicht haltbar herausstellte. Aus den genannten Gründen wird in dem hier vorgestellten, vorläufigen PM für den MM-Test nur die Beeinträchtigung der Differenzierung berücksichtigt, die in einer etwas besseren Prädiktion resultierte als die Zytotoxizität. Es werden im MM-Test bei der weiteren Prüfung der verbleibenden 14 Prüfsubstanzen jedoch nach wie vor die Hemmung von Differenzierung und Vitalität der Zellen aus embryonalen Extremitätenknospen geprüft, um das vorläufige PM nach Abschluß der Validierungsstudie unter Berücksichtigung der beiden wichtigsten Parameter in der Embryotoxizität optimieren zu können.

Bei der Bewertung des embryotoxischen Potentials der ersten sechs Testchemikalien mit Hilfe der PM erwies sich in einigen Fällen die Unterscheidung zwischen *nicht* und *schwach embryotoxischen* Substanzen als schwierig. Diese Beobachtung kann jedoch bei der begrenzten Zahl von Prüfsubstanzen nicht generalisiert werden, da dieses Ergebnis auch durch die speziellen Eigenschaften der sechs getesteten Substanzen bedingt sein kann. Eine abschließende, umfassende Diskussion der drei *in vitro* Embryotoxizitätstests und die Bewertung der Güte und Trennkraft der PM ist daher erst nach Abschluß der Studie im Januar 2000 möglich. Die nächsten Zwischenergebnisse der Testung von weiteren sechs Substanzen wird im Mai 1999 erwartet. Erst dann kann die weitere Evaluierung der PM erfolgen.

Literatur

Amacher, D. E., Stadler, J., Schomaker, S. J., Verseil, C. (1996). The comparative testing of eight coded chemicals in the rat limb bud micromass and rat embryo culture systems. *ATLA 24,* 945-952.

Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (1996). *Multivariate Analysemethoden.* 8. Auflage, 591 Seiten. Berlin, Heidelberg, New York, London: Springer Verlag.

Balls, M., Blaauboer, B., Brusik, D., Frazier, J., Lamp, D., Pemberton, M., Reinhardt, C., Robertfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stammati, A.L. and Walum,E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA 18,* 313-337.

Balls, M., Blaauboer, B. J., Fentem, J. H., Bruner, L., Combes, R. D., Ekwall, B., Fielder, R. J., Guillouzo, A., Lewis, R. W., Lovell, D. P., Reinhardt, C. A., Repetto, G., Sladowski, D., Spielmann, H. und Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations ECVAM workshop 5. *ATLA 23,* 129-147.

Bechter, R., and Schön, H. (1988). Use of the whole embryo culture system in drug safety assessment. *Toxic. in vitro 2,* 195-203.

Bojic U., Ehlers K., Ellerbeck U., Bacon C.L., O'Driscoll E., O'Connell C., Berezin V., Kawa A., Lepekhin E., Bock E., Regan C.M. and Nau H. (1998). Studies on the teratogen pharmacophore of valproic acid analogues: evidence of interactions at a hydrophobic centre. *Eur. J. Pharmacol.* 354, 289-299.

Bojic U., Elmazar M.M.A., Hauck R.-S. and Nau H. (1996). Further branching of valproate-related carboxylic acids reduces the teratogenic activity, but not the anticonvulsant effect. *Chem. Res. Toxicol.* 9, 866-870.

Bortz, J. (1993). *Statistik.* 3. Auflage, 753 Seiten. Berlin, Heidelberg, New York, London: Springer Verlag.

Brown, N. A. and Fabro, S. (1981). Quantitation of rat embryonic development in vitro: a morphological scoring system. *Teratology 24,* 65-78.

Brown, N. A., Spielmann, H., Bechter, R., Flint, O. P., Freeman, S. J., Jelinek, R. J., Koch, E., Nau, H., Newall, D. R., Palmer, A. K., Renault, J. Y., Repetto, M., Vogel, R. und Wigler, R. (1995). Screening chemicals for reproductive toxicity: the current alternatives. The report and recommendations of an ECVAM/ETS workshop (ECVAM workshop 12). *ATLA 23,* 868-882.

Curren, R. D., Southee J. A., Spielmann, H., Liebsch, M., Fentem, J. H. und Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods (the first report of the ECVAM Task Force on Prevalidation). *ATLA 23,* 211-217.

Flint, O. P. (1993). *In vitro* tests for teratogens: desirable endpoints, test batteries and current status of the micromass teratogen test. *Reprod.Toxicol. 7,* 103-111.

Flint, O. P., Orton, T. C. (1984). An *in vitro* assay for teratogens with cultures of rat embryo mid-brain and limb cells. *Toxicol. Appl. Pharmacol. 76,* 383-395.

Kistler, A. (1987). Limb bud cell cultures for estimating the teratogenic potential of compounds: Validation of the test system with retinoids. *Archiv. Toxicol.*; *60,* 403-414.

Kistler A, Tsuchiya T, Tsuchiya M, Klaus M (1990). Teratogenicity of arotinoids (retinoids) in vivo and in vitro. *Arch Toxicol* 64, 616-22.

Kucera, P.; Cano, E.; Honegger, P.; Schilter, B.; Zijlstra, J.A. and Schmid, B. (1993). Validation of whole chick embryo cultures, whole rat embryo culture and aggregating embryonic brain cell cultures using six pairs of coded compounds. *Toxicology in Vitro 7, 785-798.*

Norušis, M. J. (1994). *SPSS Professional Statistics*™ 6.1. Chicago, Illinois: SPSS Inc.

OECD (1996). Final reprt of the OECD Workshop on Harmoniztion of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. OECD Publications Office, Paris, France, 60pp.

Piersma, A. H., Attenon, P., Bechter, R., Govers, M. J. A. P., Krafft, N., Schmid, B. P., Stadler, J., Verhoef, A., Verseil, C. (1995). Interlaboratory evaluation of embryotoxicity in the postimplantation rat embryo culture. *Reprod. Toxic. 9,* 275-280.

Piersma, A. H., Bechter, R., Krafft, N., Schmid, B. P., Stadler, J., Verhoef, A., Verseil, C. and Zijlstra, J.A. (1996). An interlaboratory evaluation of five pairs of teratogens and teratogens in postimplantation rat embryo culture. *ATLA* 24, 201-209.

Scholz G., Pohl I., Seiler A., Bremer S., Brown N.A., Piersma A.H., Holzhütter H.G. und Spielmann H. (1998). Ergebnisse der ersten Phase des ECVAM Projektes zur Prävalidierung und Validierung von drei *in vitro* Embryotoxizitätstests. *ALTEX* 15, 3-8.

Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Moldenhauer, F., Southee, J. and Spielmann, H. (1999). Prevalidation of the Embryonic Stem Cell Test (EST), a new in vitro Embryotoxicity Test. *Toxicology in Vitro (in press).*

Shepard T.H. (1992). *Catalog of Teratogenic Agents*, 7th edition. Baltimore, London: The Johns Hopkins University Press.

Smith, M.K., Kimmel, G.L., Kochhar, D.M., Shepard, T.H., Spielberg, S.P. and Wilson, J.G. (1983). A selection of candidate compounds for in vitro teratogenesis test validation. *Teratogenesis, Carcinogenesis, und Mutagenesis* 3, 461-480.

Spielmann H., Steinhoff H., Schaefer C. and Bunjes R. (1998). *Arzneiverordnung in Schwangerschaft und Stillzeit*, 5. Aufl.. Stuttgart, Jena, Ulm, Lübeck: Gustav Fischer Verlag.

Spielmann, H. (1998). Reproduction and development. *Environmental Health Perspectives* 106, Suppl. 2, 571-576.

Spielmann, H., Pohl, I., Döring, B., Liebsch, M., and Moldenhauer, F. (1997). The embryonic stem cell test (EST), an in vitro embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicol.,10*, 119-127.

Uphill, P. F., Wilkins, S., Allen, J. A. (1990). *In vitro* micromass teratogen test: results from a blind trial of 25 compounds. *Toxicol. In Vitro 4*, 623-626.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. und Gough, N. M. (1988). Myeloid leucemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature 336,* 684-68

Danksagung

Die Autoren danken Heinz Nau (Tierärztlich Hochschule Hannover) und Heinrich Bürgin (Hoffmann La Roche, Basel, CH) für die freundliche Bereitstellung von Testchemikalien sowie der ECETOC (*European Chemical Industry Ecology and Toxicology Center*) *Task Force on Reproductive Toxicology* für die kritische Prüfung der Chemikalienliste. Die Studie wurde durch einen Vertrag mit der EU/ECVAM (Vertrag Nr. 12225-96-10-F1ED ISP D) finanziert.

Korrespondenzadresse:
Prof. Dr. Horst Spielmann
ZEBET/BgVV
Diedersdorfer Weg 1
D-12277 Berlin
Tel. +49-30-8412 2270
Fax +49-30-8412 2958
E-mail zebet@bgvv.de

Abkürzungen: EST=Embryonaler Stammzelltest, ES-Zellen=Embryonale Stammzellen, HTS=*high throughput screening,* MM-Test=*Micromass-Test,* WEC-Test=*Whole Embryo Culture-Test,* LIF=*Leukemia Inhibitory Factor,* SOP= *Standard Operation Protocol,* PM=Prädiktionsmodell