2.0 *IN VITRO* SCREENING METHODS FOR ASSESSING ACUTE TOXICITY

2.1 Introduction

Since the early work of Pomerat and Leake (1954), Eagle and Foley (1956), and Smith and colleagues (1963), research over the last 50 years has been conducted to evaluate the potential use of *in vitro* cell systems for predicting acute toxic effects *in vivo*. Significant correlations between cytotoxicity *in vitro* and animal lethality have been demonstrated on numerous occasions (for reviews see Phillips et al., 1990; Garle et al., 1994), as have correlations between cytotoxicity *in vitro* and systemic and topical effects from acute exposures to chemicals. Several newer initiatives directed toward reducing and replacing the use of laboratory animals for acute toxicity testing have emerged (Curren et al., 1998; Ohno et al., 1998; Spielmann et al., 1999; Ekwall et al., 2000); these initiatives were reviewed as part of the charge given to Breakout Group 1 (*In Vitro* Screening Methods) at this Workshop.

2.1.1 *Charge to the Breakout Group*

Breakout Group 1 (BG1) was asked to evaluate the validation status of available *in vitro* methods for estimating *in vivo* acute toxicity and was requested to identify methods and appropriate validation studies that might be completed within the next one to two years. It was also envisaged that the Breakout Group would evaluate potential uses of QSAR as part of an *in vitro* strategy.

2.1.2 *Objectives*

The specific objectives of the Workshop pertinent to the charge given to BG1 were given as follows:

(1) Review the validation status of available *in vitro* screening methods for their usefulness in estimating *in vivo* acute toxicity.
(2) Recommend candidate methods for future evaluation in prevalidation and validation studies.
(3) Recommend validation study designs that can be used to adequately characterize the usefulness and limitations of proposed *in vitro* methods.
(4) Identify priority research efforts necessary to support the development of mechanism-based *in vitro* methods to assess acute systemic toxicity.

In its opening deliberation on these objectives, BG1 members decided to limit the review to methods for reducing or replacing animal use for determining acute lethality with the understanding that Breakout Group 3 would focus on methods for assessing acute systemic toxicity.

2.2 Background

Cytotoxicity has been defined as the adverse effects resulting from interference with structures and/or processes essential for cell survival, proliferation, and/or function (Ekwall, 1983). These effects may involve the integrity of membranes and the cytoskeleton, cellular metabolism, the synthesis and degradation or release of cellular constituents or products, ion regulation, and cell division. Ekwall (1983) described the concept of "basal cell functions" that virtually all cells possess (mitochondria, plasma membrane integrity, etc.) and suggested that, for most chemicals, toxicity is a consequence of non-specific alterations in those cellular functions which may then lead to effects on organ-specific functions and/or death of the organism.

Ekwall drew two important inferences from his early studies: that (a) cell cultures (notably cell lines) can be used to detect basal cytotoxicity; and (b) many chemicals exert cytotoxic effects on these cultures at concentrations which would be lethal in humans. Ekwall recognized that there will be exceptions and ultimately refinements needed in the development of a test battery for predicting human lethality, as, for example, incorporating test strategies for identifying chemicals that produce cell selective (organ specific) toxicity at lower concentrations than “basal” (or general) cytotoxicity.

Others likewise concluded that, since the actions of chemicals that produce injury and death are ultimately exerted at the cellular level, cytotoxicity assays may be useful for the prediction of acute lethal potency (Grisham and Smith, 1984). Based on that premise, a considerable amount of research has been undertaken into the development and evaluation of *in vitro* tests for use as screens and as potential replacements for *in vivo* LD50 tests. Good agreement between cytotoxicity *in vitro* and animal lethality have been reported by numerous groups (see reviews by Phillips et al., 1990; Garle et al., 1994; Guzzie, 1994). However, none of the proposed *in vitro* models have been evaluated in any formal studies for reliability and relevance, and their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing have not been assessed.

More recently, Spielmann and colleagues have conducted studies to indicate that, as a first step toward replacement of LD50 tests, *in vitro* cytotoxicity data could be used now to identify the appropriate starting dose for *in vivo* studies, thereby reducing the number of animals necessary for such determinations (Spielmann et al., 1999). Other studies have indicated an association between chemical concentrations inducing cytotoxic effects *in vitro* and human lethal blood concentrations (Ekwall et al., 2000). Several groups have proposed the use of *in vitro* cytotoxicity tests in tiered testing schemes. These tests include proposed strategies for using *in vitro* test data as a basis for classifying and labeling new chemicals, thereby reducing (and possibly replacing) the need for acute toxicity tests in animals (Seibert et al., 1996) and for *in vitro* cytotoxicity data and other information in a tiered approach to replace oral LD50 tests (Curren et al., 1998). Curren and colleagues recognized that the application of their proposal was limited because of insufficient information on the many cellular mechanisms involved in chemical-induced lethality and because the most reliable *in vitro* models for gastrointestinal uptake, blood-brain barrier (BBB) passage, and biotransformation for more precise quantitative *in vivo* toxic dose/exposures were not yet identified.

To summarize, many investigations of the relationship between *in vitro* cytotoxicity and acute toxicity *in vivo* have been reported. Since it was not possible to critically review and discuss all of the published literature in the course of the Workshop, a selection of recent key activities and reports that included the most advanced and extensive efforts to develop alternative methods for lethality was made for consideration by Breakout Group 1 (Appendix D). The most intensive discussions focused on the ZEBET and MEIC approaches, which are outlined below in detail for the reader's reference (Sections 2.2.1-2.2.6 and 2.2.7, respectively).

2.2.1 *Prediction of In Vivo Starting Doses (ZEBET Approach)*

Investigators (Halle et al., 1997; Halle 1998; Spielmann et al., 1999) have proposed a strategy to reduce the number of animals required for acute oral toxicity testing. The strategy is referred to in this document as the ZEBET approach where ZEBET is the acronym for Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergaenzungsmethoden zum Tierversuch (the National Center for Documentation and Evaluation of Alternative Methods to Animal Experiments). The strategy involves using *in vitro* cytotoxicity data to determine the starting dose for *in vivo* testing. They report the findings of an initial study conducted to assess the feasibility of applying the standard regression between mean IC50 values (i.e., IC50*x*, the mean concentration estimated to affect the endpoint in question by 50%) and acute oral LD50 data included in the Register of Cytotoxicity (RC) to estimate the LD50 value which can then be used to determine the *in vivo* starting dose.

The RC is a database of acute oral LD50 data from rats and mice (taken from the NIOSH Registry of Toxic Effects of Chemical Substances [RTECS]) and IC50*x* values of chemicals and drugs from *in vitro* cytotoxicity assays (Halle and Goeres, 1988; Halle and Spielmann, 1992). It currently contains data on 347 chemicals (Halle, 1998; Spielmann et al., 1999). The main purpose of establishing the RC was to evaluate, with a large amount of non-selected data from various chemicals with different systemic oral toxicities, whether basal

cytotoxicity (averaged over various cells, cell lines, and/or toxicity endpoints) is a sufficient predictor for acute systemic toxicity.

Apart from the fact that basal cytotoxicity was an acceptable predictor (i.e., LD50 values localized in the dose range around the regression line by the empirical factor $F_G \leq \log 5$) of the LD50 for 74% of the RC chemicals (Halle and Spielmann, 1992), the predicted LD50 value can be used as a starting dose in acute oral toxicity testing to reduce the number of animals. This concept was first discussed at an ECVAM workshop (Seibert et al., 1996) as it related to refinements of *in vivo* acute toxicity tests by the use of new sequential dosing methods such as the Acute Toxic Class method ([ATC; OECD TG 423] OECD, 1996) and the Up-and-Down Procedure ([UDP; OECD TG 425] OECD, 1998b). In these tests, the number of animals needed depends upon the correct choice of the starting dose, since the number of consecutive dosing steps would be reduced as the starting dose more closely approximates the true toxicity class (ATC), or the true LD50 (UDP) (i.e., the more precisely the starting dose is predicted, the fewer animals that need to be used).

2.2.2 *Characterization of the RC*

The first registry, RC-I (Halle and Göeres, 1988), contained 117 chemicals and served as a training data set to establish a linear regression model for predicting oral LD50 values. A second data set of 230 chemicals, RC-II, verified the regression obtained with RC-I (Halle, 1998). Currently, a third RC of 150 chemicals that will increase the number of chemicals to almost 500 is in preparation. It is important to note that, in order to keep the registry unbiased, published data that were complete and met the acceptance criteria described below were included in the RC without further restriction. Thus, the RC contains data of nonselected chemicals. However, it has to be noted that selecting only published data may be a slight bias in itself because it identifies chemicals of scientific interest, public concern, etc., so that pharmaceuticals, pesticides, consumer products (e.g., cosmetics, food additives, etc.), and biocides are over-represented compared to industrial chemicals; the majority of the latter are of low toxicity (I. Gerner, BgVV, personal communication, as cited in Spielmann et al., [1999]).

The acceptance criteria for the *in vitro* cytotoxicity data were defined as follows:

- At least two different IC50 values were available, either from different cell types, or from different cell lines, or from different cytotoxicity endpoints.
- Only cytotoxicity data obtained with mammalian cells were accepted.
- Cytotoxicity data obtained with hepatocytes were not acceptable.
- The chemical exposure time in the cytotoxicity tests was at least 16-hr.

Only the following cytotoxicity endpoints were accepted:

- Cell proliferation: cell number, cell protein, DNA content, DNA synthesis, colony formation;
- Cell viability, metabolic indicators: MIT-24, MTT, MTS, XTTC;
- Cell viability, membrane indicators: Neutral Red Uptake (NRU), Trypan blue exclusion, cell attachment, cell detachment;
- Differentiation indicators.

The acceptance criteria for the *in vivo* data were defined as follows:

- Only LD50 values published in RTECS were used.
- If different issues of RTECS reported different LD50 values, then the first LD50 value was used for the RC. This value is also the highest value reported, since NIOSH replaces an LD50 value whenever a smaller value is available in the literature. A continuous change of *in vivo* data in the RC would not have been acceptable because the RC database had to be 'closed' to form a training data set (RC-I) and later a verification data set (RC-II). Therefore, since the beginning of data collection for RC-II, all LD50 values were only taken from

the 1983 RTECS issue, and later issues were not used.

The IC50 values from RC-I and RC-II, for a total of 347 chemicals, were obtained from 157 original publications in the literature. In the regression analysis for 347 chemicals, 1,912 single IC50 values were averaged (geometric means) per chemical to one IC50*x* value and then paired with 347 *in vivo* acute oral LD50 values. Whenever obtainable from RTECS, oral *in vivo* LD50 data from the rat were used (282 values). As a second priority, LD50 data from the mouse were used (65 values). Before data of rats and mice were merged in the RC, regression analyses performed separately with rat and mouse data justified this procedure (Halle, 1998). Although, by pairing 347 *in vitro* IC50*x* data with 347 *in vivo* LD50 data, an equal weight is given to each chemical, it has been criticized by reviewers that the IC50*x* is the geometric mean of a few up to many single data [minimum: n = 2, maximum: n = 32] per chemical. However, if the RC regression is recalculated with the means of only the smallest and the largest IC50 values per chemical, there are no differences in the regression function (Halle, personal communication).

To obtain a prediction model, a linear regression was derived from pairs of the log-transformed IC50*x* values and oral LD50 values (in mmol/kg), where 'a' is the intercept and 'b' is the regression coefficient, to produce the regression model [log (LD50) = b x log (IC50*x*) + a] shown graphically in Figure 2.1:

log (LD50) = 0.435 x log (IC50*x*) + 0.625

To allow comparison of the predictive value of the RC (or parts of the RC) with other similar approaches (prediction of the LD50 from basal cytotoxicity), an empirical linear-shaped prediction interval of a factor (F_G) of ± log 5 was defined (Figure 2.1). The linear-shaped boundaries should not be confused with the curved boundaries of a probability-based confidence interval. Halle defined this interval empirically as an acceptability measure based on information of the required and expected precision of rodent oral LD50 data (Halle and Spielmann,1992).

To evaluate the validity of the regression model, the key parameters of the regression for RC-I, RC-II, and RC-I+II (Table 2.1) were compared with the regression parameters obtained with single mammalian cell lines. Table 2.1 shows that all regression lines have essentially identical intercepts and regression coefficients (slopes) regardless of whether single parts of the RC or the whole RC were analyzed, or whether data from single studies with only one cell line were used. In addition, the percentage of data within the defined prediction interval (± log 5) is almost constant (73%-77%). In summary, the regression function derived from the RC, and from the RC subsets, seems to be a reliable description of the general relationship between basal cytotoxicity and rodent oral systemic LD50 values. This relationship can consequently be used as a mathematical model for prediction of rodent oral LD50 values from basal cytotoxicity.

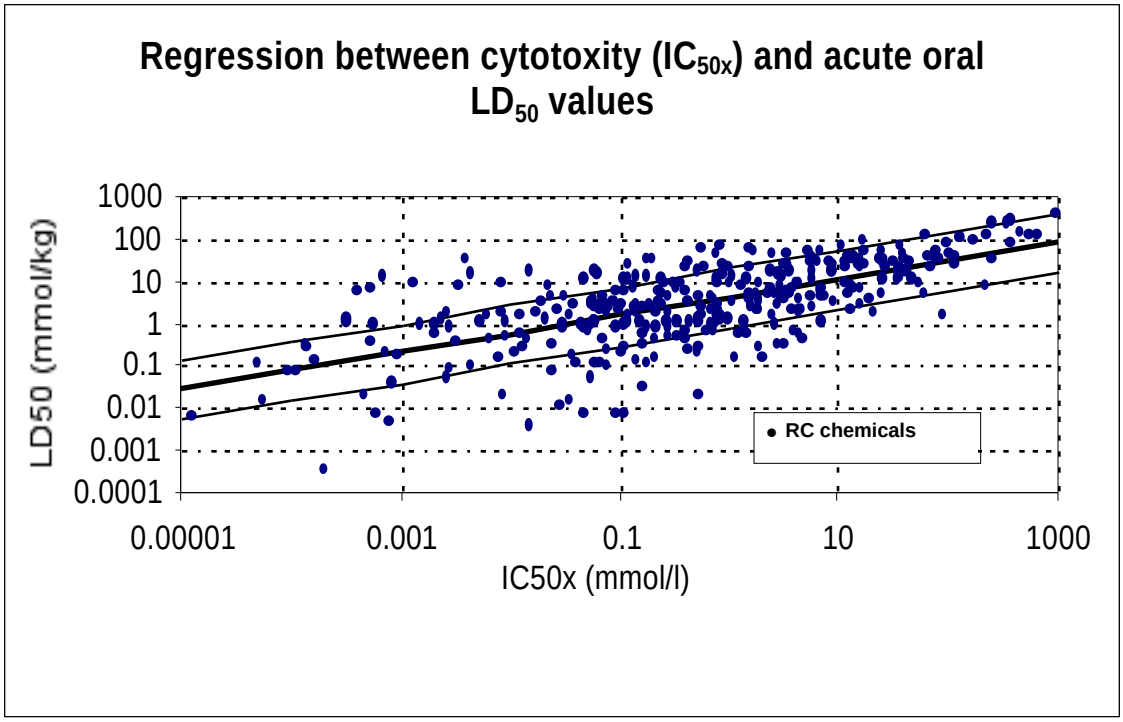

Figure 2.1. Registry of Cytotoxicity regression between cytotoxicity (IC50*x*) and rodent acute oral LD50 values of 347 chemicals
The heavy line represents the fit of the data to a linear regression model (r=0.67); the two additional lines represent the boundaries of ± log 5, an acceptance interval for this prediction model (Halle and Spielmann, 1992). This factor, F_G = ± log 5, was established based on information of the required and expected precision of LD50 values from rodent studies. The equation of the regression line (prediction model) reads: log (LD50) = 0.435 x log (IC50*x*) + 0.625.

Table 2.1. Linear regression parameters of two RC issues and two single studies using one cell line and one cytotoxicity endpoint

RC or Cell line**	Number of Chemicals (n)	Correlation Coefficient (r)	Intercept (a)	Regression Coefficient (b)	% Chemicals in Prediction Interval[a]	Reference[b]
RC-I *	117	0.667	0.637	0.477	74	1
RC-II *	230	0.666	0.634	0.414	73	2
RC-I+II *	347	0.672	0.625	0.435	73	2, 3, 4
BCL-D1**	22	0.720	0.536	0.633	77	5
3T3-L1 **	91	0.720	0.631	0.427	74	6

[a]Prediction interval for regression line is ± $F_G \leq$ log 5.
[b]References: 1 = Halle and Göeres, 1988; 2 = Halle, 1998; 3 = Halle et al., 1997; 4 = Spielmann et al., 1999; 5 = Knox et al., 1986; 6 = Clothier et al., 1988.

2.2.3 *Influence of the Starting Dose in the Acute Toxic Class (ATC) Method.*

Introductory note: The current accepted version of the ATC is the version adopted by the OECD in 1996 (OECD TG 423; OECD, 1996). Several updated drafts have been created since the OECD endorsed a new Globally Harmonized System (GHS) for the classification of chemicals in November 1998 (OECD, 1998a). The most recent draft of TG 423 was issued after the ICCVAM Workshop was held (OECD, October, 2000; http://www.oecd.org/ehs/test/health.htm). Consequently, the following analysis focuses on the 1996 OECD version of TG 423, but also attempts to address recent developments.

Following a national and an international experimental validation study of the ATC Method (Schlede et al., 1992, 1994; Diener et al., 1995), the ATC was accepted by the OECD (OECD TG 423; OECD, 1996) as an alternative to the classical LD50 test for acute oral toxicity. In the TG 423 procedure, a substance is tested in a stepwise dosing procedure with each step using three animals of a single sex at the same time. The proportion of survivors dosed at one step determines the next step, which is: (a) no further testing, or (b) dose three additional animals with the same dose, or (c) dose three additional animals at the next higher or the next lower dose. Originally, the method was developed and experimentally validated with two sexes and three different fixed starting doses (25, 200, and 2000 mg/kg body weight [b.w.]) reflecting the European Union (EU) hazard classification system. A thorough biometrical analysis (Diener et al., 1995) showed that the ATC is applicable to all hazard classifications currently in use.

Figure 2.2 shows, for example, that to classify a chemical as "toxic" or "very toxic", 1-2 consecutive steps could be saved if 25 mg/kg b.w. was used as the starting dose instead of the medium dose. With increasing distance between the true toxicity class and the starting dose, the number of dosing steps increases. This effect is shown in more detail in Table 2.2, which shows the expected number of animals used and the number that died in relation to starting dose and true LD50 for a dose-mortality slope of $\beta = 2$. Biometrical calculations with other slopes (from $\beta = 1$ to $\beta = 6$) revealed the dependency in Table 2.2 is only slightly affected by the dose-mortality slope (for details see Diener et al., 1995).

In summary, one to three dosing steps can be avoided if the optimum starting dose can be predicted from a preceding cytotoxicity test. Taking into account that approximately 75% of the LD50 values predicted from basal cytotoxicity tests are expected to fall within the prediction interval of $\pm$ log 5 (see Table 2.1), and, moreover, that the space between the three starting doses (25, 200, 2000 mg/kg b.w.) is a factor of about 10, it was anticipated that, for most chemicals, the starting dose predicted from cytotoxicity would have been the dose requiring the fewest consecutive steps to reach a classification.

In November 1998, the GHS for the classification of chemicals, which uses four toxicity classes instead of the three used by the current EU system, was endorsed by the OECD (OECD, 1998a). A fifth toxicity class (> 2000 – 5000 mg/kg b.w.) was additionally introduced for special regulatory purposes. As a consequence, the current updated Draft OECD TG 423 (OECD, October, 2000; http://www.oecd.org/ehs/test/health.htm) now uses four different starting doses (5, 50, 300, and 2000 mg/kg b.w.), but the upper boundary of the fifth class of 5000 mg/kg b.w. is not used as a starting dose. Figure 2.3 shows the proposed revision of the ATC.

For the version of the revised ATC to be consistent with the OECD GHS classification system, biometrical calculations of the expected number of animals used and dead in relation to starting dose, true LD50, and dose-mortality slope, have been published (Diener and Schlede, 1999). While any increase in the number of possible starting doses theoretically increases the potential to save dosing steps when using the optimal starting dose, only a small decrease in animal numbers is expected compared to the current ATC method because (a) the number of starting doses has been increased at the toxic end of the scale, where the prediction of the LD50 by IC50 is less accurate than at the non-toxic end of the scale, and (b) the entire scale is still about the same length.

INTERPRETATION OF RESULTS BASED ON OPTION 1 TESTING FOR COMMONLY USED CLASSIFICATION SYSTEMS

Starting dose: 200 mg/kg body weight

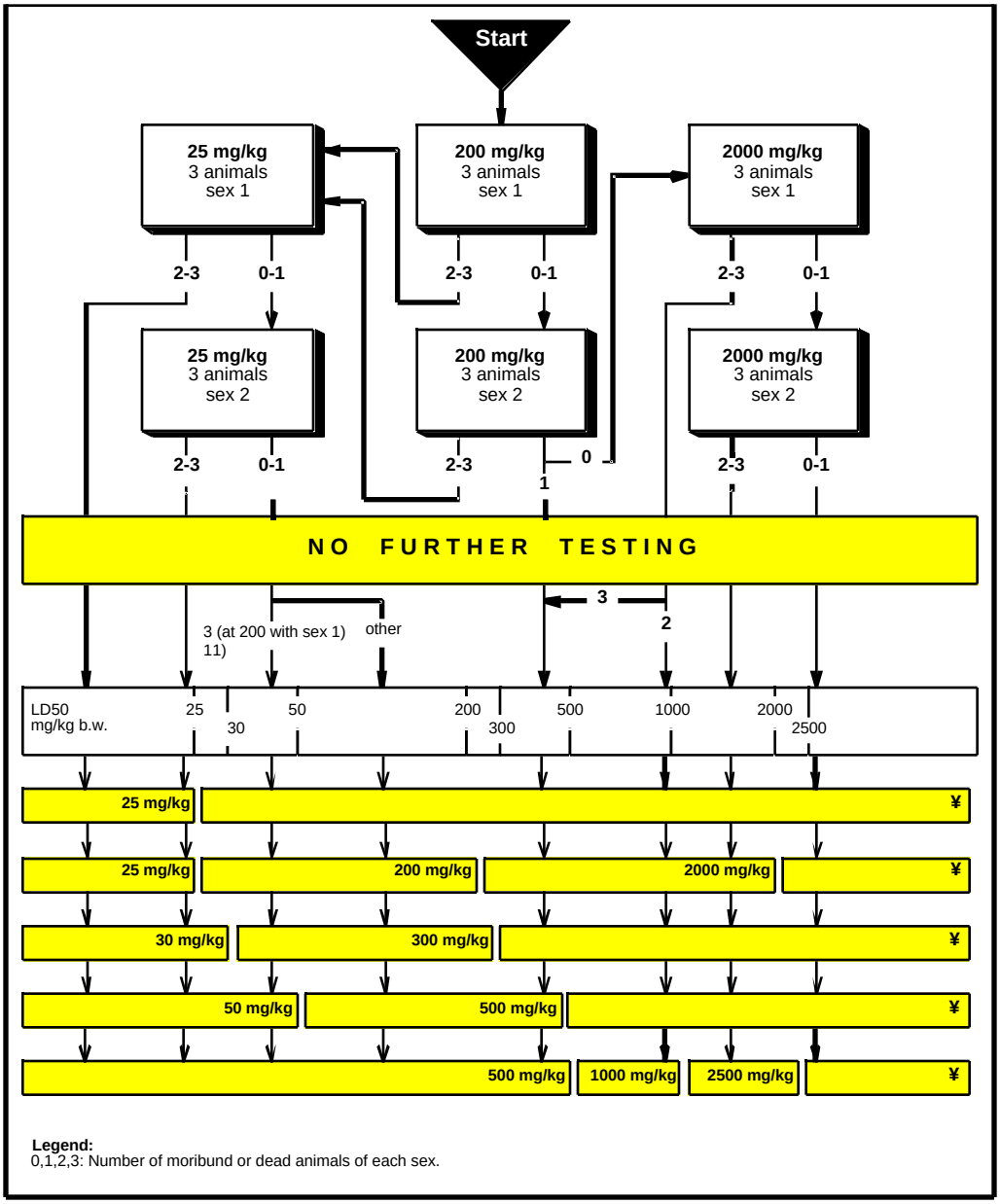

Figure 2.2 Principle of the Acute Toxic Class (ATC) method: medium starting dose
Source: OECD TG 423, Annex 3b (OECD, 1996). Example shows the possible dosing steps when 200 mg/kg b.w. is used as the starting dose. Depending on the toxicity of the test substance, 2 to 4 steps may be necessary to reach a classification according to hazard classification systems currently in use.

Table 2.2. Influence of the ATC starting dose on total number of animals (used and dead) in relation to the true LD50 for slope = 2[a]

	Starting dose in mg/kg body weight					
	25		200		2000	
True LD50	Used	Dead	Used	Dead	Used	Dead
1	3.0	3.0	6.0	6.0	9.0	9.0
2	3.0	3.0	6.0	6.0	9.0	9.0
5	3.1	2.8	6.1	5.8	9.1	8.8
10	3.4	2.7	6.4	5.6	9.4	8.6
20	4.6	2.8	7.2	5.3	10.2	8.3
50	7.5	3.3	8.6	4.2	11.6	7.2
100	9.3	3.2	9.3	3.3	12.2	6.2
200	11.2	3.2	9.7	3.1	12.0	5.3
500	14.0	3.3	9.3	3.3	10.0	3.9
1000	14.9	2.6	9.1	2.6	9.2	2,7
2000	15.4	1.8	9.4	1.8	9.3	1.8
5000	16.5	1.0	10.5	1.0	9.0	1.0
10000	17.3	0.4	11.3	0.4	7.7	0.4
20000	17.8	0.1	11.8	0.1	6.6	0.1
50000	18.0	0.0	12.0	0.0	6.1	0.0
100000	18.0	0.0	12.0	0.0	6.0	0.0

[a]Presented by W. Diener at the OECD ad hoc expert meeting on evaluation of the ATC in Berlin, Germany, 1994.

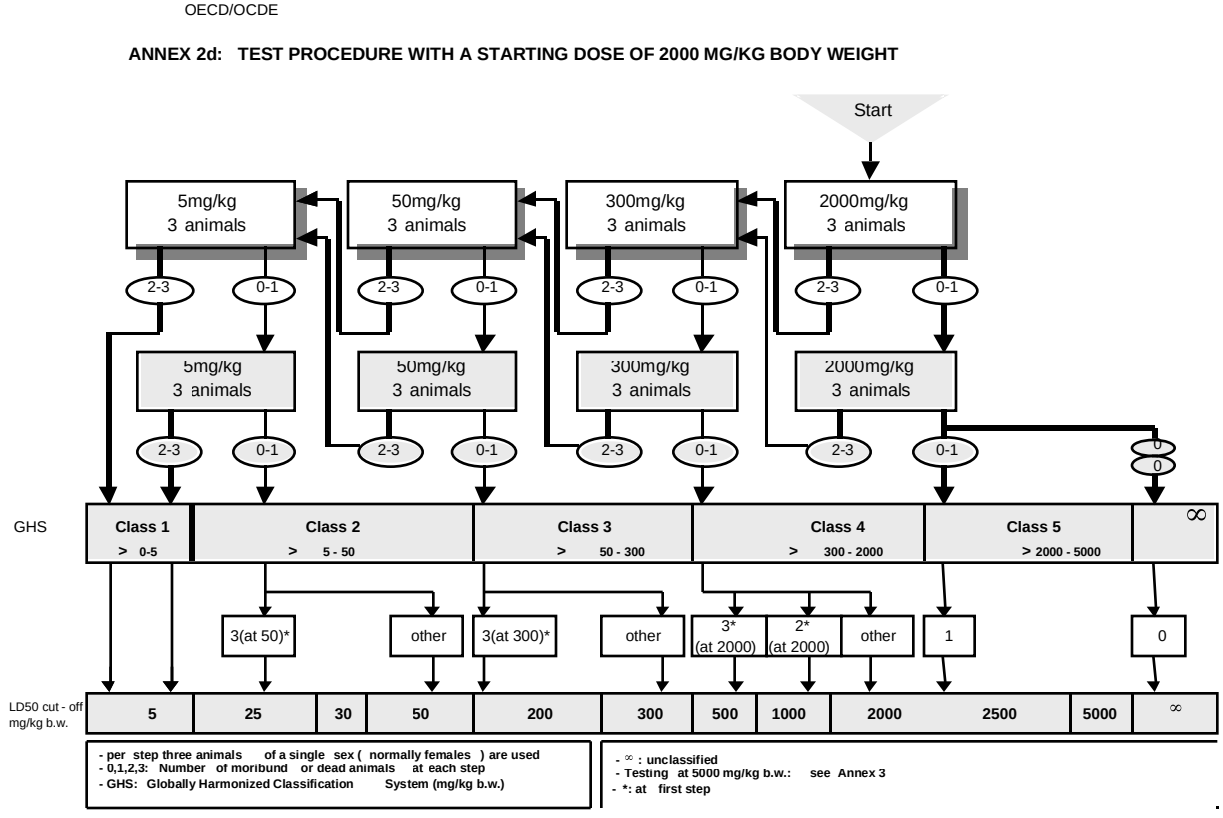

Figure 2.3. Proposed revision of the ATC to meet requirements of the OECD GHS
Source: OECD, Draft TG 423 (OECD, 2000). The number of new starting doses and spaces between have been changed so that the results from this test will allow a substance to be ranked and classified according to the GHS for the classification of chemicals which cause acute toxicity (OECD, 1998a).

2.2.4 *Influence of the Starting Dose in the Up-and-Down-Procedure (UDP)*

Introductory note: The current accepted version of the UDP is the version adopted by the OECD in 1998 (OECD TG 425; OECD, 1998b). Updated drafts of TG 425 have been created to allow for assessment of the confidence interval for the LD50 point estimate, and to include the application of new stopping rules and a larger dose progression factor, both of which tailor the UDP to the most efficient use of animals and improve the point estimate obtained. The most recent draft of TG 425 was issued after the ICCVAM Workshop was held (OECD, October 2000; http://www.oecd.org/ehs/test/health.htm). The analysis of the possible number of animals saved in a tiered approach is therefore based on the currently adopted 1998 OECD version of TG 425, but the significance for both versions can be assumed.

The concept of the up-and-down testing approach was first described by Dixon and Mood (Dixon and Mood, 1948; Dixon, 1965; 1991a, 1991b) and was later proposed to be used for the determination of acute toxicity of chemicals (Bruce, 1985). Apart from many biometrical publications refining the method (not cited here), a key review paper (Lipnick et al., 1995a) compared the results obtained with the UDP, the conventional LD50 test ([TG 401] OECD, 1981) and the Fixed Dose Procedure ([FDP; TG 420] OECD, 1992).

In principle, all versions of the UDP are stepwise procedures that use (as opposed to the ATC) single animals with the first animal receiving a dose at the best estimate of the LD50 (adopted TG 425, OECD 1998b), or one dosing step below the best estimate of the LD50 (most recent draft TG 425). Depending on the outcome for the first animal, the dose for the next is increased or decreased, either by a factor of 1.3 (adopted TG 425), or by a factor of 3.2 (recent draft TG 425). This sequence continues until there is a reversal of the initial outcome (i.e., the point where an increasing dose results in death rather than survival, or decreasing dose results in survival rather than death). After reaching the first reversal of the initial outcome, four additional animals are dosed following the up-down principle according to the adopted TG 425 (OECD, 1998b). In the most recent draft, however, a combination of stopping criteria is used to keep the number of animals to a minimum, while adjusting the dosing pattern to reduce the effect of a poor starting value or low slope. When one of the following criteria is satisfied, dosing is stopped and estimates of the LD50 and confidence interval are calculated according to the maximum likelihood method.

Three stopping criteria are defined in the draft UDP test guideline as follows:

(1) Three consecutive animals survive at the upper bound;
(2) Five reversals occur in any six consecutive animals tested (not just the first six);
(3) At least four animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. (Calculations are made at each dose following the fourth animal after the first reversal.)

Under certain circumstances, which are defined in the draft Guideline, statistical computation will not be possible or will likely give erroneous results. For most applications, testing will be completed with only four to six animals after an or the initial reversal in animal outcome [stopping rule (c)]

Since the UDP test guideline ([TG 425] OECD, 1998b) clearly states that the test performance of the method is optimal if the investigator's best estimate is used as a starting dose, Spielmann et al. (1999) have investigated the quality of LD50 estimates derived from the RC (Halle, 1998) for several chemicals used to validate the UDP (Lipnick et al., 1995a). Of the 35 chemicals used in the UDP validation study (Lipnick et al., 1995a), nine chemicals were also part of the RC (acetonitrile, p-aminophenol, caffeine, coumarin, dimethyl-formamide, mercury (II) chloride, nicotine, phenylthiourea and resorcinol). For four chemicals, the LD50 values predicted by the RC were almost exactly the same as those determined with the UDP *in vivo*, (i.e., the LD50 values determined in the UDP were on the regression line of the RC) (see Figure 1 in Spielmann et al., 1999). For three chemicals,

the predicted LD50 values were within the prediction interval of $\pm$ log 5, and for two chemicals (p-aminophenol and caffeine), the predicted LD50 values differed from the *in vivo* LD50 values by one order of magnitude (Spielmann et al., 1999). Thus, even in this small set of data, the 'basic rule' derived from the RC that about 75% of the LD50 values predicted from cytotoxicity (see Section 2.2.2, Table 2.1) are acceptable, was confirmed. This indicates that cytotoxicity assays could be successfully used to determine starting doses, and can reduce the number of animals for *in vivo* studies, particularly the UDP.

To date, no computer simulations have been performed to estimate the possible reduction in animal numbers if the combined *in vitro*/*in vivo* approach is applied to the UDP. Thus, the Workshop discussions were based on computations taken from the ICCVAM background document for the peer review of a recent revision of the UDP (ICCVAM, 2000) which are shown in a slightly improved way in Figure 2.4a and Figure 2.4b. Figure 2.4a applies to the stopping rule defined in the adopted TG 425 (OECD, 1998b), and Figure 2.4b shows the effect when the likelihood-ratio (LR) stopping-rule (current draft OECD TG 425) applies.

Since the LR rule is only one out of three stopping rules that should be applied in an adaptive way, additional computation will be needed to assess the influence of the starting dose on animal usage. The upper curves of both figures depict the numbers of animals used if the starting dose is two logs from the true LD50 (1/100 LD50) while the lower curves show the number of animals used if the true LD50 is used as a starting dose. The percentage of animals saved when the starting dose equals the true LD50 value is about 30% in Figure 2.4a, and independent of the dose mortality slope; whereas in the case of the LR stopping rule (Figure 2.4b), 25 to 40% fewer animals may be used, depending on the slope.

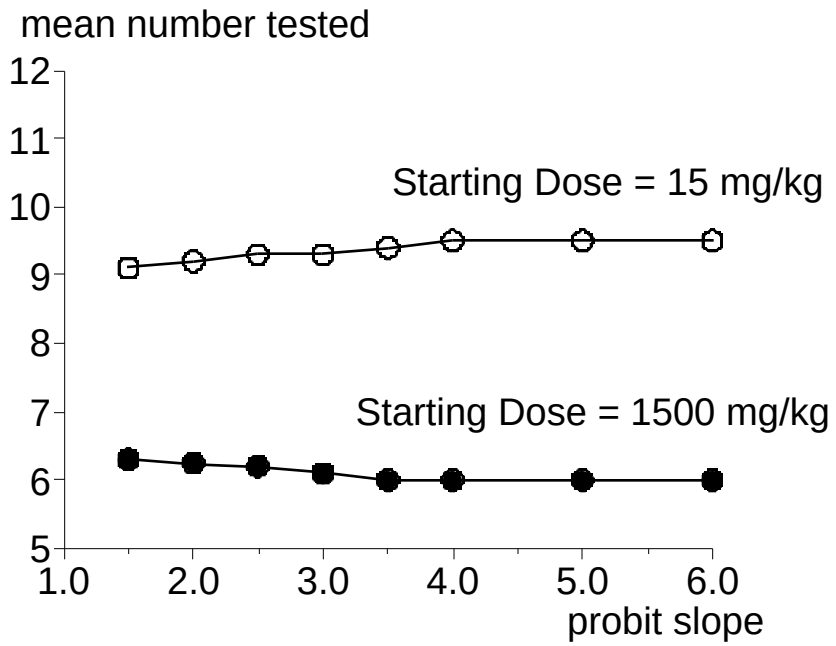

Figure 2.4a.Number of animals needed in relation to the starting dose for UDP adopted TG 425 (OECD 1998b) for LD50 = 1,500 mg/kg b.w.

The figure shows the number of animals needed if the LD50 is used as starting dose (lower curve), or if 1/100 of the LD50 is used as starting dose (upper curve). For details on the stopping rule applied see text.

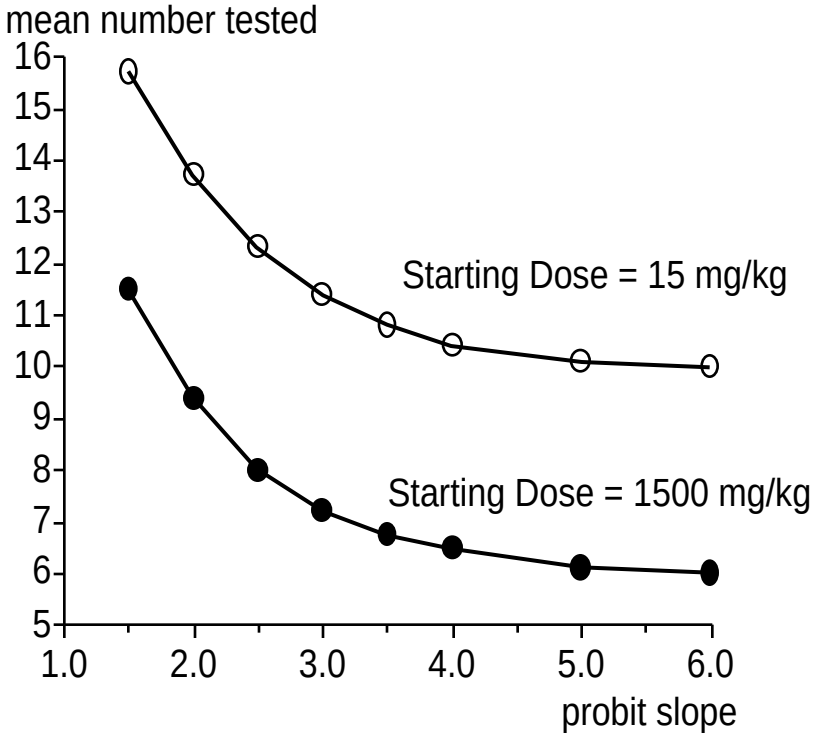

Figure 2.4b. Number of animals needed in relation to the starting dose for UDP draft TG 425 (OECD, 2000) for LD50 = 1,500 mg/kg b.w.
Figure shows the number of animals needed if the LD50 is used as starting dose (lower curve), or, if 1/100 of the LD50 is used as starting dose (upper curve) if the LR stopping rule singularly applies. For details see text.

2.2.5 *Prediction of a Limit Test Value from Basal Cytotoxicity Data*

According to a personal communication (Ingrid Gerner, BgVV) published by Spielmann et al. (1999), the notification process of new chemicals in the EU since 1982 revealed an unbalanced frequency distribution of the toxicity of industrial chemicals. No chemicals were classified "very toxic" (LD50 $\leq$ 25 mg/kg). Only 3% of the chemicals were classified "toxic" (LD50 > 25-200 mg/kg), while 21% were classified "harmful" (LD50 >200-2000 mg/kg), and the vast majority (76%) remained unclassified (LD50 > 2000 mg/kg). In other words, in the world of new industrial chemicals a clear majority are candidates for performing a 'limit test' where only the defined highest dose (2000 mg/kg most often, and occasionally 5000 mg/kg) is applied and no or marginal mortality occurs. Limit tests are defined in all OECD guidelines for acute oral toxicity testing (TG 401, TG 420, TG 423, and TG 425).

It must be emphasized that, if the limit dose defined in these guidelines is applied to all chemicals without knowledge of their toxicity, it would be correct for 76% of the chemicals, while 24% of the chemicals would cause avoidable deaths. It is therefore recommended to perform a limit test only if the prediction from a preceding basal cytotoxicity test suggests an LD50 value larger than the defined limit test dose. Special notice should be given to the fact that the precision of the prediction of low systemic toxicity from cytotoxicity data is much better than the precision of high systemic

toxicity. This is empirically supported by data from the RC (Halle, 1998) shown in Figure 2.1. The main factors affecting a strict log-linear relationship between basal cytotoxicity and systemic toxicity, bioavailability, and in some cases, biotransformation, play a minor role if a chemical is of low basal cytotoxicity.

2.2.6 *Evaluation of a Cytotoxicity Test Intended to be Used for Prediction of a Starting Dose*

This section describes how basal cytotoxicity data can be used to predict a starting dose for an *in vivo* lethality assay. Theoretically, any *in vitro* test that is capable of determining basal cytotoxicity could be used for determining the best estimate of a starting dose for acute testing in the UDP and ATC method. In addition, if the LD50 value predicted from cytotoxicity is high (≥ 2000 mg/kg b.w.), any of the currently used *in vivo* test protocols, including the FDP (OECD, 1992), would allow for performing an *in vivo* limit test without a proceeding sighting study.

In order to apply predictions of LD50 values obtained with experimental cytotoxicity data in the proposed tiered testing strategy as starting doses for the ATC or UDP methods, Spielmann et al. (1999) suggested a procedure shown in Figure 2.5. The authors suggested selecting 10-20 reference chemicals from the RC (Halle, 1998) and testing them in a standardized cytotoxicity test (Figure 2.5, Step 1). A promising candidate would be the BALB/c 3T3 NRU test that has proved robust in several validation studies. To allow comparison of the regression obtained with the in-house test (Figure 2.5, Step 2), reference chemicals should be selected to cover the entire range of cytotoxicity and to be as close as possible to the RC regression line.

Next, the in-house regression equation should be calculated by linear regression (least square method) using the new in-house IC50 values for the reference chemicals and the corresponding LD50 values from the RC. The resulting regression is then compared with the RC regression (Figure 2.5, Step 3). If the regression function obtained with the in-house cytotoxicity test is parallel to the RC regression and within the defined prediction interval, then the test is regarded suitable to be used without modification in applying the RC regression for future predictions of starting doses (Figure 2.5, Step 4). If the in-house regression shows a significantly higher or lower slope, then it may be possible to adjust the in-house test to a higher or lower sensitivity. However, it is likely that a more efficient approach would be to use a cell line and protocol, which have produced results that closely reproduce the RC data (recommended in the Guidance Document, ICCVAM, 2001).

The procedure of evaluating the usability of an in-house cytotoxicity test is explained in full detail in a special Guidance Document from this Workshop (ICCVAM, 2001), in which a set of 11 well-selected reference chemicals from the RC is recommended, and new experimental data obtained by testing the chemicals are presented. The data confirm that an in-house NRU cytotoxicity test, performed either with normal human keratinocytes (NHK) or with BALB/c 3T3 mouse cells, produces a regression line which matched the RC regression line ($R^2 > 0.9$).

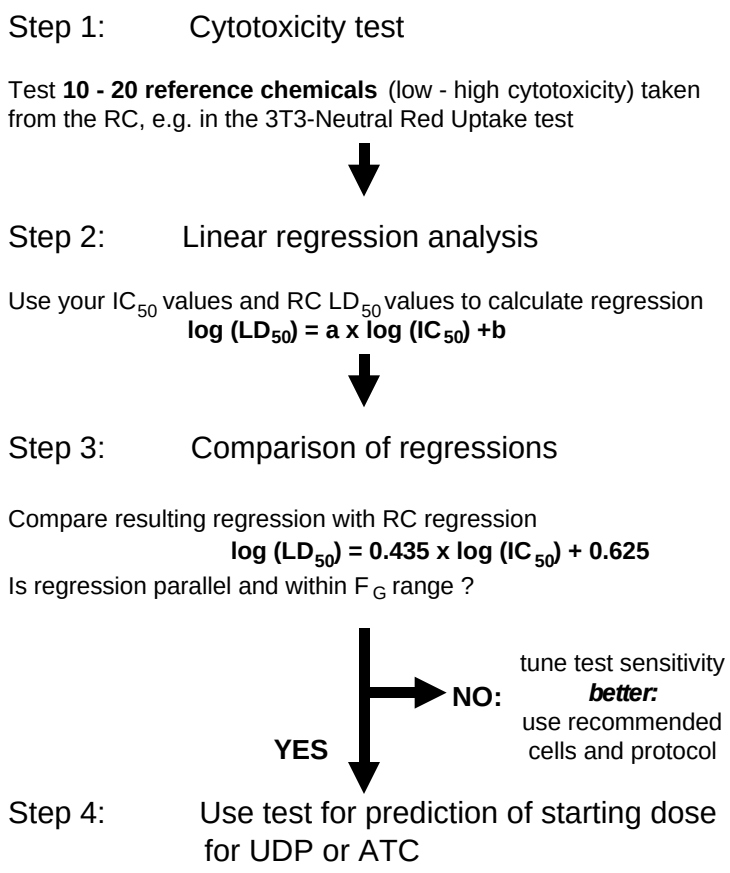

Figure 2.5. Procedure for evaluating a cytotoxicity test for tiered *in vitro*/*in vivo* testing for acute oral toxicity testing (slightly modified version of the scheme presented by Spielmann and colleagues).
Note: based on the expectation that many valid cytotoxicity tests would match with the RC regression, Spielmann et al. (1999) defined only the "yes" option between steps 3 and 4. A "no" option has been added here for clarity.

2.2.7 *Multicenter Evaluation of In Vitro Cytotoxicity (MEIC Approach)*

The MEIC program was established by the Scandinavian Society for Cell Toxicology in 1989. The intention of the program was to investigate the relevance of *in vitro* test results for predicting the acute toxic action of chemicals in humans directly rather than in rodents. Batteries of existing *in vitro* tests that have the potential to serve as replacements for acute toxicity tests were identified. The program was designed as an open study with all interested laboratories worldwide invited to participate and test 50 preselected reference chemicals in their particular *in vitro* toxicity assays (Bondesson et al., 1989). Minimal methodological directives were provided in order to maximize protocol diversity among the laboratories. Eventually, some 96 laboratories participated in this voluntary undertaking.

The 50 reference chemicals were selected to represent different classes of chemicals, with the availability of good data on acute toxicity (lethal blood [or serum] concentrations [LC] in humans; oral LD50 values in rats and mice) being a key determinant. Since the LC data available from clinical toxicology handbooks are average values with a wide variation, they were found to be sub-optimal for comparative purposes. Therefore, during 1995-97, the MEIC management team collected case reports from human poisonings with the reference

chemicals to provide LC data with known times between ingestion and sampling/death. The aim was to compile enough case reports to be able to construct time-related LC curves for comparison with the IC50 values for different incubation times *in vitro*. The results were presented and analyzed in a series of 50 MEIC Monographs (referred to as the MEMO subproject by the organizers).

When the MEIC project finished in 1996, all 50 reference chemicals had been tested in 61 different *in vitro* assays. Twenty of these assays used human-derived cells, 18 of which were cell lines and two were primary cell cultures. In 21 of the assays, the cells were of animal origin (12 cell lines and 9 primary cell cultures). Eighteen of the assays were ecotoxicological tests, and two were cell-free test systems. The majority of the assays were based on measurement of effects on cell viability or cell growth (or a combination of the two).

The test results submitted to MEIC were analyzed statistically using analysis of variance (ANOVA), principal component analysis (PCA), and partial least square analysis (PLS) techniques. The analyses conducted were based on *in vitro* cytotoxicity data presented as IC50 values. The predictability of *in vivo* acute toxicity from the *in vitro* IC50 data was assessed against human lethal blood concentrations compiled from three different data sets: clinically measured acute lethal serum concentrations, acute lethal blood concentrations measured post-mortem, and peak lethal concentrations derived from approximate LC50 curves over time after exposure (Ekwall et al., 1998a).

Statistical analysis of results from the 61 assays using the PLS model predicted the three sets of lethal blood concentrations well (R^2 = 0.77, 0.76 and 0.83, Q^2 = 0.74, 0.72, and 0.81, respectively, where R^2 is the determination coefficient and Q^2 is the predicted variance according to cross-validation in the PLS model used) (Ekwall et al., 2000). A two-component PLS model of the prediction of lethal doses in humans from published oral rodent LD50 values for the 50 MEIC compounds was less effective (R^2 = 0.65, Q^2 = 0.64) (Ekwall et al., 1998a; Ekwall et al., 2000).

The analysis showed that *in vitro* assays that were among the most predictive generally used human cell lines (6 of the 18 assays using them gave the highest determination coefficients, vs. 1 of 12 rat cell line assays that performed comparably). Two of 9 non-human primary cell assays analyzed also performed well. Assays that did not perform well were primarily ecotoxicological assays using bacteria or plant cells and, in general, assays with very short exposure times (up to a few hours). Two human primary cell assays, both of which utilized PMN leukocytes and involved 3-hour exposure times, also performed relatively poorly. These results led the authors to note that human-derived cells appeared to be the most predictive for human acute toxicity.

The exposure time for the *in vitro* assays was most often 24 hours, but ranged from 5 minutes to 6 weeks. For 22 of the 50 reference chemicals, the toxicity *in vitro* increased with increasing exposure time. However, high predictivity was generally observed in vertebrate cell assays with 24 to 168 hours exposure. The actual endpoint measurements (cell viability assays) used with the *in vitro* tests were not crucial. Typically, different endpoint measurements gave approximately the same result, suggesting that basal (general) cytotoxicity can be assessed using many mammalian cell lines and almost any growth/viability endpoint.

To select an optimal battery for predicting acute toxicity in humans, the MEIC management team further evaluated various combinations of assays using PLS models and 38 chemicals deemed to have the most reliable and relevant lethal peak concentration data (see Ekwall et al., 2000, for the detailed procedure). From their analysis, the most predictive and cost-effective test battery consisted of four endpoints/two exposure times (protein content/24 hours; ATP content/24 hours; inhibition of elongation of cells/24 hours; pH change/7 days) in three human cell line tests. The test battery (designated 1,5,9/16) was found to be highly predictive of the peak human lethal blood concentrations of all 50 chemicals (R^2 = 0.79, Q^2 = 0.76) when incorporated into an algorithm developed by the team. The R^2 value was further improved to 0.83 when information

on BBB penetration was added to the battery results.

It was noted that passage across the BBB can be predicted from the chemical formula and/or physico-chemical properties, or from *in vitro* tests in appropriate model systems; however those methods were not used in the MEIC analysis. The MEIC team proposed that the cell battery they identified could be used immediately for many non-regulatory purposes in a multistep testing strategy and urged its formal validation (and/or that of other promising cell assays also identified in the MEIC program) as soon as possible (Ekwall et al., 2000). Test protocols for evaluating the proposed assays in a validation exercise remain to be developed and optimized.

In summarizing, the MEIC team concluded that their study yielded a limited battery of *in vitro* assays using human cell lines that showed very good performance and were cost effective for predicting acute lethality in humans (Ekwall et al., 2000). However, to further improve the predictive capability of this proposed battery, and to take into account non-basal cytotoxicity factors as a full replacement for acute animal tests, further, targeted development of *in vitro* methods for other particular endpoints is needed. An evaluation-guided development of new *in vitro* tests (EDIT) has been proposed to address these requirements (Ekwall et al., 1999), which includes, as most urgently needed, *in vitro* assays for:

- Assessing passage through the BBB;
- Predicting gut absorption;
- Distribution volume;
- Biotransformation.

The results of the MEIC program have appeared in a series of publications in the open literature (Clemedson et al., 1996a; Clemedson et al., 1996b; Clemedson et al., 1998a; Clemedson et al., 1998b; Ekwall et al., 1998a; Ekwall et al., 1998b; Ekwall et al., 1999; Clemedson et al., 2000; Ekwall et al., 2000). Additional information about MEIC, MEMO and EDIT, as well as the MEMO database, can be found at the following Internet address:

http://www.cctoxconsulting.a.se/nica.htm

2.3 Identifying Needs

In the area of human health effects, the overall aim is to reliably and accurately predict the potential for human acute toxicity. The Breakout Group noted that there is extensive documentation showing that human outcomes from chemical exposure are not predicted well by studies in rodent species (see, e.g., Ekwall et al. [2000] and the recent survey by Olson et al. [2000] on target organ toxicity). Consequently, it was agreed that the long-term goal (the ideal approach) should be the use a battery of *in vitro* tests employing human (rather than rodent or other animal) cells and tissues to provide data which when combined with information derived from other sources (e.g., on key physico-chemical parameters, kinetics, and dynamics) could more accurately predict human acute toxic effects including lethality. However, in the near term, the Breakout Group considered it appropriate and more pragmatic to concentrate on ways to reduce and replace animal use in acute oral toxicity tests as detailed in OECD TG401, TG420, TG423, and TG425.

The Breakout Group was fully aware that rather more information than just an (approximate) LD50 value can be obtained and used from a properly conducted rodent acute toxicity test (such as clinical signs, dose-response relationships, possible target organs, etc.); however, it received reassurance from the U.S. regulatory agencies represented at the Workshop that if there was a validated *in vitro* cytotoxicity test which could accurately predict the approximate rodent LD50 value *in vivo*, then its implementation would result in a significant reduction in animal use. Thus, the primary focus of Breakout Group 1 was to identify and evaluate candidate *in vitro* cytotoxicity tests that could possibly serve as reduction and replacement alternatives for current rodent acute oral toxicity tests for determining LD50 values.

2.3.1 *Near-term (< 2 years) Goals and Potentially Attainable Objectives*

The Breakout Group participants started from the premise that it is biologically plausible that

cell death (cytotoxicity) *in vitro* could be used to predict acute lethality. The many studies that show relatively good correlations between *in vitro* IC50 values and *in vivo* LD50 data support this view (e.g., Phillips et al., 1990; Garle et al., 1994). Thus, the near-term focus should be on conducting studies aimed at reducing and replacing animal use for determining LD50 values of chemical substances.

The Breakout Group agreed that standardized *in vitro* test protocols were available but probably not optimized, and that prediction models were needed for predicting acute oral LD50 values. Consequently, a prevalidation study, which would include several promising candidate *in vitro* cytotoxicity tests, would have to be undertaken in order to determine which tests should go forward to the validation stage. Partly because of this, the development of a practical replacement test will take time. As a parallel activity, the ZEBET method for generating cytotoxicity data to help establish the starting dose for *in vivo* testing of new chemical substances (Spielmann et al., 1999) should be seriously considered as an interim measure to potentially reduce the numbers of animals used in the *in vivo* tests.

2.3.2 *In Vitro Endpoints for Assessing In Vivo Acute Toxicity*

There is considerable literature covering a large variety of endpoints and endpoint measurements that have been evaluated for *in vitro* cytotoxicity testing (e.g., Phillips et al., 1990; Balls and Fentem, 1992; Garle et al., 1994; Itagaki et al., 1998a; 1998b; Ohno et al., 1998a; 1998b; 1998c; Tanaka et al., 1998; Clemedson and Ekwall, 1999; Ekwall, 1999). Some of these citations were provided to the Breakout Group members for reference, but time did not allow a systematic assessment of the literature on this topic. It was noted nevertheless that, in practice, basal function endpoints (such as NRU or MTT reduction and/or inhibition of cell proliferation), even though they may measure different cellular functions, have been commonly used with a reasonable degree of success; where cell lines are concerned, the endpoints typically assess a combination of both cell death and cell growth/proliferation. Since the events are based on cellular events that have circumstantial if not direct relevance to cellular responses to chemicals *in vivo*, model cell systems incorporating these "nonspecific" endpoints may satisfy requirements for fidelity and discrimination for alternative methods that have been set forth earlier (Blaauboer et al., 1998). The need for cell-specific or functional endpoints in acute toxicity assays was considered to be on a case-by-case basis and more relevant to studying target organ-specific toxicities (Breakout Group 3's charge).

2.3.3 *Other Issues for Selecting Protocols*

The key components of the protocols for *in vitro* cytotoxicity tests were considered to be the appropriate choice of: (a) cell type (human or animal, cell line or primary cultures) and its characteristics (stability, origin, characterization, availability); (b) exposure period(s) – (i.e., duration cells are exposed to the test chemical); and (c) endpoint measurement(s) – (i.e., cell viability assays such as NRU, lactate dehydrogenase [LDH] leakage, ATP content) (Borenfreund and Puerner, 1986; Riddell et al., 1986; Phillips et al., 1990; Balls and Fentem, 1992; Garle et al., 1994; Ekwall, 1999; Ohno et al., 1998a; Ekwall, 1999; Ekwall et al., 2000). In addition, the inclusion of a prediction model, evidence of repeatability, and facility of transfer between laboratories are important considerations (Balls et al., 1995; Bruner et al., 1996; Archer et al., 1997; ICCVAM, 1997). Ease of automation/high throughput where applicable should offer attractive additional cost benefits but is not a requirement for validation purposes.

2.3.4 *QSAR Models for Predicting Acute Toxicity*

The Breakout Group was requested to assess the role of QSAR, or related models such as structure-activity relationships (SAR) in predicting acute toxicity. While SAR methods involve qualitative assessment of chemical features that confer biological properties, QSAR approaches develop a quantitative relationship between physico-chemical or structural properties and biological activity (Albert, 1985; Barratt et al., 1995). QSAR models are usually developed for sets of chemically similar

compounds on the assumption that they will have the same mechanism of action. Any compounds that do not act by the same mechanism are likely to fit the correlation poorly, and thus their effects would not be predicted accurately. Although defining chemical classes or commonality of mechanisms of action are not trivial due to the multidimensional nature of both characteristics, a review of QSAR studies for predicting LD50 values concluded that QSAR methods have shown some success in relating LD50 values to certain physico-chemical properties of a compound, especially lipophilicity (Phillips et al., 1990).

In contrast, QSAR approaches appear to be less successful in correlating electronic properties of molecules (related to reactivity), or structural variables, with LD50 values, and their use with certain important chemical classes, (e.g., pesticides), is problematic. However, the Breakout Group felt that it lacked sufficient expertise in the field to evaluate the potential of QSAR as a replacement test for lethality and suggested that the topic be reviewed more thoroughly by a more appropriate scientific body. The review should include coverage of commercially available models (e.g., TOPKAT, CASE).

The Breakout Group did recognize that these methods might play key roles as adjuncts to improve LD50 predictions and to reduce animal usage. As noted by others (e.g., Barratt et al., 1998; Lipnick et al., 1995b), QSAR can aid in a number of areas, including the selection of test chemicals for validation studies, the interpretation of outliers, and the grouping of chemicals by structure and biological mechanisms. In addition, looking to future requirements to improve the predictive capability of *in vitro* cytotoxicity data for *in vivo* LD50 values, the Breakout Group agrees with Breakout Group 2 in recommending a more thorough evaluation of QSARs for predicting gut absorption and passage across the BBB. These applications were discussed at length by Breakout Group 2.

The Breakout Group noted that, in principle, expert systems, neural networks, and classical structure-activity approaches might be developed and validated for predicting specific systemic effects (Barratt, 2000; Dearden et al., 1997; Phillips et al., 1990). Requirements for the successful development and use of QSAR methods have been identified and include the following:

- A well-defined mechanism of action for the compound(s) used to derive the QSAR model;
- Use of congeneric, pure compounds and not mixtures;
- A common site of action for the biological effect;
- For comparative purposes, expressing concentrations or doses in molar (not weight) units;
- Validation of each model by investigating its predictive capability using a different set of compounds from its learning (i.e., training) set;
- Use of the same ranges of parameter space as the original test chemicals; and
- The QSAR should not be applied outside of its domain of validity (Phillips et al., 1990; Barratt et al., 1995; Worth et al., 1998).

The limitations or general applicability of each model for different chemical classes will need to be established. The application of QSAR procedures for identifying potential systemic effects was considered by Breakout Group 2.

2.4 Current Status

Many investigations of the relationship between *in vitro* cytotoxicity and acute toxicity *in vivo* have been reported. It was not possible to critically review and discuss all of the literature during the course of the Workshop, so the Workshop organizers made a selection of recent key activities and reports for consideration by Breakout Group 1. The Breakout Group made note of the fact that many of these recent initiatives build upon the conclusions of studies conducted, in particular, during the 1980s (e.g., Balls et al., 1992; Balls and Clothier, 1992; Balls and Fentem, 1992; Borenfreund and Puerner, 1986; Clothier et al., 1987; Dierickx, 1989; Ekwall, 1983; Ekwall et al., 2000; Fentem et al., 1993; Fry et al., 1988; Fry et al., 1990; Garle et al., 1987; Garle et al., 1994;

Gülden et al., 1994; Guzzie, 1994; Halle and Spielmann, 1992; Hopkinson et al., 1993; Hulme et al., 1987; Ohno et al., 1998a; Phillips et al., 1990; Riddell et al., 1986; Seibert et al., 1996; Spielmann et al., 1999; Wakuri et al., 1993; Zanetti et al., 1992).

The studies and approaches considered were:

- Studies conducted by FRAME and partners (e.g., Balls et al., 1992; Fry et al., 1990; Hulme et al., 1987; Riddell et al., 1986);
- The MEIC scheme (e.g., Clemedson and Ekwall, 1999; Ekwall et al., 2000);
- Japanese Society of Alternatives to Animal Experiments (JSAAE) activities (e.g., Ohno et al., 1998a);
- The ZEBET approach for predicting *in vivo* starting doses (Halle et al., 2000; Halle and Goeres, 1988; Spielmann et al., 1999);
- Testing strategy outlined in ECVAM Workshop Report 16 (Seibert et al., 1996);
- Testing framework proposed under the auspices of SGOMSEC (Curren et al., 1998);
- TestSmart acute systemic toxicity initiative to determine whether cellular changes can predict acute system failure *in vivo* (A. Goldberg, personal communication).

The MEIC and ZEBET approaches were presented to the Breakout Group as specific proposals for adoption as alternative methodologies by regulatory authorities, and therefore received the most attention.

2.4.1 *In Vitro Methods for Estimating Acute In Vivo Toxicity*

There are more than 80 variations of *in vitro* basal cytotoxicity tests, employing a variety of cell lines (e.g., HeLa, HL-60, BALB/c 3T3, Chang cells) and endpoint measurements (e.g., MTT reduction, NRU, ATP content, LDH leakage). From the results of the MEIC and ZEBET programs it appears that basal cytotoxicity can be determined using almost any cell line and almost any toxicity endpoint measurement that correlates well with cell death and/or growth inhibition. Standard protocols are available for some of these methods (e.g., via the *INVITTOX* database run by ECVAM, from the JSAAE validation study, and by slight modification of test protocols used for other purposes such as phototoxicity or eye irritation testing), but these have not necessarily been optimized for predicting rodent oral LD50 values.

Typically, prediction models have not been explicitly defined, although they are usually based on the IC50 value derived in the *in vitro* cytotoxicity assay. Some of these initiatives made note of that and tried to define useful testing strategies that incorporated *in vitro* assays. An example was the ECVAM Workshop report, which to some extent was based on work from the University of Kiel, recognizing the importance of including biokinetic parameters alongside *in vitro* cytotoxicity data to improve the predictions (Seibert et al., 1996).

2.4.2 *Strengths and Limitations of Available In Vitro Cytotoxicity Assays*

Sufficient information was presented to the Breakout Group for evaluating the merits of the MEIC and ZEBET proposals and the JSAAE study in that the information could be adapted and utilized for evaluating assays designed to predict acute lethality.

The MEIC proposal was that a battery of three human cell-based tests (HepG2, protein content, 24 hr exposure; HL-60, ATP content, 24-hr exposure; Chang liver cell morphology, 24 and 168-hr exposure) could be used to predict human lethal blood concentrations and be a surrogate for the LD50 test (Ekwall et al., 2000). Although the MEIC program was not set up as a validation study and assessing reproducibility was not an objective, the Breakout Group agreed with the following MEIC conclusions:

(1) There is a strong correlation between concentrations of chemicals causing cytotoxicity *in vitro* and human lethal serum concentrations.

(2) Metabolism may not play a role *in vivo* as frequently as thought.

(3) Specificity of action requiring many types of differentiated cells is not as significant a problem as may initially have been envisaged.
(4) Some simple corrections of the data, such as for BBB passage, improve the correlations observed.

The key strengths of the MEIC approach are the comparison of acute cytotoxicity data with human exposure data and the database on human lethal concentrations, kinetic profiles, etc., which has been generated and is available as MEMO monographs for others to evaluate and use. The Breakout Group agreed that attempts be made to extend this human database, and that it should be subjected to independent peer review. The outcome of the MEIC program in general was considered to provide strong support for the concept of basal cytotoxicity first proposed by Ekwall in 1983.

Several issues were raised concerning the MEIC proposal and the use of such an approach as an alternative to animal tests. Various limitations of the approach were cited, including the following:

(1) Because the program was not intended to be a validation study, it was not conducted under controlled conditions.
(2) Replicate assays were generally not performed, hence there is limited information on intra-laboratory assay repeatability and inter-laboratory reproducibility. Nevertheless, there is a large body of evidence from other validation studies that *in vitro* cytotoxicity assays are highly reproducible and relatively easy to transfer between laboratories.
(3) The chemicals tested in the different laboratories were probably from different batches and sources (allowed by MEIC for practical purposes, and because the human case exposures likely involved different materials and sources also).
(4) Statistical analyses were often performed on groups of tests rather than on individual assays.
(5) In many of the assays, not all 50 chemicals were tested. This impacts on the conclusions being made on the basis of correlation coefficients;
(6) There is a tendency for the data to be over-interpreted and some of the conclusions have been over-stated in the publications.
(7) Prediction models were not defined for any of the *in vitro* assays. This would be a pre-requisite for a validation study.

There were also specific confounding factors in relation to the 1, 9, 5/16 battery proposed by Ekwall and colleagues (Ekwall et al., 2000). The assay battery was selected using data from 38 of the 50 MEIC chemicals, and the predictivity for all 50 chemicals reassessed by PLS analysis. The values obtained were: R^2=0.84, 38 chemicals; R^2=0.77, 50 chemicals; R^2=0.88, 38 chemicals + BBB correction; R^2=0.83, 50 chemicals + BBB correction. However, it was noted that: (a) results for test 1 were reported for only 45 chemicals, and 3 of the missing 5 results were for chemicals included in the first set of 38, thus n=35 and n=45; in addition, three other *in vitro* tests employing HepG2 cells and a 24-hr exposure time were evaluated in the MEIC program, and the data vary considerably, particularly for some of the reference chemicals; (b) results for test 9 were reported for only 46 chemicals, and all 4 of the missing results are for chemicals included in the first set of 38, thus n=34 and n=46; and (c) tests 5/16 used Chang liver cells, which are known to possess several HeLa markers. In addition, only single data points for each combination of *in vitro* test and chemical have been reported, meaning that there is no way to evaluate the variability in the assay results which would necessarily impact upon the robustness of the conclusions drawn by the MEIC management team.

A major strength of the ZEBET RC approach is the extensive database underpinning the strategy proposed (Spielmann et al., 1999). The database includes IC50 values derived from numerous *in vitro* cytotoxicity tests on more than 300 chemicals. The actual data are used in a very defined way in trying to predict starting doses for *in vivo* testing, and the simplicity of the concept, flexibility in choice of potentially useful cell systems, and ease of validating and applying the cell systems in practice are attractive features of the approach.

One disadvantage of the ZEBET approach at the present time is the lack of information on the variability in both the *in vitro* and *in vivo* data. In addition, the use of LD50 values from RTECS is perhaps a problem because of this. The Breakout Group suggested that several follow-up actions be undertaken immediately after the Workshop to update and improve the understanding of the applicability of this approach: (a) the examples shown for using *in vitro* cytotoxicity data to identify the starting dose for the ATC or UDP *in vivo* study should be updated to bring them in line with the new draft guidelines, which have now been modified to incorporate the OECD harmonized hazard classification system (OECD, 1998a); and (b) additional simulation modeling should be undertaken to demonstrate the actual reduction in animal use which is expected to be achieved by implementing the approach, and real-life worked examples should be provided to serve as guidance for those adopting and evaluating the approach in the future (See Section 2.6).

2.4.3 *Validation Status of Available In Vitro Screening Methods*

The Breakout Group considered the validation status of the *in vitro* cytotoxicity assays evaluated in the MEIC program, and those used to generate the data included in the RC, relative to the ICCVAM Validation Criteria (ICCVAM, 1997) and the ICCVAM Evaluation Guidelines (ICCVAM, 1999; Section 11, Appendix E). It was concluded that no single *in vitro* cytotoxicity test, or test battery, has yet been formally validated for the specific purpose of replacing the rodent LD50 test. Upon completion of the MEIC study, Ekwall suggested that the battery of three tests proposed should now undergo formal validation (Ekwall et al., 2000). Typically, data on the intra- and inter-laboratory reproducibility of the *in vitro* assays, generated in a structured manner, are lacking, and further work is still needed to fully evaluate the predictive ability of *in vitro* cytotoxicity tests for acute toxicity *in vivo*.

Since several *in vitro* cytotoxicity assays have been included in formal validation studies on eye irritation and phototoxicity (e.g., various test protocols using BALB/c 3T3 mouse fibroblasts or keratinocytes and NRU as the endpoint measurement [Balls et al., 1995b; Brantom et al., 1997; Spielmann et al., 1996; Spielmann et al., 1998]), objective data on the intra-laboratory and inter-laboratory reproducibility of these tests are available for test materials which were coded and tested in at least three laboratories. The Breakout Group proposed that a Working Group be established to evaluate this information and to undertake a paper exercise to determine the capability of these particular *in vitro* cytotoxicity tests for predicting rodent LD50 values rather than Draize rabbit eye irritation scores. It was envisaged that LD50 data would be available for most of the chemicals tested in the EC/HO and BgVV eye irritation validation studies.

A validation study on five *in vitro* cytotoxicity tests (endpoint measurements: colony formation, crystal violet staining, LDH release, MTT, and NRU) has been conducted under the auspices of the JSAAE (Ohno et al., 1998a). Six chemicals (Tween 20, Tween 80, sucrose fatty acid ester, propylene glycol, cetylpyridinium chloride, and sodium lauryl sulfate) were tested. The LDH release endpoint measurement was not reproducible, and the crystal violet staining assay was deemed to be the most reliable of the *in vitro* cytotoxicity tests evaluated (Ohno et al., 1998a). The colony formation assay in HeLa S3 (SC) and BALB/c 3T3 A31-1-1 cell lines was reported to be the most sensitive, but also showed the largest variation (Tanaka et al., 1998).

Disadvantages of the colony formation assay are that it is time-consuming (7 to 13 days culture time, depending on the cell line) and cannot be conducted in 96-well plates and, hence, cannot be readily automated. Although the focus of the study was on comparisons with Draize eye irritation scores and not acute lethality *in vivo*, the study does provide another source of objective information on the general reproducibility and transferability of *in vitro* cytotoxicity tests (Ohno et al., 1998a). In that sense, the Working Group should also examine the data from this study for how well they predict rodent LD50 values for the test chemicals.

Based on consideration of the studies referred to in previous sections, it was concluded that none of the available *in vitro* methods or proposed

testing strategies had been adequately evaluated for implementation to reduce and/or replace animal use for acute systemic toxicity testing. However, it was suggested that the ZEBET approach, using *in vitro* cytotoxicity data to predict *in vivo* starting doses, should be implemented relatively quickly once a guidance document had been prepared (see Section 2.6). The rapid adoption of the ZEBET approach into general practice would enable data to be generated in a relatively short time to fully establish its usefulness and accuracy with a large number of test chemicals.

2.4.4 *Selection of the Most Appropriate Cell Type*

The selection of the most appropriate cell type depends on the objective. Thus, for the prediction of rodent LD50 values in a replacement test, one would conceptually favor a rodent cell line; for the human situation, human cell lines would be more appropriate. Although the MEIC results tend to support this view, the Breakout Group did not feel the data were strong enough (for the reasons given above) to come to a definitive conclusion on this point. Further evidence of this was provided by an analysis of the ZEBET RC data relative to IC50 data generated using a human cell line evaluated in the MEIC program (Clemedson et al., 1998a; Clemedson et al., 1998b). The correlation between the IC50*x* (RC) and IC50*m* (MEIC human cell line) values for the 50 MEIC chemicals was extremely high (R^2=0.90; see Addendum to this report). Consequently, where the objective is to reduce animal numbers required for lethality tests, the apparent difference is too small to rule out the use of a human cell line if that cell line offers other particular advantages or performs acceptably for that purpose.

The current *in vitro* basal cytotoxicity tests do not take into account metabolism-mediated toxicity. It is widely accepted that simple predictive systems (*in vitro* or *in silico*) will need to be developed for early identification of those substances likely to be metabolized to more toxic or less toxic species than the parent chemical (e.g., Fentem et al., 1993; Seibert et al., 1996; Curren et al., 1998; Ekwall et al., 1999). It should be noted that in Ekwall's early studies, approximately 20% of the chemicals assayed in HeLa cell cultures did not fit the basal cytotoxicity concept (Ekwall, 1983). It is expected from the existing literature that "biotransformation screens" will provide valuable data to supplement *in vitro* cytotoxicity results for improving predictions of LD50 values for a significant fraction of those chemicals.

2.5 Future Directions

The Breakout Group concentrated its efforts mainly on short-term approaches to reduce and replace animal use in acute oral toxicity tests, leaving the discussion of longer-term research needs and priorities to Breakout Groups 2 (biokinetics) and 3 (specific organ toxicity and mechanisms). However, it was agreed that the long-term goal (i.e., the ideal approach) should be to develop and use a battery of *in vitro* tests employing human cells and tissues, and integrate this information with that derived from other sources (e.g., on key physico-chemical parameters, kinetics, and dynamics) to predict human acute toxicity, including systemic target organ effects.

2.5.1 *Most Promising In Vitro Methods for Further Evaluation to Reduce and/or Refine Animal Use for Acute Toxicity*

The Breakout Group considered that, in the absence of other information which enables the dose to be set with confidence (e.g., acute toxicity data on structurally related chemicals, physico-chemical or other information), *in vitro* cytotoxicity data generated using the proposed ZEBET approach should be useful for predicting starting doses for *in vivo* studies. The proponents presented supporting data indicating that this approach would result in a further reduction and refinement in animal use for acute toxicity testing. By judicious use of time and resources, initial cytotoxicity assays need not slow the overall developmental or evaluation processes and in fact may actually expedite it where several chemicals can be tested *in vitro* at the same time.

To use the approach, test laboratories should evaluate and compare the performance of several *in vitro* cytotoxicity tests with the

existing RC data (Figure 2.1). For example, a protocol employing the BALB/c 3T3 mouse fibroblast cell line, a 24-hour exposure time, and NRU as the endpoint measurement is appropriate, but other cell lines and cell viability assays could serve the same purpose equally well. The main considerations are:

- The selection of cell type for assessing general cytotoxicity (e.g., rodent fibroblast cell line, human epithelial cell line; monolayer or suspension [e.g., HL60 human acute leukemia cell line] cultures);
- Exposure period (a minimum of 24 hours, but consideration of longer exposures [e.g., 72 hours] as well, if appropriate);
- Endpoint (cell viability/growth);
- Endpoint measurement (e.g., NRU, MTT, ATP, protein).

Since the choice of endpoint measurement does not appear to be critical to the correlative power of the tests (Garle et al., 1994; Ohno et al., 1998a; Spielmann et al., 1999; Ekwall et al., 2000), the simplest, cheapest, most reproducible, with least interference by test chemicals, and, especially where large numbers of chemicals or materials are to be tested, most easily automated endpoint measurements would be the most practical option.

An *in vitro* cytotoxicity test could be implemented in a tiered testing strategy (in the context of predicting starting doses for a subsequent *in vivo* test) in the short-term, without needing to await the outcome of formal validation activities (Section 2.5.2; see below). The main prerequisite would be the production of a guidance document, including details of test protocols considered to be appropriate, and worked examples illustrating the practical application of the strategy.

2.5.2 *Most Promising In Vitro Methods for Further Evaluation to Replace In Vivo Acute Toxicity Test Methods*

The Breakout Group did not evaluate individual test protocols or proposals as candidates for replacement of *in vivo* acute toxicity tests and therefore could not address this question directly. As noted earlier, *in vitro* tests do not currently provide all the information that can be obtained from an *in vivo* study. However, the accumulated results of many cytotoxicity studies and the ZEBET/MEIC initiatives do suggest that, in general, we may be able to obtain reasonable estimates of LD50 values if this parameter is the primary one required for regulatory decisions. Certainly by applying one or more reasonably predictive assays of the LD50 to test the considerable number of chemicals on which such risk assessment data are needed, (e.g., high production volume [HPV] chemicals), it should be possible to make a truly significant reduction in animal usage.

The Breakout Group agreed that a prevalidation study should be initiated at the earliest possible date to identify the most promising *in vitro* cytotoxicity tests for further validation. The study should include a comparison of different cell types (as a minimum, one rodent and one human cell line), exposure periods, and endpoint measurements. Regarding exposure times to evaluate, it was evident from the data available that a minimum exposure of 24 hours should be recommended (Garle et al., 1994; Hopkinson et al., 1993; Riddell et al., 1986), plus an additional "expression" period during which the previously treated cells are cultured in the absence of test material. There may be a need to evaluate several exposure times, as the most appropriate will depend on the cell type chosen, the kinetics of the test chemical, and the sensitivity of the endpoint measured (e.g., Ohno et al., 1998a).

The Breakout Group urged that a Working Group be established to follow up on its conclusions and recommendations at this Workshop (Section 2.6), and specifically, to define the details of the test protocols to be included in any prevalidation study. The selection of basal cytotoxicity tests to be included should be justified with reference to the scientific literature. It was also suggested that the statistical analyses of the MEIC program results be reviewed, so that the basis for the selection of the test battery is fully transparent.

The Breakout Group anticipates that the general performance of the assay or combination/battery of cytotoxicity assays determined from the validation study to be the

best predictor of *in vivo* lethality can be enhanced further by supplementation with other information or data. In this respect, immediate research and development needs of particular importance relate to identifying, standardizing, and validating simple predictive systems for gut absorption, BBB passage, kinetics, and metabolism. These are all important parameters which have been identified as improving the predictive ability of *in vitro* cytotoxicity data for *in vivo* LD50 values (Curren et al., 1998; Seibert et al., 1996; Ekwall et al., 1999). A new initiative on acute systemic toxicity, being undertaken as part of the TestSmart activities, has been established to address the question "can one measure cellular changes that will predict acute system failure?" The successful development of this system would complement basal cytotoxicity assays for predicting acute toxicity *in vivo* (Goldberg, personal communication).

In the longer-term, preferably undertaken as a parallel activity, the focus should be on the development and validation of human test systems for predicting human acute toxicity, integrating the approaches suggested by Breakout Groups 2 and 3. In this respect, there are numerous mechanism-based endpoints that need to be identified and evaluated in future studies.

The Breakout Group recognizes the potential impact genomics and proteomics technologies may have in many areas of toxicology, but feels these technologies could only lead to the identification of new endpoints and screening methods in the long-term, and that acute toxicity testing is not currently an area of high priority for the application of these new technologies. Investigations of changes in gene expression (e.g., using microarrays) are better targeted to more specific toxicological effects rather than general responses such as acute lethality.

2.5.3 *Ways to Evaluate the Usefulness of In Vitro Assays in an Overall Acute Toxicity Testing Strategy*

The evaluation of the usefulness of *in vitro* cytotoxicity assays in the overall testing strategy can be achieved in two ways, as indicated above. Firstly, a prospective evaluation "in practice" (in this case by implementing the use of an *in vitro* cytotoxicity test in the strategy proposed by ZEBET [Spielmann et al., 1999]) can be made once the necessary guidance document, including worked examples, has been produced. Once a sufficient body of data has been collected, the *in vitro* cytotoxicity tests can be evaluated retrospectively to determine the validity and practical usefulness of the strategy and to assess whether the predicted starting dose for an *in vivo* study is accurate for a sufficiently large enough percentage of test chemicals to continue its use.

Secondly, a formal validation activity (of which prevalidation would be an initial step; Curren et al., 1995; ICCVAM, 1997) could be conducted in which the test protocols and prediction models are evaluated independently in a multi-laboratory study involving testing of coded chemicals for the reproducibility of their responses, within and among laboratories, and the ability to predict rodent LD50 values (Balls et al., 1995a; ICCVAM, 1997).

2.6 Summary

2.6.1 *Conclusions*

The Breakout Group agreed that its primary objective was to identify and evaluate candidate *in vitro* cytotoxicity tests that could possibly serve as reduction and replacement alternatives for rodent acute oral toxicity tests for determining LD50 values. Despite the considerable research efforts by a large number of laboratories from different sectors, no standardized *in vitro* cytotoxicity assays, with optimized protocols and prediction models for the determination of LD50 values, have yet been validated. It appears from the number of studies showing positive correlations between cytotoxicity results *in vitro* and acute toxic effects *in vivo* that the application of such *in vitro* methods does have the potential to reduce and refine, and, if properly developed, ultimately replace the use of laboratory animals in acute lethality tests.

A strategy was devised by the Breakout Group that was considered to offer realistic short-term and long-term solutions to address the need for

prevalidation and validation of *in vitro* cytotoxicity tests (Figure 2.6). In the short-term, the Breakout Group concluded that the ZEBET approach (Section 2.2.1) had the potential to produce modest reductions in animal use in the ATC and UDP (OECD TG 423 and TG 425) *in vivo* tests (and in the FDP [OECD TG 420] to obviate the need for any initial sighting study). Thus, it is suggested that an *in vitro* cytotoxicity test be used in a tiered testing scheme as proposed by Spielmann et al, (1999).

The Breakout Group concluded that a guidance document with test protocol details, supporting information, and worked examples should be produced and disseminated as quickly as possible. The testing strategy should be implemented as soon as this guidance was available, without the need for a validation study. This conclusion is based on the Breakout Group's awareness of the large database on *in vitro* cytotoxicity and its demonstrated correlative power with rat acute oral LD50 values, particularly the MEIC and RC approaches. The validity of the *in vitro* cytotoxicity data in establishing appropriate starting doses for *in vivo* studies (and hence its direct predictive capability for the LD50) should be assessed retrospectively by evaluating the data generated on a sufficiently large number of substances according to pre-defined criteria for judging the acceptability of the approach. The implementation of such a testing strategy was considered to be relatively inexpensive and simple, and would not compromise the actual outcome of the *in vivo* test.

In vitro assays to replace animal tests for acute lethality will require more time to implement. The information and time available to the Breakout Group was inadequate to recommend specific cytotoxicity assays for prevalidation and validation, although the major considerations and suggestions for possible assays (e.g., a BALB/c 3T3 mouse fibroblast NRU assay) have been documented (Section 2.5.1). An additional Working Group will need to be convened for this purpose at the earliest possible date to maintain momentum and to make progress in the near term.

The scheme conceptualizing the Breakout Group's conclusions as to how cytotoxicity tests can reduce/refine and ultimately replace animal use for acute toxicity (LD50) testing (Figure 2.6) indicates what needs to be done and the projected timings for reaching that point. Each pathway involves a stepwise approach to addressing the issue. Step 1 in any testing scheme would be the collection and integration of information on the physical/chemical properties of a compound, including literature reviews and analysis of structure-activity relationships whenever possible. Most companies currently do this as a preliminary step in their evaluation of new candidate compounds for commercial development. In addition, the likelihood that acute toxicity could be metabolism-mediated needs to be considered at this early stage, and here it would be useful to integrate data derived from simple *in vitro* or *in silico* screens for biotransformation (bioactivation or detoxification). Step 2 would involve conducting an *in vitro* basal cytotoxicity test to provide data, either for correct selection of the *in vivo* starting dose (enabling an immediate reduction and refinement of animal use in the interim) or in lieu of animal testing for estimating rodent LD50 values (once the battery of *in vitro* tests required to do this had been validated for this purpose).

In the left-hand pathway in Figure 2.6, *in vivo* studies are still performed and provide supplementary information on dose response, clinical signs, and target organ effects from acute exposure for those agencies or organizations that need this additional information. However, it is anticipated that conducting a preliminary cytotoxicity test for starting dose selection would result in a modest, but cumulatively appreciable, reduction in animal numbers at minimal cost and with negligible impact on chemical or product development time. It is further projected that the ZEBET approach can be proved effective in a straightforward exercise, and Guidance for applying the approach prepared within a short period of time (i.e., 2 to 3 months).

In the right-hand pathway of Figure 2.6, the steps required for validating one or more *in vitro* cytotoxicity assays to replace animal testing for acute lethality are shown (Balls et al., 1995;

ICCVAM, 1997). This goal will take longer to achieve in light of the current state of the art. It will first be necessary to design and conduct a prevalidation study on those *in vitro* assays that are considered promising (Curren et al., 1995). Then the *in vitro* test protocol(s) and prediction models would be subjected to full validation studies to provide the necessary supporting data for assay evaluation, and eventual regulatory acceptance.

It was considered that, if the commitment to conducting a formal validation study was strong enough, the scientific resources could be harnessed for this effort with facility and the *in vitro* tests studied proved good enough, a replacement test battery might be achieved in as short a time as 2-3 years. However, past experience indicates that the formal acceptance of this battery might require substantial additional time. All prevalidation and validation studies should be conducted in compliance with the ICCVAM and ECVAM guidelines (Balls et al., 1995; ICCVAM, 1997), following the designs of similar validation studies conducted on *in vitro* tests for eye irritation (e.g., Brantom et al., 1997), skin corrosion (Fentem et al., 1998) phototoxicity (Spielmann et al., 1998), and a prevalidation study for skin irritation (Fentem et al., 2001).

In summary, it was concluded that initially a prevalidation study should be undertaken for several promising candidate *in vitro* cytotoxicity tests. Meanwhile, as a parallel activity, the generation of *in vitro* cytotoxicity data to help establish the starting dose for *in vivo* testing of new chemical substances (Spielmann et al., 1999) should be strongly encouraged as a means to potentially reduce the numbers of animals used in LD50 tests (Figure 2.6).

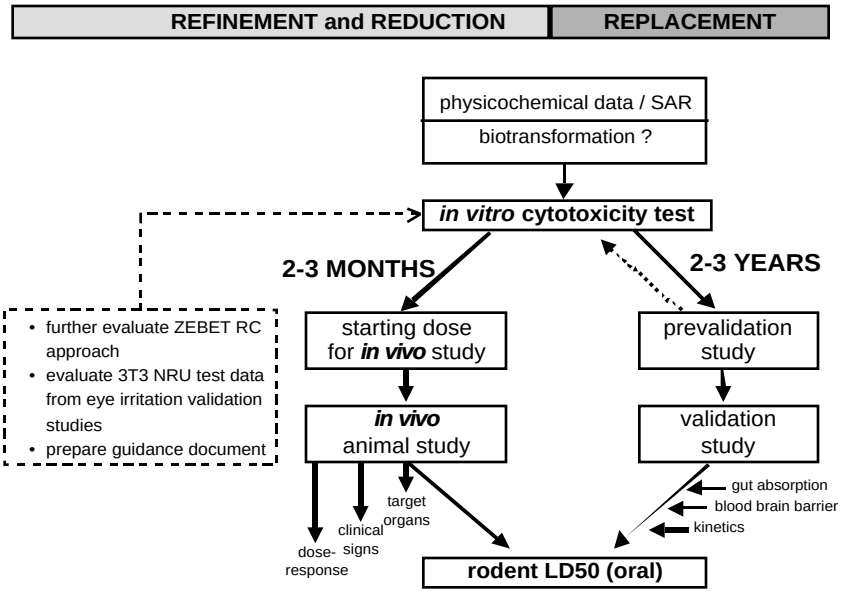

Figure 2.6. Strategy for the reduction, refinement and replacement of animals in acute LD50 testing

2.7 Recommendations

Breakout Group 1 made the following recommendations for the prevalidation, validation, and future development of *in vitro* assays for acute lethal toxicity.

2.7.1 *Short-term Activities*

- A guidance document on the application of *in vitro* cytotoxicity data for predicting *in vivo* starting doses, to include details of current test protocols considered appropriate and their application, and worked examples, should be prepared.
- A Working Group of scientific experts should be established to identify and/or define specific test protocols for inclusion in a prevalidation study. The Working Group should design and plan the study in detail. This Group should take into account the suggestions on cell type, exposure period, and endpoint measurement made by BG1 in this report.

2.7.2 *Intermediate-term Activities*

- It is anticipated that simple systems that predict gut absorption, BBB passage, key kinetic parameters, and metabolism will be needed to improve the capability of *in vitro* cytotoxicity assays to predict rodent LD50 values, or any *in vivo* toxic effects. Continued development and optimization of such systems for this application is encouraged and should receive regulatory support.
- QSAR approaches, including expert systems and neural networks, could be developed and validated as adjunct systems for predicting acute systemic toxicity. The development of commercial QSAR packages should be encouraged. As an initial step in the development of these approaches, an up-to-date review of current QSAR systems for predicting rodent oral LD50 values should be undertaken. In addition, QSARs for predicting gut absorption, metabolism, and BBB passage should be developed and evaluated.

2.7.3 *Longer-term Activities*

- The ultimate objective is the prediction of acute toxicity in humans. For this purpose, the development of simple predictive models for human acute toxicity should be a major focus.
- The evaluation and ultimate acceptance of *in vitro* assays for human acute toxicity will need a larger reference database than is presently available for validation purposes. The MEIC human database should be peer-reviewed, modified if needed, and expanded as soon as possible in order to have the data available for future validation studies.
- Other mechanism-based *in vitro* methods or endpoints, in particular resulting from the application of genomics/proteomics, may provide data that enhances the information that can be derived from cytotoxicity tests. Such research efforts should continue to be encouraged and financially supported.

2.8 References

Albert, A. 1985. Selective Toxicity, 7th edition. Chapman and Hall, London.

Archer, G., M. Balls, L.H. Bruner, R.D. Curren, J.H. Fentem, H-G. Holzhütter, M. Liebsch, D.P Lovell, and J.A. Southee. 1997. The Validation of Toxicological Prediction Models. ATLA 25: 505-516.

Balls, M., and R.H. Clothier. 1992. Cytotoxicity Assays for Intrinsic Toxicity and Irritancy. In: *In Vitro* Methods of Toxicology. (R.R. Watson, ed). CRC Press, Boca Raton, FL. pp. 37-52.

Balls, M., and J.H. Fentem. 1992. The Use of Basal Cytotoxicity and Target Organ Toxicity Tests in Hazard Identification and Risk Assessment. ATLA 20: 368-388.

Balls, M., C. Atterwill, J.Fentem, M. Garle, and F. Wiebel. 1992. Evaluation of Alternatives to Animal Tests for Assessing the Acute Lethal Potency and Neurotoxic Potential of Chemicals. A report prepared for DGXI, CEC; Contract Number B4-3081/91/8678. FRAME, Nottingham.

Balls, M., B.J. Blaauboer, J.H. Fentem, L. Bruner, R.D. Combes, B. Ekwall, R.J. Fielder, A. Guillouzo, R.W. Lewis, D.P. Lovell, C.A. Reinhardt, G. Repetto, D. Sladowski, H. Spielmann and F. Zucco. 1995a. Practical Aspects of the Validation of Toxicity Test Procedures. The report and recommendations of ECVAM Workshop 5. ATLA 23: 129-147.

Balls, M., P.A.Botham, L.H. Bruner, and H. Spielmann. 1995b. The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test. Toxicol. *In Vitro* 9: 871-929.

Barratt, M.D., J.V. Castell, M. Chamberlain, R.D. Combes, J.C. Dearden, J.H. Fentem, I. Gerner, A. Giuliani, T.J.B. Gray, D.J. Livingstone, W. McLean Provan, F.A.J.J.L. Rutten, H.J.M. Verhaar, and P. Zbinden. 1995. The Integrated Use of Alternative Approaches for Predicting Toxic Hazard: The report and recommendations of ECVAM Workshop 8. ATLA 23: 410-429.

Barratt, M.D., P.G. Brantom, J.H. Fentem, I. Gerner, A.P. Walker, and A.P. Worth. 1998. The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 1. Selection and distribution of the test chemicals. Toxicol. *In Vitro* 12: 471-482.

Barratt, M. 2000. Prediction of toxicity from chemical structure. Cell Biol. Toxicol. 16: 1-13.

Blaauboer, B.J., Balls, M., Barratt, M., Casati, S., Coecke, S, Mohamed, M.K., Moore, J., Rall, D., Smith, K.R., Tennant, R., Schwetz, B.A., Stokes, W.S., Younes, M. 1998. 13th meeting of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Alternative testing methodologies and conceptual issues. Environ. Health Persp. 106 (Suppl. 2): 413-418.

Bondesson, I., B. Ekwall, S. Hellberg, L. Romert, K. Stenberg, and E. Walum. 1989. MEIC - A New International Multicenter Project to Evaluate the Relevance to Human Toxicity of *In Vitro* Cytotoxicity Tests. Cell Biol. Toxicol. 5: 331-347.

Borenfreund, E., and J.A. Puerner. 1986. Cytotoxicity of Metals, Metal-Metal and Metal-Chelator Combinations Assayed *In Vitro.* Toxicology 39: 121-124.

Brantom, P.G., L.H. Bruner, M. Chamberlain, O. DeSilva, J. Dupuis, L.K. Earl, D.P. Lovell, W.J.W. Pape, M. Uttley, D.M. Bagley, F.W. Baker, M. Bracher, P.Courtellemont, I. Declercq, S. Freeman, W. Steiling, A.P. Walker, G.J. Carr, N. Dami, G. Thomas, J. Harbell, P.A. Jones, U. Pfannenbecker,

J.A. Southee, M. Tcheng, H. Argembeaux, D. Castelli, R. Clothier, D.J. Esdaile, H. Itigaki, K. Jung, Y. Kasai, H.Kojima, U. Kristen, M. Larnicol, R.W. Lewis, K. Marenus, O. Moreno, A. Peterson, E.S. Rasmussen, C. Robles, and M. Stern. 1997. A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test. Toxicol. *In Vitro* 11: 141-179.

Bruce R.D. 1985. An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Tox. 5: 151-157.

Bruner, L.H., G.J. Carr, M. Chamberlain, and R.D. Curren. 1996. Validation of Alternative Methods for Toxicity Testing. Toxicol. *In Vitro* 10: 479-501.

Clemedson, C., and B. Ekwall. 1999. Overview of the Final MEIC Results: I. The *In Vitro-In Vivo* Evaluation. Toxicol. *In Vitro* 13: 1-7.

Clemedson, C., E. McFarlane-Abdulla, M. Andersson, F.A. Barile, M.C. Calleja, C. Chesné, R. Clothier, M. Cottin, R. Curren, E. Daniel-Szolgay, P. Dierickx, M. Ferro, G. Fiskesjö, L. Garza-Ocanas, M.J. Gómez-Lechón, M. Gülden, B. Isomaa, J. Janus, P. Judge, A. Kahru, R.B. Kemp, G. Kerszman, U. Kristen, M. Kunimoto, S. Kärenlampi, K. Lavrijsen, L. Lewan, H. Lilius, T. Ohno, G. Persoone, R. Roguet, L. Romert, T. Sawyer, H. Seibert, R. Shrivastava, A. Stammati, N. Tanaka, O. Torres Alanis, J.-U. Voss, S. Wakuri, E. Walum, X. Wang, F. Zucco, and B. Ekwall. 1996a. MEIC Evaluation of Acute Systemic Toxicity. Part I. Methodology of 68 *In vitro* toxicity assays used to test the first 30 reference chemicals. ATLA 24 (Suppl. 1): 249-272.

Clemedson, C, E. McFarlane-Abdulla, M. Andersson, F.A. Barile, M.C. Calleja, C. Chesné, R. Clothier, M. Cottin, R. Curren, P. Dierickx, M. Ferro, G. Fiskesjö, L. Garza-Ocanas, M.J. Gómez-Lechón, M. Gülden, B. Isomaa, J. Janus, P. Judge, A. Kahru, R.B. Kemp, G. Kerszman, U. Kristen, M. Kunimoto, S. Kärenlampi, K. Lavrijsen, L. Lewan, H. Lilius, A. Malmsten, T. Ohno, G. Persoone, R. Pettersson, R. Roguet, L. Romert, M. Sandberg, T. Sawyer, H. Seibert, R. Shrivastava, M. Sjöström, A. Stammati, N. Tanaka, O. Torres Alanis, J.-U. Voss, S. Wakuri, E. Walum, X. Wang, F. Zucco, and B. Ekwall. 1996b. MEIC Evaluation of Acute Systemic Toxicity. Part II. *In vitro* results from 68 toxicity assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. ATLA 24 (Suppl. 1): 273-311.

Clemedson, C., F.A. Barile, B. Ekwall, M.J. Gómez-Lechón, T. Hall, K. Imai, A. Kahru, P. Logemann, F. Monaco, T. Ohno, H. Segner, M. Sjöström, M. Valentino, E. Walum, X. Wang, and B. Ekwall. 1998a. MEIC Evaluation of Acute Systemic Toxicity. Part III. *In vitro* results from 16 additional methods used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. ATLA 26 (Suppl. 1): 91-129.

Clemedson, C., Y. Aoki, M. Andersson, F.A. Barile, A.M. Bassi, M.C. Calleja, A. Castano, R.H. Clothier, P. Dierickx, B. Ekwall, M. Ferro, G. Fiskesjö, L. Garza-Ocanas, M.J. Gómez-Lechón, M. Gülden, T. Hall, K. Imai, B. Isomaa, A. Kahru, G. Kerszman, P. Kjellstrand, U. Kristen, M. Kunimoto, S. Kärenlampi, L. Lewan, H. Lilius, A. Loukianov, F. Monaco, T. Ohno, G. Persoone, L. Romert, T.W. Sawyer, R. Shrivastava, H. Segner, H. Seibert, M. Sjöström, A. Stammati, N. Tanaka, A. Thuvander, O. Torres-Alanis, M. Valentino, S. Wakuri, E. Walum, A. Wieslander, X. Wang, F. Zucco, and B. Ekwall. 1998b. MEIC Evaluation of Acute Systemic Toxicity. Part IV. *In vitro* results from 67 toxicity assays used to test reference chemicals 31-50 and a comparative cytotoxicity analysis. ATLA 26 **(**Suppl. 1): 131-183.

Clemedson, C., F.A. Barile, C. Chesné, M. Cottin, R. Curren, Ba. Ekwall, M. Ferro, M.J. Gomez-Lechon, K. Imai, J. Janus, R.B. Kemp, G. Kerszman, P. Kjellstrand, K. Lavrijsen, P. Logemann, E. McFarlane-Abdulla, R. Roguet, H. Segner, H. Seibert, A. Thuvander, E. Walum, and Bj. Ekwall. 2000. MEIC Evaluation of Acute Systemic Toxicity. Part VII. Prediction of human toxicity by

results from testing of the first 30 reference chemicals with 27 further *in vitro* assays. ATLA 28 (Suppl. 1): 161-200.

Clothier, R.H., L.M.Hulme, M. Smith, and M. Balls. 1987. Comparison of the *In Vitro* Cytotoxicities and Acute *In Vivo* Toxicities of 59 Chemicals. Molecular Toxicol. 1: 571-577.

Clothier, R.H., L.M.Hulme, A.B. Ahmed, H.L. Reeves, M. Smith, and M. Balls. 1988. *In Vitro* Cytotoxicity of 150 Chemicals to 3T3-L1 Cells Assessed by the FRAME Kenacid Blue Method. ATLA 16: 84-95.

Curren, R.D., J.A. Southee, H. Spielmann, M. Liebsch, J.H. Fentem, and M. Balls. 1995. The Role of Prevalidation in the Development, Validation and Acceptance of Alternative Methods. ATLA 23: 211-217.

Curren, R., L. Bruner, A. Goldberg, and E. Walum. 1998. 13th meeting of the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Validation and acute toxicity testing. Environ. Hlth Perspect. 106 (Suppl. 2): 419-425.

Dearden, J.C., M.D. Barratt, R. Benigni, D.W. Bristol, R.D. Combes, M.T.D.Cronin, P.N. Judson, M.P. Payne, A.M. Richard, M. Tichy, A.P. Worth, and J.J. Yourick. 1997. The Development and Validation of Expert Systems for Predicting Toxicity. The report and recommendations of an ECVAM/ECB workshop (ECVAM workshop 24). ATLA 25: 223-252.

Diener, W., U. Mischke, D. Kayser, and E. Schlede . 1995. The Biometric Evaluation of the OECD Modified Version of the Acute-Toxic-Class Method (Oral). Arch. Toxicol. 69: 729-734.

Diener W. and E. Schlede. 1999. Acute toxic class methods: alternatives to LD/LC_{50} tests. ALTEX 16: 129-134.

Diener, W., and E. Schlede. 2000. Acute Toxic Class Methods: Biometric Evaluations and Test Procedures for the New International Classification Systems. Manuscript in preparation.

Dierickx, P.J. 1989. Cytotoxicity Testing of 114 Compounds by the Determination of Protein Content in HepG2 Cell Cultures. Toxicol. *In Vitro* 3: 189-193.

Dixon, W.J., and A.M. Mood. 1948. A Method for Obtaining and Analyzing Sensitivity Data. J. Amer. Statist. Assoc. 43: 109-126.

Dixon, W.J. 1965. The Up-and-Down Method for Small Samples. J. Amer. Statist. Assoc. 60: 967-978.

Dixon, W.J. 1991. Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev. 15: 47-50.

Dixon, W.J. 1991. Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.

Eagle, H., and G.E. Foley. 1956. The Cytotoxic Action of Carcinolytic Agents in Tissue Culture. Amer. Jour. of Med. 21: 739-745.

Ekwall, B. 1983. Screening of Toxic Compounds in Mammalian Cell Cultures. Ann. New York Acad. Sci. 407: 64-77.

Ekwall, B. 1999. Overview of the Final MEIC Results: II. The *In Vitro/In Vivo* Evaluation, Including the Selection of a Practical Battery of Cell Tests for Prediction of Acute Lethal Blood Concentrations in Humans. Toxicol. *In Vitro* 13: 665-673.

Ekwall, B., C. Clemedson, B. Crafoord, Ba. Ekwall, S. Hallander, E. Walum, and I. Bondesson. 1998a. MEIC Evaluation of Acute Systemic Toxicity. Part V. Rodent and Human Toxicity Data for the 50 Reference Chemicals. ATLA 26 (Suppl. 2): 569-615.

Ekwall, B., F.A. Barile, A. Castano, C. Clemedson, R.H. Clothier, P. Dierickx, Ba. Ekwall, M. Ferro, G. Fiskesjö, L. Garza-Ocanas, M.J. Gómez-Lechón, M. Gülden, T. Hall, B. Isomaa, A. Kahru, G. Kerszman, U. Kristen, M. Kunimoto, S. Kärenlampi, L. Lewan, A. Loukianov, T. Ohno, G. Persoone, L. Romert, T.W. Sawyer, H. Segner, R. Shrivastava, A. Stammati, N. Tanaka, M. Valentino, E. Walum, and F. Zucco. 1998b. MEIC Evaluation of Acute Systemic Toxicity. Part VI. Prediction of human toxicity by rodent LD50 values and results from 61 *In vitro* tests. ATLA 26 (Suppl. 2): 617-658.

Ekwall, B., C. Clemedson, B. Ekwall, P. Ring, and L. Romert. 1999. EDIT: A new international multicentre programme to develop and evaluate batteries on *in vitro* tests for acute and chronic systemic toxicity. ATLA 27: 339-349.

Ekwall, B., B. Ekwall, and M. Sjostrom. 2000. MEIC Evaluation of Acute Systemic Toxicity. Part VIII. Multivariate partial least squares evaluation, including the selection of a battery cell line tests with a good prediction of human acute lethal peak blood concentrations for 50 chemicals. ATLA 28 (Suppl. 1): 201-234.

Fentem, J., J. Fry, M. Garle, M. Gülden, H. Seibert, J-U. Voss, O. Wassermann, M. Perchermeier, and F.Wiebel. 1993. An International Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic Toxicity. A report prepared for DGXI, CEC; Contract Numbers B92/B4-3063/14086 & B92/B4-3040/14087. FRAME, Nottingham.

Fentem, J.H., G.E.B. Archer, M. Balls, P.A. Botham, R.D. Curren, L.K.Earl, D.J.Esdaile, H-G. Holzhütter, and M. Liebsch. 1998. The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 2. Results and evaluation by the Management Team. Toxicology *In Vitro* 12: 483-524.

Fentem, J.H., D. Briggs, C.Chesné, G.R. Elliott, J.W. Harbell, J.R. Heylings, P. Portes, R. Roguet, J.J.M. van de Sandt, and P.A.Botham. 2001. A Prevalidation Study on *In Vitro* Tests for Acute Skin Irritation: results and evaluation by the Management Team. Toxicology *In vitro* 15:57-93.

Fry, J.R., M.J. Garle, and A.H. Hammond. 1988. Choice of Acute Toxicity Measures for Comparison of *In Vivo-In Vitro* Toxicity. ATLA 16: 175-179.

Fry, J.R., M.J. Garle, A.H. Hammond, and A. Hatfield. 1990. Correlation of Acute Lethal Potency with *In Vitro* Cytotoxicity. Toxicol. *In Vitro* 4: 175-178.

Galson, S. 2000. Historical and current regulatory perspectives. Opening Plenary Session, ICCVAM International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, October 17-20, 2000.

Garle, M.J., A.H. Hammond, and J. R. Fry. 1987. The Cytotoxicity of 27 Chemicals to V79 Chinese Hamster Cells. ATLA 15: 30-32.

Garle, M., J.H. Fentem, and J.R. Fry. 1994. *In Vitro* Cytotoxicity Tests for the Prediction of Acute Toxicity *In Vivo*. Toxicol. *In Vitro* 8: 1303-1312.

Grisham, J.W., and G.J. Smith. 1984. Predictive and Mechanistic Evaluation of Toxic Responses in Mammalian Cell Culture Systems. Pharmacolog. Rev. 36 (Suppl.): 151S-171S.

Gülden, M., H. Seibert, and J.-U. Voss. 1994. Inclusion of Physicochemical Data in Quantitative Comparisons of *In Vitro* and *In Vivo* Toxic Potencies. ATLA 22: 185-192.

Guzzie, P.J. 1994. Lethality Testing. In: *In Vitro Toxicology*. (S.C. Gad, ed). Raven Press, New York. pp. 57-86.

Halle, W., and E. Goeres. 1988. Register der Zytotoxizität (IC50) in der Zellkultur und Möglichkeiten zur Abschätzung der akuten Toxizität (LD50). In: Beiträge zur Wirkstoffforschung, Institute für Wirkstoffforschung. (Oehme, P., H. Loewe, and E. Goeres, eds). Berlin, Germany.

Halle, W., and H. Spielmann. 1992. Two Procedures for the Prediction of Acute Toxicity (LD50) from Cytotoxicity Data. ATLA 20: 40-49.

Halle, W., M. Liebsch, D. Traue, and H. Spielmann. 1997. Reduktion der Tierzahlen bei der Einstufung von Stoffen in die EU-Toxizitätsklassen für akute orale Toxizität mit Hilfe von Daten aus dem Register der Zytotoxizität (RC). ALTEX 14: 8-15.

Halle, W. 1998. Toxizitätsprüfungen in Zellkulturen für eine Vorhersage der akuten Toxizität (LD50) zur Einsparung von Tierversuchen. Life Sciences/ Lebens-wissenschaften, Volume 1, 94 pp. Jülich: Forschungszentrum Jülich.

Halle, W., H. Spielmann, and M. Liebsch. 2000. Prediction of Human Lethal Concentrations by Cytotoxicity Data from 50 MEIC Chemicals. ALTEX 17: 75-79.

Hopkinson, D., R. Bourne, and F.A. Barile. 1993. *In vitro* Cytotoxicity Testing: 24-hour and 72-hour studies with cultured lung cells. ATLA 21: 167-172.

Hulme, L.M., H.L. Reeves, R.H. Clothier, M. Smith, and M. Balls. 1987. Assessment of Two Alternative Methods for Predicting the *In Vivo* Toxicities of Metallic Compounds. Molecular Toxicol. 1: 589-596.

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

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). 1999. Evaluation of the validation status of toxicological methods: General Guidelines for Submissions to ICCVAM Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication 99-4496. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. Available on the Internet at http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm.

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). 2000. The revised Up-and-Down Procedure: a test for determining the acute oral toxicity of chemicals and products. Proposed test method and background review document, April 14, 2000. ICCVAM / NIEHS, Research Triangle Park, NC 27709. Available on the Internet at http://iccvam.niehs.nih.gov/methods/udpdocs/AllBRDlk.pdf.

Itagaki, H., T.Ohno, M. Hatao et al. 1998a. Validation Study on Five Cytotoxicity Assays by JSAAE. V. Details of the crystal violet staining assay. AATEX 5: 87-98.

Itagaki, H., T.Ohno, M. Hatao et al. 1998b. Validation Study on Five Cytotoxicity Assays by JSAAE. VII. Details of the MTT assay. AATEX 5: 119-130.

Klaassen, C.D., and D.L. Eaton. 1991a. Principles of Toxicology. In: Caserett and Doull's Toxicology: The Science of Poisons. 4th Ed (Amdur, M.O., J. Doull, and C.D. Klaassen, eds). Pergamon Press Inc., New York, pp. 16-17.

Klaassen, C.D., and D.L. Eaton. 1991b. Principles of Toxicology. In: Caserett and Doull's Toxicology: The Science of Poisons. 4th Ed (Amdur, M.O., J. Doull, and C.D. Klaassen, eds). Pergamon Press Inc., New York, pp. 22.

Knox, P., P.F. Uphill, J.R. Fry, J. Benford, and M. Balls. 1986. The FRAME Multicentre Project on *In Vitro* Cytotoxicity. Food Chem. Toxicol. 24: 457-463.

Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. 1995. Comparison of the Up-and-Down, Conventional LD50, and Fixed-Dose Acute Toxicity Procedures. Food Chem. Toxicol. 33: 223-231.

Lipnick, R.L., M. Zeeman, and J.A. Cotruvo. 1995b. Structure-activity relationships in the validation of in vitro toxicology tests. In: *Animal Test Alternatives: Refinement Reduction Replacement* . H. Salem, ed., pp 47-55. Marcel Dekker, New York.

MEIC, MEMO, and EDIT information: http://www.cctoxconsulting.a.se/nica.htm

NIEHS (National Institute of Environmental Health Sciences), 2001. Guidance document on using in vitro data to estimate in vivo starting doses for acute toxicity. NIH Publication 01-4500. NIEHS, Research Triangle Park, North Carolina.

OECD. October, 2000; Draft OECD Guideline for Testing of Chemicals, 423: Acute Oral toxicity - Acute Toxic Class Method; http://www.oecd.org/ehs/test/health.htm.

OECD. 1998a. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as Endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11. OECD, Paris. http://www.oecd.org/ehs/class/HCL6htm

OECD. 1998b. OECD Guideline for Testing of Chemicals. 425. Acute Oral Toxicity - Up-and-Down Procedure. OECD. Paris.

OECD. 1981. OECD Guideline for Testing of Chemicals. 401. Acute Oral Toxicity. OECD, Paris. [updated in 1987]

OECD. 1987. OECD Guideline for Testing of Chemicals. 401. Acute Oral Toxicity. OECD, Paris.

OECD. 1992. OECD Guideline for Testing of Chemicals. 420. Acute Oral Toxicity - Fixed Dose Method. OECD, Paris.

OECD. 1996. OECD Guideline for Testing of Chemicals. 423. Acute Oral Toxicity - Acute Toxic Class Method. OECD, Paris.

Ohno, T., M. Asakura, T. Awogi et al. 1998a. Validation Study on Five Cytotoxicity Assays by JSAAE. I. Overview of the study and analyses of variations of ED50 values. AATEX 5: 1-38.

Ohno, T., Y. Futamura, A. Harihara et al. 1998b. Validation Study on Five Cytotoxicity Assays by JSAAE. VI. Details of the LDH release assay. AATEX 5: 99-118.

Ohno, T., Y. Futamura, A. Harihara et al. 1998c. Validation Study on Five Cytotoxicity Assays by JSAAE. VIII. Details of the Neutral Red Uptake assay. AATEX 5: 131-145.

Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes. 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul. Toxicol. Pharm. 32: 56-67.

Phillips, J.C., W.B. Gibson, J. Yam, C.L. Alden, and G.C. Hard. 1990. Survey of the QSAR and *In Vitro* Approaches for Developing Non-Animal Methods to Supersede the *In Vivo* LD50 Test. Food Chem. Toxicol. 28: 375-394.

Pomerat, C., and C.D. Lake. 1954. Short Term Cultures for Drug Assays: general considerations. Ann. New York Acad. Sci. 58: 1110-1128.

Riddell, R.J., D.S.Panacer, S.M. Wilde, R.H. Clothier, and M. Balls. 1986. The Importance of Exposure Period and Cell Type in *In Vitro* Cytotoxicity Tests. ATLA 14: 86-92.

Schlede, E., U. Mischke, R. Roll, and D. Kayser. 1992. A National Validation Study of the Acute-Toxic Class Method - An Alternative to the LD50 Test. Arch. Toxicol. 66: 455-470.
Schlede, E., U. Mischke, W. Diener, and D. Kayser. 1994. The International Validation Study of the Acute-Toxic Class Method (Oral). Arch. Toxicol. 69: 659-670.

Seibert, H., M. Balls, J.H. Fentem, V. Bianchi, R.H. Clothier, P.J. Dierickx, B. Ekwall, M.J. Garle, M.J. Gómez-Lechón, L. Gribaldo, M. Gülden, M. Liebsch, E. Rasmussen, R. Roguet, R. Shrivastava, and E. Walum. 1996. Acute Toxicity Testing *In Vitro* and the Classification and Labeling of Chemicals: The report and recommendations of ECVAM Workshop 16. ATLA 24: 499-510.

Smith, C.G., J.E. Grady, and J.I. Northam. 1963. Relationship between Cytotoxicity *In Vitro* and Whole Animal Toxicity. Cancer Chemother. Rep. 30: 9-12.

Spielmann, H., M. Liebsch, and S. Kalweits. 1996. Results of a Validation Study in Germany on Two *In Vitro* Alternatives to the Draize Eye Irritation Test, the HET-CAM Test and the 3T3-NRU Cytotoxicity Test. ATLA 24: 741-858.

Spielmann, H., M. Balls, J. Dupuis, W.J.W. Pape, G. Pechovitch, O. de Silva, H-G. Holzhütter, R. Clothier, P. Desolle, F. Gerberick, M. Liebsch, W.W. Lovell, T. Maurer, U. Pfannenbecker, J.M. Potthast, M. Csato, D. Sladowski, W. Steiling, and P. Brantom. 1998. EU/COLIPA "*In Vitro* Phototoxicity" Validation Study, Results of Phase II (blind trial), Part 1: the 3T3 NRU phototoxicity test. Toxicol. *In Vitro* 12: 305-327.

Spielmann, H., E. Genschow, M. Leibsch, and W. Halle. 1999. Determination of the Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data. ATLA 27: 957-966.

Tanaka, N. M. Asakura, C. Hattori et al. 1998. Validation Study on Five Cytotoxicity Assays by JSAAE. IV. Details of the colony formation assay. AATEX 5: 74-86.

Trevan, J.W. 1927. The Error of Determination of Toxicity. Proceedings of the Royal Society (London). Series B 101: 483-514.

Wakuri, S., J. Izumi, K. Sasaki, N. Tanaka, and H. Ono. 1993. Cytotoxicity Study of 32 MEIC Chemicals by Colony Formation and ATP Assays. Toxicol. *In Vitro* 7: 517-521.

Worth, A.P., M.D. Barratt, and J.B. Houston. 1998. The Validation of Computational Prediction Techniques. ATLA 26: 241-247.

Zanetti, C., I. De Angelis, A-L. Stammati, and F. Zucco. 1992. Evaluation of Toxicity Testing of 20 MEIC Chemicals on Hep-2 Cells Using Two Viability Endpoints. ATLA 20: 120-125.

Zbinden, G., and M. Flury-Roversi. 1981. Significance of the LD50 Ttest for the Toxicological Evaluation of Chemical Substances. Arch. of Toxicol. 47: 77-99.

ADDENDUM

Combined analyses of the ZEBET Register of Cytotoxicity (RC) and MEIC data

The predictions of acute lethality *in vivo* from the RC and MEIC cytotoxicity data have been analyzed. The correlation for the 50 MEIC chemicals (IC50 *in vitro* vs rodent oral LD50 *in vivo*), including the RC cytotoxicity data for various mammalian cell lines (dark triangles, dark linear regression line) and the MEIC program cytotoxicity data for various human cell lines (circles, gray linear regression line; taken from Clemedson et al., 1998a; Clemedson et al., 1998b), are shown in Figure A.1. Similar standard regression lines, with comparable data fits, were obtained for the RC values (mean IC50*x* data) and the MEIC values (IC50m) for the 50 chemicals (Table A.1).

A similar comparison of the correlations for the 50 MEIC chemicals (RC mammalian *in vitro* values and MEIC human *in vitro* values from Clemedson et al. [1998a; 1998b]) was also undertaken for *in vitro* IC50 vs human peak lethal blood concentrations *in vivo* (Ekwall et al., 1998a). Again, similar standard regression lines, with comparable fits, were obtained (Table A.1):

RC: log (peak concentration) =
0.822 x log (IC50*x*) - 0.437; r=0.81; R^2=0.66

MEIC: log (peak concentration) =
0.913 x log (IC50*m*) - 0.702; r=0.86; R^2=0.74

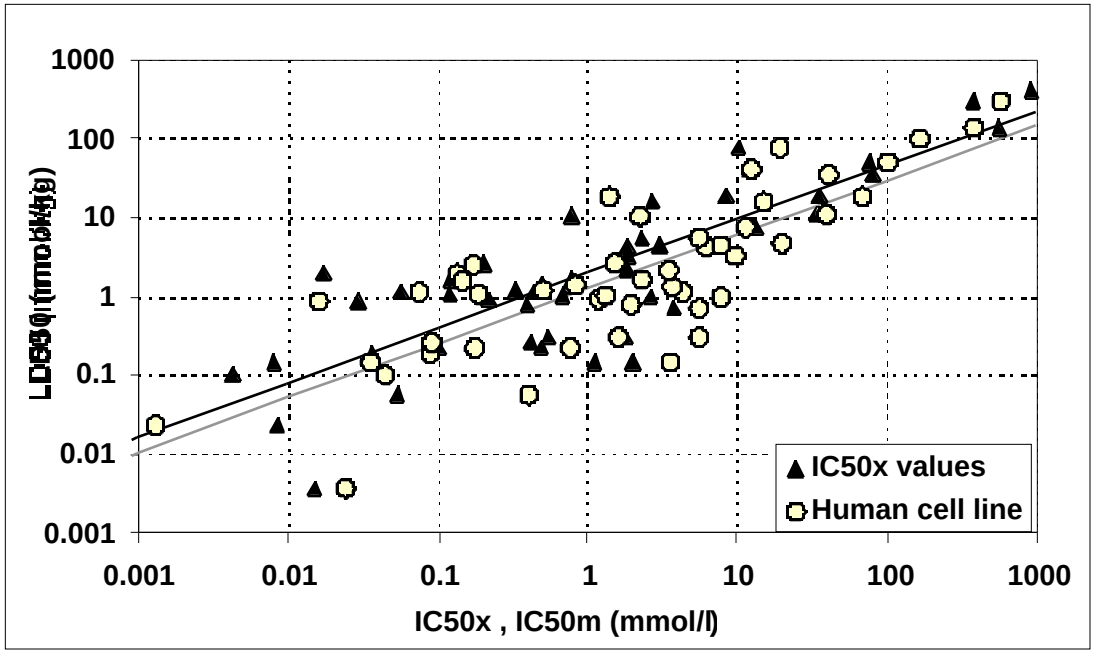

Figure A.1. Regression between Cytotoxicity (IC50) and rodent acute oral LD50 for the 50 MEIC chemicals
RC: log (LD50) = 0.689 x log (IC50*x*) + 0.276; r=0.84; R^2=0.71
MEIC: log (LD50) = 0.690 x log (IC50*m*) + 0.080; r=0.81; R^2=0.66

Table A.1. Summary of linear regression analyses (RC vs MEIC)

Chemicals	x	y	slope	constant	r	R^2
347 non-selected (RC)	IC50*x*	LD50	0.435	0.625	0.67	0.45
50 MEIC (RC)	IC50*x*	LD50	0.689	0.276	0.84	0.71
50 MEIC (human cell lines)	IC50*m*	LD50	0.690	0.080	0.81	0.66
50 MEIC (RC)	IC50*x*	human lethal	0.822	- 0.437	0.81	0.66
50 MEIC (human cell lines)	IC50*m*	human lethal	0.913	- 0.702	0.86	0.74
50 MEIC	LD50	human lethal	0.879	- 0.669	0.71	0.50

To set these results in context, the predictivity of the rat LD50 for human peak lethal concentration was assessed for the MEIC chemicals (Figure A.2; Table A.1). The correlation was not as good as that found with the IC50 values.

The 50 MEIC chemicals are a subset of the RC; the overall predictivity of the entire RC (347 chemicals) for rodent LD50 values is lower than that of the 50 MEIC chemicals (Figure A.3; Table A.1). The relationship between *in vitro* IC50 values and *in vivo* LD50 values should be investigated further by employing multiple regression techniques rather than simple linear regression. In addition, cluster analysis could also be undertaken.

To investigate how basal cytotoxicity data obtained from various human cell lines (IC50m) in the MEIC program (part III and IV) compares with basal cytotoxicity data from various mammalian cell lines (IC50*x*), the correlation between IC50*x* and IC50m is shown in Figure A.4. The correlation is judged very high by R^2 = 0.90, and suggests that basal cytotoxicity data obtained with either human cells or other mammalian cells may be similar and equivalent for the prediction of *in vivo* lethality measures.

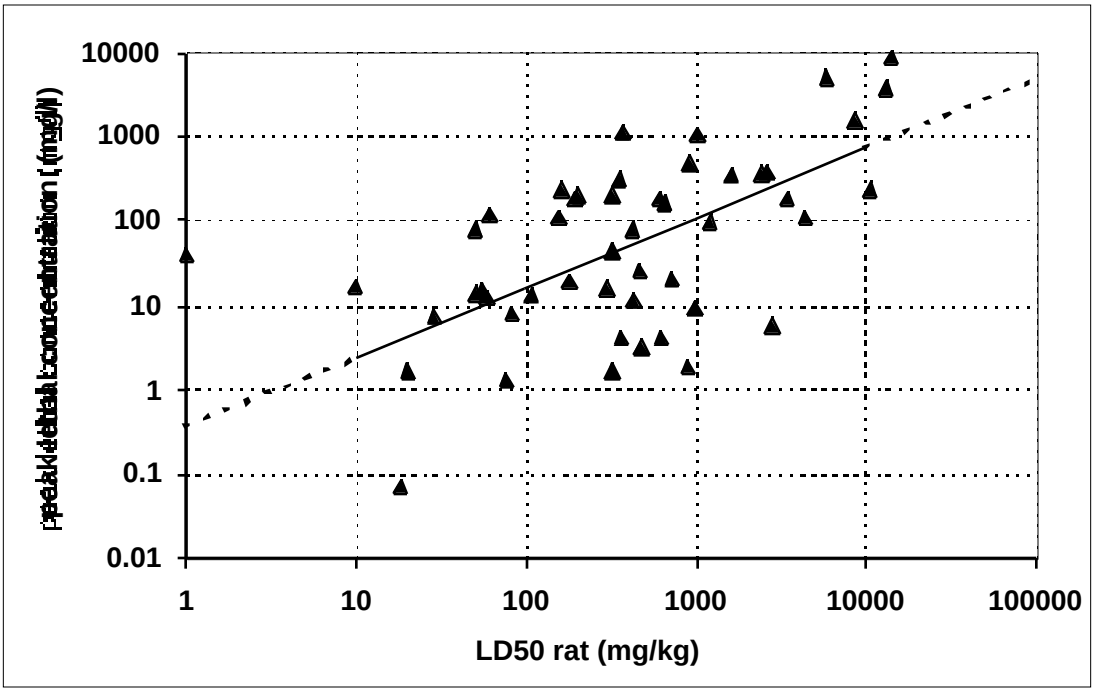

Figure A.2. Regression between rodent acute oral LD50 values and human peak lethal concentrations for the 50 MEIC chemicals.
Regression equation: log (peak conc.) = 0.879 x log (LD50) – 0.669; r=0.71; R^2=0.50.

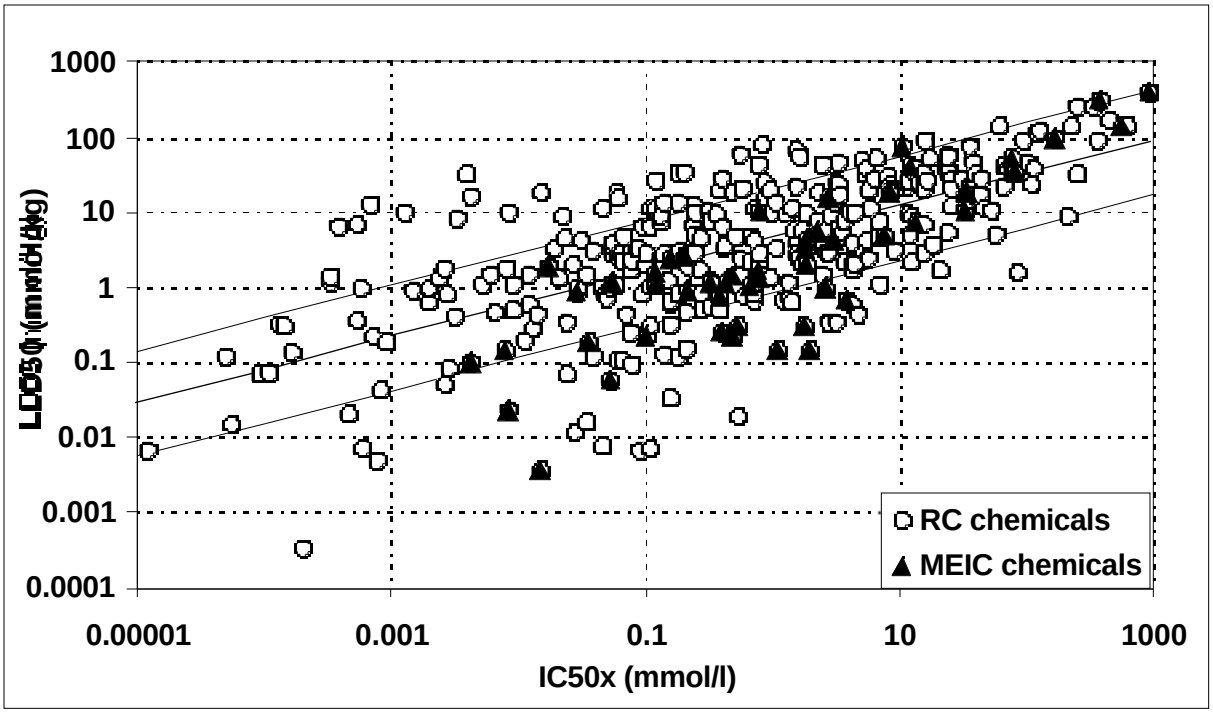

Figure A.3. Regression between Cytotoxicity (IC50) and rodent acute oral LD50 values for the RC database showing the 50 MEIC chemicals as a subset of the 347 chemicals in the RC

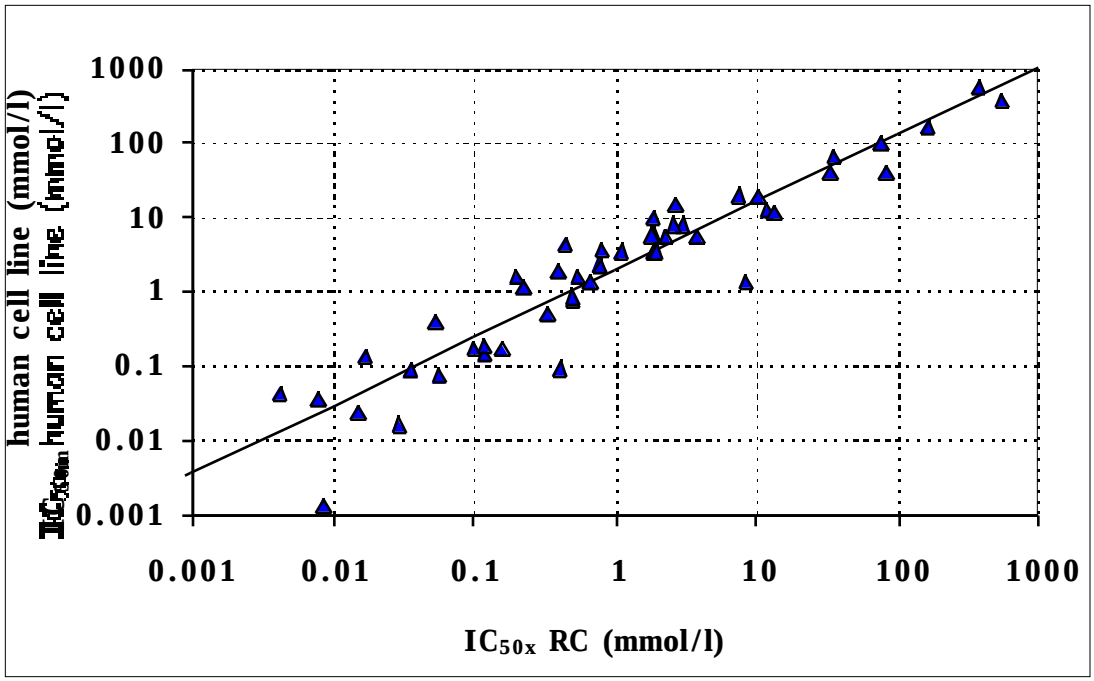

Figure A.4. Correlation between IC50*x* (averaged from various mammalian cell lines) of the RC and IC50m (from various human cell lines) is shown for the 50 MEIC chemicals
The linear correlation coefficient is high (r = 0.95) and judged by an R^2 = 0.90.