[This Page Intentionally Left Blank]

1.0 INTRODUCTION AND RATIONALE FOR THE USE OF *IN VITRO* NEUTRAL RED UPTAKE CYTOTOXICITY TEST METHODS TO PREDICT STARTING DOSES FOR *IN VIVO* ACUTE ORAL TOXICITY TESTING

Poisoning is a more serious public health problem than generally recognized. The Institute of Medicine (IOM) estimates that more than 4 million poisoning episodes occur annually in the United States (IOM 2004). In 2001, poisoning (30,800 deaths) placed second behind automobile accidents (42,433 deaths) as the leading cause of injury-related death (IOM 2004). To reduce the risk for accidental poisonings, various regulatory agencies in the United States (e.g., the Environmental Protection Agency [EPA], the Consumer Products Safety Commission [CPSC]), require the testing of marketed products for acute oral toxicity in rodents. Increasing societal concerns about animal use have lead to the development and evaluation of alternative *in vitro* test methods that might refine, reduce, or replace acute oral toxicity test methods[1].

The purpose of this background review document (BRD) is to:

- Describe a validation study organized and managed by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM) to evaluate the ability of two *in vitro* basal cytotoxicity test methods to predict starting doses for rodent acute oral toxicity tests
- Provide the results of an evaluation of the accuracy and reliability of the two *in vitro* basal cytotoxicity test methods, as well as of the animal savings that would occur if these test methods were used to predict the starting dose.

The structure of the BRD follows the requested structure of the *ICCVAM*[2] *Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods* (ICCVAM 2003).

This section provides:

- A historical perspective of scientific efforts to develop and evaluate the ability of *in vitro* cytotoxicity test methods to refine, reduce, or replace acute oral toxicity test methods
- A general review of reported correlations between *in vitro* cytotoxicity and acute oral lethality in rodents
- The regulatory requirements for rodent acute oral toxicity testing
- The scientific basis of using *in vitro* basal cytotoxicity test methods to predict the starting doses for rodent acute oral toxicity assays

---

[1] A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

[2] The Interagency Coordinating Committee for the Validation of Alternative Methods

- The intended regulatory uses and applicability of *in vitro* basal cytotoxicity test methods

1.1 Historical Background and Rationale for the Use of *In Vitro* Cytotoxicity Assays to Predict Starting Doses for Rodent Acute Oral Toxicity Tests

This section provides the historical background and rationale for the NICEATM/ECVAM validation study by summarizing several major studies promoted by the European Union (EU) to investigate the properties and capabilities of cell-based methods to predict acute toxicity. The Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) Program was initiated in 1983 to compare *in vitro* methods to acute oral lethality in humans (**Section 1.1.1**). In 1992-1993, the Fund for the Replacement of Animals in Medical Experiments (FRAME) conducted an international evaluation of selected *in vitro* toxicity test systems for predicting acute systemic toxicity (**Section 1.1.2**). Dr. Willi Halle published a monograph regarding the development of the Registry of Cytotoxicity (RC) database to evaluate whether basal cytotoxicity data could accurately predict acute oral lethality in rats and mice (**Section 1.1.3**). ECVAM organized a workshop in 1994 to evaluate the use of *in vitro* data for the classification and labeling of chemicals and reviewed the assessment of acute oral toxicity using *in vitro* data. Workshop participants suggested that the use *in vitro* data to determine starting doses for acute oral toxicity tests would reduce the use of animals. The German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) then recommended that *in vitro* basal cytotoxicity data be used with the RC millimole regression, which is referred to as the ZEBET approach (**Section 1.1.4**), to determine starting doses for acute oral toxicity tests. **Section 1.1.5** provides background on an international workshop that reviewed and evaluated the EU studies above and **Section 1.1.6** describes the NICEATM/ECVAM *in vitro* cytotoxicity validation study that expands upon the EU studies.

1.1.1 The Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) Program

The Scandinavian Society for Cell Toxicology established the MEIC program in 1983 to investigate the ability of *in vitro* cytotoxicity test methods to predict acute oral lethality in humans (Bondesson et al. 1989). MEIC was based on the following assumptions:

- *In vitro* cell culture systems could be used to model *in vivo* acute oral toxicity.
- The basal cytotoxicity detected by these *in vitro* test methods is responsible for a large proportion of *in vivo* toxic effects[3].

The MEIC program was an open study that invited laboratories worldwide to participate in testing 50 reference substances using laboratory-specific *in vitro* cytotoxicity assays. Although the MEIC management team requested that all participating laboratories test chemicals with high purity, no effort was made to assure that the substances tested were purchased from the same supplier or were of the same purity (Clemedson et al. 1996a). Minimal methodological directives were provided so as to maximize protocol diversity among the 96 participating laboratories.

[3] Basal, or general, cytotoxicity was described as toxicity resulting from interference with basic cellular structures and functions, such as cell membranes, metabolism, ion regulation, and cell division that are common to all human and animal cells.

The reference substances were selected to represent different chemical classes for which reference acute oral lethality data existed in humans (i.e., lethal doses, kinetics, and lethal blood/serum concentrations [LC]) and rodents (oral median lethal dose [LD_{50}] values) (Bondesson et al. 1989). The MEIC management team collected human data from clinical and forensic toxicology handbooks and case reports of human poisonings (Ekwall et al. 1998a). The resulting data were presented and analyzed in a series of 50 MEIC Monographs. Rat and mouse oral LD_{50} data were collected from the Registry of Toxic Effects for Chemical Substances (RTECS®)[4].

The 50 reference substances were tested in as many as 61 different *in vitro* assays (Ekwall et al. 1998b). The metric of interest was the IC_{50} (i.e., the concentration that inhibited the response measured by 50%) for the endpoint measured. Of the 20 test methods that used human-derived cells, 18 used cell lines and two used primary cell cultures. Of the 21 test methods that used mammalian (but other than human) cells, 12 used cell lines and nine used primary cell cultures. Eighteen test methods were ecotoxicological in nature and two used cell-free systems. Cell viability and/or cell growth were the endpoints of choice in the majority of the cell-based systems. The chemical exposure duration ranged from 5 minutes to 6 weeks, but most frequently was 24 hours (Clemedson et al. 1996).

The ability of the *in vitro* IC_{50} data to predict human acute oral lethality was assessed using human LC values compiled from three different data sets (Ekwall et al. 2000):

- Clinically measured acute lethal serum concentrations
- Acute LC values measured post-mortem
- Peak LC values derived from approximate LC_{50} curves over time after exposure

A partial least squares (PLS) analysis indicated that the IC_{50} data generated from as many as 61 test methods predicted the three sets of LC data 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). A two component PLS model using rat and mouse oral LD_{50} values less accurately predicted human LC values (R^2=0.65, Q^2=0.64). These results suggested that *in vitro* basal cytotoxicity assays might be more effective in estimating human acute oral lethality than rodent acute oral toxicity test methods.

Because the MEIC study showed that the *in vitro* test methods with the best predictivity generally used human cell lines (Ekwall et al. 1998b), the MEIC management team identified a battery of *in vitro* assays using three human cell lines that had maximal performance for predicting peak acute LC values in humans (R^2=0.79 and Q^2=0.76) (Ekwall et al. 2000). However, it was concluded that improvements in the prediction of human acute oral lethality were necessary before *in vitro* cytotoxicity assays could replace animal tests. To adjust for lethality produced by mechanisms other than basal cytotoxicity, the Evaluation-guided Development of New *In Vitro* Tests (EDIT) program was proposed to address targeted

[4] RTECS® was originally published by the U.S. National Institute for Occupational Safety and Health (NIOSH) and is currently licensed to MDL Information Systems, Inc.

development of *in vitro* test methods for other endpoints, including biokinetics (e.g., gut absorption, distribution, clearance), biotransformation, and target organ toxicity (Clemedson et al. 2002).

1.1.2 An International Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic Toxicity

FRAME organized an international collaborative study conducted in 1992 - 1993 to evaluate the prediction of rodent acute oral lethality by *in vitro* test methods (Fentem et al. 1993)[5]. The objective of the study was to identify *in vitro* systems and strategies that could be used for the classification and labelling of new chemicals, thereby reducing, and possibly replacing, the use of animals for acute oral toxicity testing.

The 42 substances tested in the study comprised a diverse group of organic and inorganic chemical classes, including surfactants, pharmaceuticals, and pesticides (Fentem et al. 1993). *In vitro* toxicity assays using different mammalian cell lines, exposure periods, and toxicity endpoints were evaluated, including:

- Two cell proliferation assays (total protein in mouse BALB/c 3T3 fibroblast cells and MTT[6] reduction in Chinese hamster fibroblastoid V79 cells after a 72-hour exposure period)
- Two cytolethality assays (MTT reduction in V79 cells and lactate dehydrogenase [LDH] release from primary rat hepatocytes after a 24-hour exposure period)
- A cell function assay (myotube contractility inhibition in rat skeletal muscle cells)

The resulting *in vitro* IC_{50} data were linearly regressed against the lowest available rat or mouse oral LD_{50} values for each test substance. There were no significant differences among the IC_{50}-LD_{50} regressions for the different *in vitro* test methods.

A subset of 26 to 40 of the 42 test substances, based on the availability of European Union (EU) hazard classification data, was used to evaluate two approaches for using *in vitro* IC_{50} data to classify chemicals into the four hazard categories used by the EU for acute oral toxicity labelling (Fentem et al. 1993). One approach used the IC_{50} values obtained from the five different *in vitro* test methods for each test substance to predict the LD_{50} value and hazard category from the IC_{50}-LD_{50} regression. The accuracy of hazard classification for the five *in vitro* tests was from 43 to 65%. The other approach used toxicokinetic parameters for 31 to 38 substances to convert the IC_{50} values to effective dose (i.e., ED_{50}) values. Hazard classification accuracy was 43 to 55%.

---

[5] The collaborative study was conducted by the Institute of Toxicology, Kiel, Germany; the University of Nottingham, United Kingdom; and the Gesellschaft für Strahlen- und Umweltforschung, Neuherberg, Germany (Society for Radiological and Environmental Research, which later changed its name to Center for Environmental and Health Research [Forschungszentrum für Umwelt und Gesundheit])

[6] MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide is metabolized by the mitochondrial succinate dehydrogenase of proliferating cells to yield a purple formazan reaction product.

In addition, to ensure that a variety of toxic mechanisms were evaluated during *in vitro* testing, the lowest predicted LD_{50} or ED_{50} from the results of a battery of three tests: a cell proliferation assay (total protein for 3T3 cells); a cytotoxicity/cytolethality assay using primary rat hepatocytes (LDH release); and the rat skeletal muscle cell contractility assay, was used also. The lowest predicted LD_{50} or ED_{50} of the three tests was then used to predict toxicity classification. The accuracy of classification using this approach was 48% for the ED_{50} and 45% for the predicted LD_{50} values.

Based on the results obtained, a battery of *in vitro* tests was recommended for classifying chemicals for their acute lethal potency in rodents (Fentem et al. 1993). The first order test in the battery measures basal cytotoxicity. This study observed no major differences in the performances of the *in vitro* test methods that measure inhibition of cell growth regardless of the cell line (V79, 3T3-L1, or BALB/c 3T3), exposure duration (24-72 hours), or endpoint measurement technique (MTT reduction, neutral red uptake [NRU], or protein concentration). The second order test in the battery assesses hepatocyte-specific toxicity and the role of biotransformation in cytotoxic activity. Co-cultures of rodent hepatocytes with proliferating cells such as 3T3 cells were recommended because the use of hepatocytes alone would not indicate that a chemical requires bioactivation to produce its toxic effects. The third order test in the battery detects chemicals that interfere with electrically excitable membranes at non-cytotoxic concentrations (e.g., a contractility assay using primary cultures of rat muscle cells) (Fentem et al. 1993).

1.1.3 The Registry of Cytotoxicity (RC)

The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS®, and published IC_{50} values from *in vitro* cytotoxicity assays that used a variety of cell lines and cytotoxicity endpoints for substances with known molecular weights (Halle 1998, 2003). The main purpose for compiling the RC was to evaluate, using data from substances with a wide range of rodent acute oral toxicities, whether basal cytotoxicity (averaged over various cell types, cell lines, and/or toxicity endpoints) accurately predicted acute oral lethality in rats and mice. The RC currently contains data for 347 different substances (Halle 1998, 2003) and efforts are underway to increase the number to 500 (ICCVAM 2001a). The RC does not contain data on chemical mixtures.

The RC contains cytotoxicity data for substances that met the following criteria (Halle 1998, 2003):

- At least two different IC_{50} values needed to be available, from studies using either different cell types, different cell lines, or different cytotoxicity endpoints
- Data had to be generated using mammalian cells only (although data from studies using hepatocytes or related cells were excluded)
- The chemical exposure duration had to be at least 16 hours, with no upper limit

The following cytotoxicity endpoints were accepted:

- Cell proliferation: cell number; cell protein; DNA content; DNA synthesis; ^{3}H-thymidine intake; colony formation

- Cell viability/metabolic indicators: metabolic inhibition test (MIT-24); mitochondrial reduction of tetrazolium salts into an insoluble (MTT) or soluble (2,3-bis(2-methoxy-4-nitro-5- sulfophenyl)-2H-tetrazolium-5-carboxanilide [XTT]) dye
- Cell viability/membrane indicators: NRU; trypan blue exclusion; cell attachment; cell detachment
- Differentiation indicators, such as functional and/or morphological changes among and within cells

IC_{50} values (1,912) for 347 substances were obtained from 157 original publications (Halle 1998, 2003). The two to 32 IC_{50} values for each substance were averaged as geometric means to produce one IC_{50x} value for each substance. The rodent LD_{50} values used in the RC were obtained from RTECS®. For the first 117 substances, designated as the training data set (RC-I), LD_{50} values were not revised when subsequent issues of RTECS® reported lower values[7]. For the most recent 230 substances, designated as the verification set (RC-II), the LD_{50} values were taken from the 1983/84 RTECS® publication. Whenever obtainable, oral LD_{50} data from rats were used (282 values). If rat data were unavailable, LD_{50} data from mice were used (65 values). Combining rat and mouse data in the regression was deemed to be justified when separate regressions for the mouse and rat LD_{50} values against the IC_{50x} values did not result in significant differences between the slopes and intercepts of the two regressions (Halle 1998, 2003).

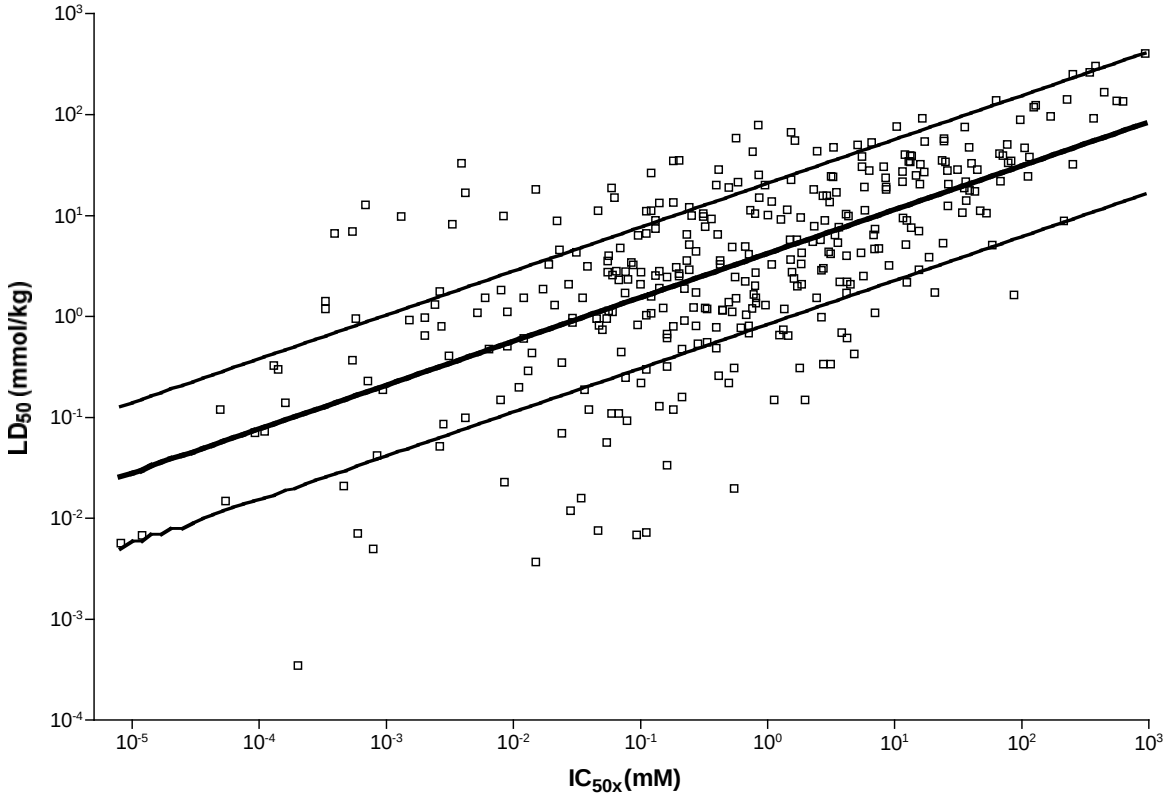
To develop a model for the prediction of acute oral LD_{50} values from IC_{50x} values, Halle (1998, 2003) calculated a linear regression from pairs of the log-transformed IC_{50x} values (in mM) and log transformed rodent oral LD_{50} values (in mmol/kg) (see **Figure 1-1**). Molar concentrations were used to allow for a comparison among chemicals based on the number of molecules rather than formula weights. The regression, referred to here as the *RC millimole regression*, has the following formula:

$$\mathbf{\log LD_{50}\ (mmol/kg) = 0.435 \times \log IC_{50x}\ (mM) + 0.625}$$

To identify an acceptability range for practical use and research purposes, the acceptable prediction interval for the LD_{50} was empirically defined as approximately one-half an order of magnitude on either side of the best-fit linear regression (i.e., $\pm \log 5$, or ± 0.699) (Halle 1998, 2003). This interval was based on eight linear regressions calculated for *in vitro* mammalian cell cytotoxicity data using various endpoints and oral LD_{50} values from rat, mouse, or rat and mouse from five publications. The prediction interval approximates the predicted LD_{50} range for the eight regressions across about eight orders of magnitude of IC_{50} values. When this approach was used, 73% (252/347) of the RC substances fall within the prediction interval.

[7] RTECS® published the lowest LD_{50} reported for a substance and updates the information periodically.

Figure 1-1 RC Millimole Regression for *In Vitro* Cytotoxicity (IC_{50x}) and Rat and Mouse Acute Oral LD_{50} Values for 347 Chemicals



Abbreviations: RC=Registry of Cytotoxicity; IC_{50x}=Geometric mean (of multiple endpoints and cell types) test substance concentration that reduces cell viability by 50%; LD_{50}=Dose producing death in 50% of the animals tested.

The heavy line shows the fit of the data to a linear regression model, log (LD_{50}) = 0.435 x log (IC_{50x}) + 0.625; r=0.67. The thinner lines show the empirical prediction interval (± log 5, or ±0.699) that is based on the anticipated precision for the prediction of LD_{50} values from cytotoxicity data (Halle 1998, 2003).

1.1.4 The ZEBET Initiative to Reduce Animal Use

ECVAM organized a workshop in 1994 to evaluate the use of *in vitro* data for the classification and labeling of chemicals (Seibert et al. 1996). Workshop participants reviewed information on the assessment of acute oral toxicity using *in vitro* data and concluded that, for *in vitro* data to be used most effectively, the following information would be necessary:

- The active concentration *in vitro* (i.e., the actual concentration available to the cultured cells)
- The *in vitro* concentrations that produce basal cytotoxicity, hepatocyte toxicity, and selective cytotoxicity (i.e., effects on cell-specific functions such as transport processes or cell-to-cell communication)
- The effect of biokinetic processes on acute oral toxicity in rodents
- *In vitro* tests that provide the physicochemical parameters needed to estimate equivalent body doses from *in vitro* data

The concept that *in vitro* data could be used to determine the starting doses for rodent acute oral toxicity tests, so as to reduce the number of animals used, was first discussed at this workshop (Seibert et al. 1996). At that time, draft Organisation for Economic Co-operation and Development (OECD) sequential rodent acute oral toxicity test guidelines (TGs) were available; these included the:

- Acute Toxic Class method (ATC; OECD draft Test Guideline [TG] 423 [ICCVAM 2001a])
- Up-and-Down Procedure (UDP; OECD draft TG 425 [ICCVAM 2001a])
- Fixed Dose Procedure (FDP; OECD draft TG 420 [ICCVAM 2001a])

Final OECD TGs now exist for these rodent acute oral toxicity tests. The number of animals needed depends upon the choice of the starting dose because the number of consecutive dosing steps, and thus the number of animals used, is reduced as the starting dose more closely approximates the true toxicity class for the ATC or the FDP, or the true LD_{50} for the UDP.

The ZEBET approach involves using an IC_{50} value from an *in vitro* basal cytotoxicity test with the RC millimole regression to predict an LD_{50} value for use as a starting dose for the ATC or UDP (Spielmann et al. 1999). Using simulation results performed to evaluate the draft UDP test method, ZEBET predicted that the use of *in vitro* cytotoxicity assays to predict a starting dose equivalent to the LD_{50} had the potential to reduce animal use in the UDP by 25-40%, depending upon the slope of the concentration response curve and the stopping rule applied (Spielmann et al. 1999; ICCVAM 2001a).

1.1.5 The International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity

In 2000, the U.S. National Institute of Environmental Health Sciences (NIEHS), the NTP, and the EPA jointly sponsored an International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity (hereafter known as Workshop 2000). This workshop evaluated:

- The ZEBET approach using the RC millimole regression to estimate LD_{50} values and set starting doses for *in vivo* testing
- A testing strategy proposed by the European Center for the Validation of Alternative Methods (ECVAM) (Siebert et al. 1996)
- Other initiatives for reducing animal use in rodent acute oral toxicity testing by using *in vitro* cytotoxicity test methods (ICCVAM 2001a)

The Workshop 2000 participants concluded that no *in vitro* cytotoxicity test methods (or battery of *in vitro* cytotoxicity test methods) existed that could replace the current *in vivo* acute oral toxicity test methods (ICCVAM 2001a). Furthermore, they concluded that none of the *in vitro* models reviewed had been adequately evaluated for reliability and relevance, and their usefulness and limitations for generating information for acute toxicity testing had not been assessed. However, there was agreement that: (1) in the near-term, *in vitro* basal cytotoxicity test methods would be useful for estimating the starting dose for rodent acute oral toxicity studies, and (2) further development, optimization, and validation of *in vitro* test methods that considered target organ specificity and *in vivo* factors like adsorption, distribution, metabolism, and excretion (ADME) that modulate the lethality of a xenobiotic were needed (ICCVAM 2001a). Furthermore, the approach proposed by ZEBET (i.e., the use

of *in vitro* basal cytotoxicity test methods to predict the starting dose for the sequential rodent acute oral toxicity test methods) (Halle 1998, 2003; Spielmann et al. 1999) was recommended for rapid adoption so that data could be generated to establish its usefulness with a larger number of substances (ICCVAM 2001a). To assist in the adoption and implementation of the ZEBET approach, several workshop participants prepared the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b).

The *Guidance Document* recommended testing 10 to 20 RC substances (of high purity) from the RC in a candidate *in vitro* basal cytotoxicity assay to be used for predicting starting doses for acute oral lethality tests (ICCVAM 2001b). The substances were to cover a wide range of toxicities and fit the RC prediction model (i.e., the linear regression line) as closely as possible. The *in vitro* test methods recommended and provided as examples were NRU assays using 3T3 and normal human epidermal keratinocytes (NHK) cells. The IC_{50} results from testing the selected substances would be used to calculate a regression against the LD_{50} values used by the RC. If the resulting regression were parallel to the RC millimole regression and within the ± log 5 (i.e., ±0.699) prediction interval for the RC, the *Guidance Document* recommended using the *in vitro* cytotoxicity assay to predict starting doses for LD_{50} assays. If the regression from the *in vitro* assay did not meet these criteria, then the *Guidance Document* advised either (a) adjusting the slope or (b) using the NRU protocols offered in the *Guidance Document* (considered the most efficient approach).

Based on the conclusions and recommendation of the Workshop 2000 participants, ICCVAM subsequently recommended that near-term validation studies should focus on two *in vitro* basal cytotoxicity assays: one using human cells and one using rodent cells. Human cells are of interest because a principal aim of rodent acute oral toxicity testing is to predict potential lethality in humans, while rodent cells may be a better predictor of lethality in rats and mice (ICCVAM 2001a).

1.1.6 The NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study

In response to the ICCVAM recommendation, NICEATM and ECVAM designed an independent[8] multi-laboratory validation study to evaluate *in vitro* basal cytotoxicity, measured as NRU, as a predictor of acute oral lethality in rodents and potentially in humans. Based on historical *in vitro* cytotoxicity data for mouse BALB/c 3T3 fibroblast cells (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996) and NHK cells (e.g., Gettings et al. 1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994), it was decided that these two cells types should be the focus of this validation effort.

The primary aim of this validation study was to determine if the NRU IC_{50} concentration of a test substance in either 3T3 or NHK cells could be used to estimate the rodent LD_{50}, as a means for predicting the starting doses for rodent acute oral toxicity studies. A secondary aim was to determine the extent to which the NRU IC_{50} in either 3T3 or NHK cells could be used

[8] "Independent" is used here to indicate that neither NICEATM nor ECVAM, nor its members, had a monetary interest in the test methods.

to estimate the blood serum concentrations associated with acute oral lethality in humans. This evaluation will be the focus of a future ECVAM report.

The specific objectives for this validation study were to:

- Further standardize and optimize the *in vitro* NRU basal cytotoxicity protocols using 3T3 and NHK cells to maximize test method reliability (intralaboratory repeatability, intra- and inter-laboratory reproducibility)
- Assess the accuracy of the two standardized *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods for estimating rodent oral LD_{50} values across the five United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity, as well as unclassified toxicities
- Estimate the reduction and refinement in animal use achievable from using the *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods to identify starting doses for *in vivo* acute oral toxicity tests, assuming that no other information were available
- Develop high quality *in vivo* acute oral lethality and *in vitro* NRU cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of *in vivo* acute oral lethality

1.1.6.1 *Study Design*

The planning phase of the validation study included the selection of reference substances for testing, which is described in **Section 3,** and the identification of rodent oral LD_{50} values for the reference substances, which is described in **Section 4**. The validation study proceeded in several phases (see **Figure 1-2**) so that the Study Management Team (SMT) could evaluate the reproducibility of results after each phase and refine the protocols, if necessary, before proceeding to the next phase. The resulting NRU data collected were used to evaluate linear regression formulas for the prediction of LD_{50} values from IC_{50} values (see **Section 6**). Computer simulation modeling of acute oral toxicity test outcomes was then performed to determine potential animal savings using the NRU-predicted starting doses compared with the default starting dose for the UDP and the ATC (see **Section 10**). Study management and study participant information is provided in **Appendix A**.

Figure 1-2 NICEATM/ECVAM Validation Study Phases

Phase Ia: Laboratory Evaluation

Development of a positive control database for each laboratory

- Perform at least 10 replicate NRU tests of the positive control substance (sodium lauryl sulfate [SLS]) with each cell type.
- Calculate mean IC_{50} value ±2 standard deviations for each cell type for each laboratory.
- Establish acceptance criteria for positive control performance in future assays.

⇓

Phase Ib: Laboratory Evaluation

Limited substance testing to demonstrate the reliability of the protocol

- Each laboratory tests the same three coded substances three times with each cell type. There was one substance each from low, medium, and high GHS toxicity categories.
- Refine protocols and repeat, if necessary, until acceptable intra- and inter-laboratory reproducibility is achieved.

⇓

Phase II: Laboratory Qualification

Evaluation of protocol refinements

- Each laboratory tests nine coded substances covering the range of GHS toxicity categories, with three replicate tests per substance in each test method.
- Assure that corrective actions taken in Phase I have achieved the desired results.
- Further refine protocols and re-test, if necessary, to achieve acceptable reliability.
- Finalize protocols for Phase III.

⇓

Phase III: Laboratory Testing Phase

Test of optimized protocols

- Each laboratory tests 60 coded substances in three replicate tests using the finalized protocol for each test method.

Abbreviations: NRU=Neutral red uptake; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005)

1.2 Regulatory Rationale and Applicability for the Use of *In Vitro* Cytotoxicity Test Methods to Predict Starting Doses for Acute Oral Toxicity Testing in Rodents

1.2.1 Current Regulatory Testing Requirements for Acute Oral Toxicity

The major regulatory need for acute oral toxicity testing is for the hazard classification and labeling of products, which is intended to alert handlers and consumers to potential toxicity hazards. The LD_{50} values from acute oral toxicity tests using rodents are used to place substances in various toxicity categories that, in turn, invoke the associated hazard phrases to be used on product labels. **Table 1-1** shows the current U.S. legislation requiring the use of acute oral toxicity testing for product labeling, and the substances regulated. **Table 1-2** shows the statutory test protocol requirements and classification systems used by each U.S. regulatory agency. Also included in this table is the UN Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures, which provides guidance to regulatory agencies on the use of the GHS (UN 2005) as an internationally comprehensible system for hazard communication (OECD 2001b).

Table 1-1 Summary of Current U.S. Legislation for Using Acute Toxicity Data for Product Labeling

Legislation (Year of Initial Enactment)	U.S. Regulatory Agency	Substances Regulated
Federal Insecticide, Fungicide and Rodenticide Act (FIFRA; 1947)	EPA	Pesticides
Federal Hazardous Substances Act (1964)	CPSC	Household products
Occupational Safety and Health Act (1970)	OSHA	Workplace materials
Federal Hazardous Material Transportation Act (1975)	DOT	Transported substances

Abbreviations: EPA=U.S. Environmental Protection Agency; CPSC=U.S. Consumer Product Safety Commission; FIFRA=Federal Insecticide, Fungicide, and Rodenticide Act; OSHA=U.S. Occupational Safety and Health Administration; DOT=U.S. Department of Transportation.
Note: The U.S. Food and Drug Administration (FDA) does not require data for from acute lethality testing, and discourages the use of animals for such testing (FDA 1993).

In addition to classification and labeling, acute oral toxicity test results may be used for:

- Establishing dosing levels for repeated dose toxicity studies or other toxicity studies
- Identifying potential target organs
- Providing information related to the mode of toxic action
- Aiding in the diagnosis and treatment of toxic reactions
- Providing information for comparison of toxicity and dose response among substances in a specific chemical or product class
- Aiding in the standardization of biological products
- Aiding in judging the consequences of single, high accidental exposures in the workplace, home, or from accidental release
- Serving as a standard for evaluating alternatives to animal tests

Table 1-2 Regulatory Classification Systems for Acute Oral Toxicity

Regulatory Agency (Authorizing Act)	Animals	Endpoint	Classification
EPA (FIFRA)	Use current EPA or OECD protocol	Death[1]	I - LD_{50} ≤50 mg/kg II - 50 < LD_{50} ≤500 mg/kg III - 500 < LD_{50} ≤5000 mg/kg IV - LD_{50} >5000 mg/kg
CPSC (Federal Hazardous Substances Act)	White rats, 200-300 g	Death[1] within 14 days for ≥ half of a group of ≥10 animals	Highly toxic - LD_{50} ≤50 mg/kg Toxic - 50 mg/kg < LD_{50} <5 g/kg
OSHA (Occupational Safety and Health Act)	Albino rats, 200-300 g	Death[1], duration not specified.	Highly toxic - LD_{50} ≤50 mg/kg Toxic - 50 < LD_{50} <500 mg/kg
DOT (Federal Hazardous Material Transportation Act)	Male and female young adult albino rats	Death[1] within 14 days of half the animals tested. Number of animals tested must be sufficient for statistically valid results.	Packing Group 1 - LD_{50} ≤5 mg/kg Packing Group II - 5 < LD_{50} ≤50 mg/kg Packing Group III - LD_{50} <500 mg/kg (liquid) LD_{50} <200 mg/kg (solid)
OECD Guidance for Use of GHS (2001b)	Protocols not specified	Not specified	I - LD_{50} ≤5 mg/kg II - 5 < LD_{50} ≤50 mg/kg III - 50 < LD_{50} ≤300 mg/kg IV - 300 < LD_{50} ≤2000 mg/kg V - 2000 < LD_{50} ≤5000 mg/kg Unclassified - LD_{50} >5000 mg/kg

Abbreviations: EPA=U.S. Environmental Protection Agency; OECD=Organisation for Economic Co-operation and Development; LD_{50}=Dose producing death in 50% of the animals tested; CPSC=U.S. Consumer Product Safety Commission; FIFRA=Federal Insecticide, Fungicide, and Rodenticide Act; OSHA=U.S. Occupational Safety and Health Administration; DOT=U.S. Department of Transportation; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).

[1]*Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation* calls for humane killing of moribund animals (OECD 2000). Moribund animals that are humanely euthanized are accepted as deaths.

1.2.1.1 *Test Methods for Assessing Acute Oral Toxicity*

The current internationally recognized test methods for acute oral toxicity testing are the FDP (OECD 2001c), the ATC (OECD 2001d), and UDP (OECD 2001a; EPA 2002a) (see **Appendix M** for test method guidelines). Information on toxic doses and signs of acute toxicity and target organs can be obtained using any of these three methods. All three methods are sequential tests in which the outcome of testing one or more animals at the first dose is used to determine the second dose that should be tested. The FDP differs from the UDP and ATC in that it involves using more animals per dose, and the primary endpoint of interest is evident toxicity[9] rather than lethality. Both the FDP and the ATC methods provide a range estimate of the LD_{50} for classification purposes. The UDP generally provides a point estimate of the LD_{50} with a confidence interval (EPA 2002a).

Each of the test method guidelines includes a limit test in which up to five or six animals are tested at the limit, or upper bound, dose depending on the dose chosen (OECD 2001a, c, d, e; EPA 2002a). The limit test can be performed using 2000 or 5000 mg/kg, depending on the regulatory need.

1.2.2 Intended Regulatory Uses for the *In Vitro* Cytotoxicity Test Methods

In vitro cytotoxicity test methods currently cannot serve as replacements for acute oral toxicity tests in animals. However, such test methods can be used as adjuncts for rodent acute oral toxicity tests. The current test guidelines for acute oral toxicity tests recommend using information from structurally-related substances and the results of any other toxicity tests (EPA 2002b), including *in vitro* cytotoxicity test method (OECD 2001a, c, d; EPA 2002a) to select the starting *in vivo* dose. The 3T3 and NHK NRU test methods may be used as part of this weight-of-evidence approach to select starting doses in order to reduce and refine the use of animals for acute oral toxicity testing.

Section 10 presents computer simulation analyses that characterize the extent of animal reduction and refinement that may occur by using the *in vitro* NRU test methods to estimate the starting doses for the UDP and the ATC method, by estimating the numbers of animals used and the numbers of animal that die. These simulations determined (1) the numbers of animals used when using the default starting dose and, (2) the number of animals used when using a starting dose determined from the *in vitro* NRU test methods. These calculations determined the reduction in animal use that can be achieved when using the *in vitro* NRU test methods. To characterize the extent of refinement produced using the NRU-determined starting dose, the number of animals that would have died with the NRU-determined starting dose was compared with the number of animals that would have died when using the default starting dose. Because there is a lack of information for specific substances about the dose at which evident toxicity occurs in relationship to the LD_{50}, the FDP will not be considered further in this document. However, the use of *in vitro* cytotoxicity data to determine starting doses may also reduce the use of animals in the FDP.

---

[9] *Evident toxicity* is a general term describing clear signs of toxicity following administration of the test substance, such that the next highest fixed dose would result in the development of severe toxic signs, and probably mortality (ICCVAM 2000).

1.2.3 Similarities and Differences in the Endpoints of the *In Vitro* Cytotoxicity Test Methods and Rodent Acute Oral Toxicity Test Methods

The endpoint measured in the *in vitro* NRU test methods is cell death. Neutral red dye is taken up and accumulated only by live cells; the primary measure of interest is the IC_{50} (i.e., the test substance concentration that causes a 50% inhibition of NRU). In contrast, the endpoint measured in acute oral toxicity assays is usually animal morbidity or death. Cell death and animal death may have similar mechanistic bases because all cells, regardless of whether they are in animals or *in vitro* cell cultures, have similar cellular mechanisms; for example, energy production and maintenance of cell membrane integrity.

Death of an animal death and death of a cultured cell due to toxicity both involve interference with vital cell processes or physical injury. Cell death in a culture system involves the death of a single cell type, but through mechanisms that also operate in the animal. In contrast, cellular injury in an animal, if sufficiently widespread or in a critical process, can lead to injury or loss of function of other cell types in a tissue not directly affected by the treatment, resulting in organ failure. Major organ system failures (e.g., liver and kidney failure), gastrointestinal corrosion, and bone marrow depression, can be fatal. Examples of mechanisms leading to such organ failures are disruption of membrane structure or function, inhibition of mitochondrial function, disturbance of protein turnover, and disruption of energy production (Gennari et al. 2004). Alternatively, the tissue injury could affect non-exposed vital organs or tissues through interference with homeostatic signaling mechanisms (Gennari et al. 2004). For example, respiratory depression leading to death may be due to depression of the central nervous system (CNS) rather than a direct assault on the respiratory system itself.

Animal and cell culture systems are also different with respect to how a substance or toxicant is delivered to the cell and how it is distributed within the cell, metabolized, and excreted. After oral administration, animals must absorb the toxicant from the gastrointestinal tract, which involves the passage through membranes, many of which are selective with respect to what molecules they will allow to pass. The toxicant may or may not be bound to serum proteins, thereby reducing its availability to the target organ. The toxicant may be metabolized before, during, and/or after its distribution to the target organs, or the toxicant or its metabolites may be excreted before reaching the target organ or reacting with its components. As a consequence, the most critical target organs may not be exposed to the active metabolite, or be exposed for only a limited time or to a relatively small fraction of the administered dose.

In contrast, in a cell culture system, the test substance is applied directly to the target cells and the only membranes that must be passed are those of the target cell and its subcellular organelles. No absorption and distribution by other cellular systems is required. Cell culture systems may or may not include serum proteins, which could reduce the availability of toxicant to the target site. For example, the 3T3 cell culture medium includes serum while the NHK cell culture medium does not. 3T3 and NHK cells have little to no capacity to metabolize xenobiotic compounds, and added cell-free metabolic activation systems, such as rat liver homogenates, may not accurately mimic all phases of *in vivo* metabolism. Excretion from the cell culture milieu is not a consideration because anything excreted from the cell

remains in the culture medium and is available to the other cells in the culture. As a result, the cells in culture (as opposed to cells in an animal) may be exposed to a test substance for the entire duration of the test protocol.

Animals and cell culture systems may also differ with respect to the target on which a toxicant acts. If a toxicant acts in a specialized organ system *in vivo*, it may not produce a toxic effect by the same mechanism in cultured cells that are derived from a tissue different from the target organ. For example, a substance that affects a neuroreceptor-mediated pathway in animals would not be expected to produce a similar toxicity in 3T3 or NHK cells, which are derived from fibroblasts and skin cells, respectively, and do not contain similar neuroreceptors; if toxicity is seen in these cell cultures, it may be from a different mechanism or in a different concentration relationship than *in vivo*. Even if a neurotoxin were applied to neuronal cells in culture, the cultured cells may not respond in the same way as neuronal cells in an animal because cells in culture, especially cell lines, may not retain the same functionalities as cells *in vivo*.

1.2.4 Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard Assessment

In the overall strategy of hazard or safety assessment, the intended regulatory use of the *in vitro* NRU test methods is to reduce and refine the use of animals in current acute toxicity assays. The *in vitro* systems would serve as adjuncts to the *in vivo* test methods but are not intended as replacements for the rodent acute oral toxicity test methods. For the OECD alternative acute oral toxicity assays (the ATC and UDP), the number of animals used depends on the starting dose. The number of dosing steps (and animals) is reduced if the starting dose is close to the true toxicity class (ATC) or the true LD_{50} (UDP) (Spielmann et al. 1999; ICCVAM 2001b).

As noted earlier, Spielmann et al. (1999) and the *Guidance Document* (ICCVAM 2001b) suggest that the RC millimole regression analysis be used with *in vitro* cytotoxicity data to predict starting doses for the ATC and UDP. The RC millimole regression cannot be applied to unknown substances or to mixtures (e.g., product formulations) because such materials cannot be assigned molecular weights. Therefore, the NICEATM/ECVAM validation study also evaluated the classification accuracy and the reduction in animal use associated with a regression based on weight units (with IC_{50} in μg/mL and LD_{50} in mg/kg) (see **Section 10**). This regression would potentially be appropriate for predicting the starting dose for mixtures and undefined substances.

1.3 Scientific Basis for the *In Vitro* NRU Test Methods

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). Ekwall (1983) described the concept of "basal cell functions" (mitochondrial activity, plasma membrane integrity, etc.) that virtually all cells possess and suggested that, for most substances, toxicity is a consequence of non-specific alterations in those cellular functions, which may then lead to adverse effects on organ-specific functions and/or death of the organism. 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) and others (e.g., Grisham and Smith 1984) concluded that, because the actions of substances that produce injury and death are ultimately exerted at the cellular level, *in vitro* cytotoxicity assays might be useful for the prediction of acute lethality potency, as well. Considerable research has been undertaken to develop and evaluate *in vitro* tests for use as screens and as potential replacements for rodent LD_{50} tests, and numerous groups have reported good agreement between *in vitro* cytotoxicity and animal lethality (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 data have not been assessed.

1.3.1 Purpose and Mechanistic Basis of the *In Vitro* NRU Test Methods

A number of basal cytotoxicity endpoints can be used to measure cell death or interference with cell proliferation. The NRU test methods were chosen for the NICEATM/ECVAM validation study because they were recommended in the *Guidance Document* for the purpose of obtaining cytotoxicity information to determine starting doses for rodent acute oral toxicity assays (ICCVAM 2001b). Both the 3T3 and NHK NRU test methods were reproducible in previous validation studies (ICCVAM 2001b). In addition, both cell types are easily obtainable from commercial sources and the *Guidance Document* provided preliminary evidence that these assays could reproduce the RC millimole regression. Additionally, the assays can be automated and they require no radioactivity or highly dangerous reagents (see **Section 2** for protocol discussion and **Appendix B** for protocols).

Neutral red is a weakly cationic water-soluble supravital dye that stains living cells (Borenfreund and Puerner 1985). It readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of neutral red retained by the culture. Borenfreund and Puerner (1985) were the first to publish a protocol for the NRU assay using 3T3 cells as a method to objectively quantify toxicity previously assessed by subjective, visual observation. The NRU assay, which was standardized for a 96-well plate format, correlated two measurements of toxicity from the exposure of 3T3 cells to six surfactants: (1) a visual morphological evaluation of the cells under an inverted phase microscope, and (2) a quantitative measurement of NRU. The visual evaluation was designed to identify the highest concentration of toxicant that causes only minimal morphological changes (i.e., the highest tolerated dose [HTD]). Because Borenfreund and Puerner (1985) found that the HTD in the NRU test was comparable to the concentration that produced 10% inhibition (i.e., the IC_{10}) compared with the controls, the IC_{10} value was deemed to be a good index for comparing the relative toxicities of experimental agents. The assay was described as a rapid, reliable, inexpensive, and reproducible *in vitro* test method for screening potentially toxic agents (Borenfreund and Puerner 1985). Furthermore, the authors suggested that the test method was a good candidate for inclusion in a battery of assays for toxicity screening with the purpose of reducing the use of animals for toxicity tests.

1.3.2 Similarities and Differences in the Modes/Mechanisms of Action for the *In Vitro* NRU Test Methods Compared with the Species of Interest

Although the ultimate species of interest for acute oral toxicity concerns is humans, labeling and hazard identification requirements are based on rodents. There are differences between humans and rodents in terms of absorption, distribution, metabolism, excretion, and the intrinsic sensitivity of target organs to xenobiotic compounds. The differences are largely substance-specific and quantitative, although there are a number of substances where the human may produce metabolites not seen in the rodent and vice versa. *In vitro* cytotoxicity studies have also noted differences in sensitivity between human cells and other mammalian cells (Clemedson et al. 1996b). It is important to note that, for certain chemicals, there can also be large differences in sensitivity among different human cell types and cell lines (Clemedson et al. 1996b, 1998a, b).

Because of the differences in sensitivity between humans and rodents, it might be likely that cultured human cells would predict human lethality better than cultured rodent cells and that cultured rodent cells would predict rodent lethality better than human cells. Ekwall et al. (1998b) showed that *in vitro* cytotoxicity test methods using human cell lines generally predicted human toxicity more accurately than did test methods using nonhuman mammalian cells.

In addition to being derived from different species, there are several other differences between 3T3 and NHK cells, all of which may contribute to differences in sensitivity.

- 3T3 cells are an immortal line, while the NHK cells are primary cells.
- The cells originate from different tissues; 3T3 cells are derived from embryonic fibroblasts, while the NHK cells are isolated from neonatal foreskin tissue.
- NHK cells grow more slowly in culture than the 3T3 cells (i.e., after seeding into 96-well plates, NHK cells require 48-72 hours for growth to the appropriate confluence while 3T3 cells require approximately 24 hours; see **Appendix B**).
- NHK cells have greater ability to metabolize xenobiotic compounds, in that they exhibit minimal cytochrome P450 activity (Babich et al. 1991), whereas 3T3 cells have practically no ability to metabolize xenobiotic compounds (INVITTOX 1991).

1.3.3 Range of Substances Amenable to the *In Vitro* NRU Test Methods

The *in vitro* NRU test methods can be applied to a wide range of substances as long as they can be dissolved in the cell culture medium or in a nontoxic solvent (at the concentration used), and do not react with the culture medium. Although these test methods may to be applicable to mixtures, none were evaluated in this validation study. The toxicity of substances that act by mechanisms not expected to be active in 3T3 or NHK cells (e.g., those that are specifically neurotoxic or cardiotoxic) will likely be underpredicted by these test methods. Therefore, until more appropriate cell lines are developed, the results from basal cytotoxicity testing with such substances may not be relevant for predicting *in vivo* effects.

Insoluble substances or those unstable in aqueous environments are not compatible with the test systems. Volatile substances may yield acceptable results if CO_2 permeable plastic film is used to seal the test plates. Testing for corrosive substances is unnecessary since there is no regulatory requirement for acute oral toxicity testing for known corrosives. The 3T3 NRU test method may underestimate the toxicity of substances that are highly bound to serum proteins because the culture medium contains 5% serum during substance exposure. The toxicity of substances that specifically affect lysosomes may be overestimated because they may affect NRU binding, and therefore, retention, in the cell. Red substances (and other colored substances) that absorb light in the optical density range of NR may interfere with the test if they remain inside the cell in sufficient amounts after washing and are soluble in the NR solvent.

[This Page Intentionally Left Blank]