ICCVAM
The Interagency Coordinating Committee on the Validation of Alternative Methods

NICEATM
The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

PEER REVIEW PANEL REPORT:
The Use of *In Vitro* Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing

June 2006

View Table of Contents
HTML Format

Peer Review Panel Report:
The Use of *In Vitro* Basal Cytotoxicity Test Methods For Estimating Starting Doses For Acute Oral Systemic Toxicity Testing

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

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

June 2006

This document is available electronically at:
http://iccvam.niehs.nih.gov/methods/invidocs/panelrpt/ATpanelrpt.htm

TABLE OF CONTENTS

IN VITRO ACUTE TOXICITY PEER PANEL ROSTER

- **David H. Blakey**, D.Phil., Health Canada, Ottawa, Ontario, Canada
- **June Bradlaw**, Ph.D., International Foundation for Ethical Research (IFER), Rockville, Maryland
- **Robert Copeland**, Ph.D., Howard University College of Medicine, Washington, DC
- **Gianni Dal Negro**, D.V.M., Ph.D., GlaxoSmithKline Medicine Research Centre, Verona, Italy
- **Marion Ehrich**, Ph.D., RPh., D.A.B.T. Virginia-Maryland Regional College of Veterinary Medicine, Blacksburg, Virginia
- **Eugene Elmore**, Ph.D., University of California, Irvine, Irvine, California
- **Benjamin Gerson**, M.D., Thomas Jefferson University School of Medicine, Philadelphia, Pennsylvania
- **Michael Greene**, Ph.D., U.S. Consumer Product Safety Commission, Bethesda, MD
- **Janice Kuhn**, Ph.D., D.A.B.T., Stillmeadow Inc., Sugar Land, Texas
- **Daniel Marsman**, D.V.M., Ph.D., D.A.B.T., Procter & Gamble Company, Cincinnati, Ohio
- **Andrew Rowan**, Ph.D., Humane Society of the United States, Washington, DC
- **Hasso Seibert**, Ph.D., University Medical School Schleswig-Holstein, Kiel, Germany
- **Nigel Stallard**, Ph.D., The University of Warwick, Coventry, United Kingdom
- **Katherine Stitzel**, D.V.M., (Panel Chair), Consultant, West Chester, Ohio
- **Shinobu Wakuri**, MSc., Hatano Research Institute, Japan
- **Daniel Wilson**, Ph.D., D.A.B.T., The Dow Chemical Company, Midland, Michigan

PREFACE

This is an independent report of the *In Vitro* Acute Toxicity Peer Review Panel ("Panel") organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The report summarizes discussions, conclusions, and recommendations of the public meeting of the Panel that was held at the National Institutes of Health in Bethesda, MD, on May 23, 2006. The ICCVAM and the Acute Toxicity Working Group (ATWG) will consider the Panel report, along with public comments, to prepare final test method recommendations for U.S. Federal agencies. ICCVAM test method recommendations will be forwarded to U.S. Federal agencies for consideration and action, in accordance with the ICCVAM Authorization Act of 2000 (P.L. 106-545).

NICEATM and the European Centre for the Validation of Alternative Methods (ECVAM) organized and conducted the NICEATM/ECVAM *In Vitro* Basal Cytotoxicity Validation Study. NICEATM, in coordination with the ATWG and ICCVAM, prepared a comprehensive draft background review document (BRD) reviewing the study. The draft BRD documents the procedures and results generated from the multi-phase study using the BALB/c 3T3 murine fibroblast (3T3) and normal human epidermal keratinocyte (NHK) neutral red uptake (NRU) test methods for the prediction of starting doses for acute oral toxicity test methods. The draft BRD was made publicly available on the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov) or from NICEATM on request.

NICEATM, in collaboration with the ATWG and ICCVAM, announced the independent Peer Panel review of the test methods in March 2005. Comments from the public and scientific community were solicited and provided to the Panel for their consideration (FR Notice Vol. 71, No. 54, pp. 14229-30, 3/21/06).

The Panel was charged with:

- Developing conclusions and recommendations regarding the usefulness and limitations of *in vitro* NRU basal cytotoxicity test methods using the 3T3 and NHK cells to estimate the rat oral acute LD_{50} for the purpose of determining the starting dose for *in vivo* acute oral toxicity test methods and thereby reducing animal use
- 'Peer reviewing' the NICEATM/ECVAM *In Vitro* Acute Toxicity Test Methods Draft BRD for completeness and for any errors or omissions
- Evaluating the information in the Draft BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003[1]) have been appropriately addressed

[1] ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC:NIEHS. The guidelines can be obtained at: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm

(validation[2] of a new test method is a prerequisite for it to be considered for regulatory decision-making)
- Considering the ICCVAM draft test method recommendations for these test methods (i.e., the proposed test method uses, the proposed recommended standardized protocols, and the proposed test method performance standards) and comment on whether the recommendations are supported by the information provided in the Draft BRD

During the public meeting on May 23, 2006, the Panel discussed the current validation status of the *in vitro* test methods. The Panel also provided formal comment on the Draft BRD and made recommendations for revisions to the Draft BRD. The Panel also provided formal comment on the ICCVAM recommendations for test method use, future studies, test method performance standards, and the cytotoxicity protocols. In addition, the public were provided time at the public meeting to comment on the Draft BRD. The Panel then provided final endorsement regarding the validation status of the test methods.

The Panel gratefully acknowledges the efforts of the NICEATM staff in coordinating the peer review logistics and accommodations and in the preparation of the Draft BRD and various other materials for the review.

[2] Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

EXECUTIVE SUMMARY

Introduction

This report describes the conclusions and recommendations of the *In Vitro* Acute Toxicity Peer Panel ("Panel") regarding the validation status of the BALB/c 3T3 murine fibroblast (3T3) and normal human epidermal keratinocyte (NHK) *in vitro* neutral red uptake (NRU) basal cytotoxicity test methods (hereafter designated as NRU test methods) and the ability to use these test methods to estimate starting doses for acute oral systemic toxicity tests. The Panel accepts the sections of the Draft Background Review Document for *In Vitro* Acute Toxicity Test Methods (BRD) for which it had no comments and recommendations as adequate and acceptably accurate.

Panel Recommendations for the BRD

The Panel stated that, in general, the information presented in the Draft BRD was sufficient for its purpose. Exceptions are noted within the body of the Panel report. The Panel concluded that the objectives of the validation study were appropriate, and agreed that the applicable validation criteria were adequately addressed in the Draft BRD for using these *in vitro* test methods to determine starting doses for acute oral systemic toxicity tests.

The Panel made numerous recommendations for additional explanations (e.g., provide the rationale for using serum that is not heat-inactivated) and clarifications (e.g., provide additional details for using the GraphPad PRISM® software to calculate IC_{50} values) to the Draft BRD that will not require additional statistical analyses. Some recommendations included presentation of the existing data in other formats (e.g., using the relative IC_{50} ratios between the reference substances and the positive control [at the level of the individual laboratory] to compare similar substances across test methods), or additional analyses (e.g., determine the usefulness of the test methods to estimate starting doses for the Fixed Dose Procedure [FDP] acute toxicity test method).

The Panel concluded that several confounding factors were not addressed in the selection or evaluation of test substances but should be. They recommended that the octanol:water coefficients and the surface-active potential (to the extent possible) for the 72 reference substances should be characterized and incorporated into the assessment of accuracy. The Panel also recommended that protein binding should also be taken into account in the data analyses (i.e., to the extent possible, the free fraction in serum corresponding to the LD_{50} should be considered). Another potential confounder was the attempt to select chemicals to prevent the entire set of reference substances from having proportionally more *outlier* substances than the Registry of Cytotoxicity (RC) linear regression.

In the evaluation of test method accuracy, substances with neurotoxic and cardiotoxic mechanisms, and those that interfere with energy utilization or that alkylate cellular macromolecules were excluded. Such substances were excluded because it was expected that these mechanisms of action could not be detected by the NRU test methods. The Panel disagreed with their exclusion because interference with energy metabolism and alkylation of proteins and deoxyribonucleic acid (DNA) represent important mechanisms of cytotoxicity

that should be detected by these two test methods. Additionally, there was consensus among the Panel members that the available data on the mechanism of acute *in vivo* toxicity were not sufficient to justify the exclusion of substances based on mechanism and/or possible involvement of biotransformation reactions. However, the Panel recommended that the properties (e.g., metabolism, receptors, transporters) of the cell types that are important for basal cytotoxicity be better characterized. Despite the fact that there was no significant difference between rat and mouse LD_{50} data from the RC, the Panel indicated that the separation of such data (in developing *in vitro-in vivo* regressions) is useful because it decreases the biological variability associated with species differences.

Although the Panel recommended additional analyses for the evaluation of intra- and inter-laboratory reproducibility (i.e., the comparison of ratios of the maxima and minima mean laboratory IC_{50} values), the Panel agreed that these would not change the conclusion that the NHK NRU test method was more reproducible than the 3T3 version. The Panel suggested that an explanation for the difference in interlaboratory reproducibility be provided.

The Panel recommended that the analyses to determine the reduction of animal use consider prevalence (i.e., the distribution of the universe of substances that are likely to be tested within each hazard classification). The Panel also recommended that animal reduction/refinement be evaluated for the use of the NRU test methods to determine the starting dose for the FDP.

The Panel suggested that costs for equipment and working time needed to perform the NRU test methods and a cost-benefit analysis, including information on the reduction of the number of animals used, should be included in the Draft BRD. The time needed to prescreen NHK culture medium should also be included.

Validation Status of the NRU Test Methods

The Panel agreed that the applicable validation criteria have been adequately addressed for using these *in vitro* test methods in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. However, the Panel was aware that validation of the two NRU test methods was carried out not only to determine if they could be used to set starting doses for *in vivo* acute toxicity studies, but also to determine the extent to which the tests could be useful step in an *in vitro* tiered testing strategy for acute toxicity. The Panel agreed the validation study showed that neither of the two NRU test methods evaluated could be used as a stand-alone replacement for the *in vivo* tests even considering the variability of the latter. The Panel encouraged future work to develop a tiered testing strategy that includes basal cytotoxicity as part of the overall strategy.

Review of the Draft Interagency Coordinating Committee on Validation of Alternative Methods (ICCVAM) Recommendations for Test Method Use

The Panel agreed that although neither of the NRU test methods can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification, the test

methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. The Panel agreed that the NRU test methods be considered before animals are used if there was no other stronger weight-of-evidence information on which to base a starting dose.

The Panel disagreed that the NRU test methods were not appropriate for substances that interfere with energy utilization or alkylation of proteins and other macromolecules and with using the revised RC regression that excluded chemicals based on mechanism of action. However, the Panel agreed with using the RC rat-only regression to estimate the LD_{50} from IC_{50} data and agreed that a regression based on weight rather than molar units would be useful for situations where the molar weight of the test substance is unknown. In situations where the molecular weight of a test substance is known, the molar regression should be used.

The Panel agreed that other *in vitro* basal cytotoxicity test methods are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

Some Panel members agreed that the 3T3 NRU, based on relative ease of performance and cost, should be recommended for general use, but cautioned that one test method should not be preferred over the other. One Panel member noted that it is important to remember that hazard assessment relates to the safety of humans, not rats. The NHK NRU IC_{50} data had a higher correlation with human LC_{50} values (R^2=0.62) than did rodent 3T3 NRU IC_{50} data (R^2=0.51) and a higher correlation than did rodent LD_{50} data with human LC_{50} values (R^2=0.56) (Casati et al. 2005).

Review of the Draft ICCVAM Recommendations for Future Studies

The Panel indicated that high quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to further evaluate the use of the these test methods for predicting the starting dose for acute oral toxicity tests. However, no Panel member recommended that *in vivo* testing be conducted solely to collect data to further assess the usefulness of the NRU test.

The Panel agreed that additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification should be investigated. The Panel also agreed that the *in vivo* database of reference substances used in the validation study be used to evaluate the utility of other non-animal approaches to estimate starting doses for rat acute oral toxicity tests.

The Panel agreed that standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included, to the extent possible, in future rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* test methods. The Panel recommended that ICCVAM consider convening a working group to explore mechanisms of

action of acute toxicity and approaches for acquiring additional information on acute toxic mechanisms during acute toxicity testing.

The Panel agreed that an expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test method development and validation studies and that there should be a concerted effort to obtain higher quality proprietary data from regulated industries.

Review of the Draft Performance Standards for *In Vitro* Acute Toxicity Test Methods and Draft Recommended Test Method Protocols

The Panel agreed that the available data from the validation study appeared to support the validity of the recommended performance standards for the two NRU test methods. The usefulness and limitations were well covered. Although the two NRU test methods may be useful, there would be cause for concern if use of the test methods were made compulsory for regulatory purposes as other information such as structure-property relationships, when available, could provide better estimates of starting doses for acute toxicity studies.

The Panel identified several aspects of the performance standards that should be clarified. Specifically, the Panel recommended that more thorough explanations and more detail for test method procedures should be added to the recommended test method protocols but that an effort should be made to streamline them, where possible, to assure easy use and transferability. Clarification of solubility procedures for the determination of test substances should be provided since the variability between laboratories in the selection of solvent indicates a possible flaw in the solvent determination procedure. The Panel also suggested including other methods for calculating the IC_{50} values and a recommendation for task-specific training for laboratory technicians.

1.0 Introduction And Rationale for the Use of *In Vitro* Neutral Red Uptake (NRU) Cytotoxicity Test Methods to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Testing

This section of the Draft *In Vitro* Acute Toxicity Test Methods Background Review Document (BRD) provided valuable historical background on the use of *in vitro* NRU test methods to predict starting doses for *in vivo* acute oral systemic toxicity. The objectives of the validation study were valid. The introduction acknowledged that *in vitro* cytotoxicity could not replace the Up-and-Down Procedure (UDP) or the Acute Toxic Class method (ATC) acute oral toxicity tests in animals. Furthermore, these *in vitro* tests would not be appropriate substitutes for any of the other standard acute toxicity tests. The Draft BRD recommended that *in vitro* cytotoxicity testing be part of a weight-of-evidence approach to determining the starting dose for *in vivo* acute oral systemic toxicity testing.

1.1 Background and Rationale for the Use of *In Vitro* Cytotoxicity Test Methods to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Tests

This section briefly mentioned the concept of using the predicted LD_{50} value as a starting dose for acute oral toxicity to reduce the number of animals. This was first discussed at a European Centre for the Validation of Alternative Methods (ECVAM) workshop in 1996 (Seibert et al. 1996). The Panel suggested that this section also include the other major conclusions and recommendations of that workshop. The 1996 ECVAM workshop arrived at a general consensus, that

- Testing for basal cytotoxicity is not sufficient for prediction of acute systemic toxicity.
- Biokinetic factors must be considered before performing *in vitro*/*in vivo* comparisons, in order to make the *in vivo* and *in vitro* data more comparable and the resulting comparison more meaningful.

The Panel also recommended including information from an international project supported by the Commission of the European Communities. The project was performed in 1992 and 1993 by the Fund for Replacement of Animals in Medical Experiments (FRAME); Institute of Toxicology, Kiel, Germany; University of Nottingham, United Kingdom (UK); and Gesellschaft für Strahlen- und Umweltforschung (Society for Radiological and Environmental Research, for which the name changed to Forschungszentrum für Umwelt und Gesundheit [Center for Environmental and Health Research]), Neuherberg, Germany. The report, *An International Evaluation of Selected in Vitro Toxicity Test Systems for Predicting Acute Systemic Toxicity* (Fentem et al. 1993), contains results on the *in vitro* cytotoxicity of 42 substances determined with a 3T3 NRU test method and several other *in vitro* systems. Many of the substances tested are identical to those tested in the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)/ECVAM validation study. Furthermore, the report contains statistical analyses of correlations between rodent LD_{50} values and *in vitro* IC_{50} values, and evaluations of the accuracy of the *in vitro* methods for predicting LD_{50} values and acute toxicity categories, respectively.

The Registry of Cytotoxicity (RC) is a registry of lethality and IC_{50} values. The Panel agreed that this database is important and that increasing the numbers of chemicals in this database would be of value. However, IC_{50} values do not indicate the steepness of slope for the cytotoxicity concentration response relationship nor the number of points the value is based on. Furthermore, the RC used many endpoints for cytotoxicity, some of which may be reversible (e.g., cell detachment, effects on cell proliferation). These deficiencies must be mentioned.

The stepwise approach for the validation study was a good approach because it allowed for the review of intermediate progress.

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

1.2.1 *Current Regulatory Testing Requirements for Acute Systemic Toxicity*

This section provided a great deal of detail regarding the context of the regulatory requirements for acute oral toxicity assays.

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

This section should clarify that the NRU cytotoxicity test methods are to be used in a weight-of-evidence approach to determining the starting dose for acute oral systemic toxicity assays. The default starting dose is usually used when there is no information upon which to base a starting dose (e.g., no toxicity information from chemicals with similar structure, etc.).

The Draft BRD indicated that the NRU cytotoxicity test methods could not be used to determine the starting dose for the Fixed Dose Procedure (FDP) because it is not possible to predict a dose that leads to non-fatal toxicity (the TD_{50}). The Panel suggested the TD_{50} and IC_{50} are highly correlated, so that, given TD_{50} data, a regression model for prediction of TD_{50} from IC_{50} could be constructed. Even in the absence of TD_{50} data, a simple procedure such as assuming that the FDP starting dose is two doses below the estimated LD_{50} would be worth investigating. The studies of one Panel member, who has compared IC_{50} values for growth inhibition and mitochondrial function of various epithelial cell lines from normal human tissues, showed that adverse events in clinical studies were observed only after plasma levels exceeded the *in vitro* IC_{50} levels by about one log or more.

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

Animal death and death of cells in culture may or may not have similarities at the cellular level. As noted in the Draft BRD, extrapolation to the whole organism may involve more than just cellular death.

The Draft BRD recognized the ability of normal human epidermal keratinocytes (NHK) to metabolize some xenobiotic substances. The fact that BALB/c mouse fibroblast 3T3 (3T3) cells and NHK cells responded differently to several of the reference substances tested could result from differences in doubling times between the two cell lines. It also could result from detoxification mechanisms or metabolites generated in the NHK cells. The use of serum can

complicate the issue of determining and/or identifying mechanism of toxicity. The 3T3 cell culture system included serum, while the NHK cell culture system did not. Mechanistic differences in cell type are recognized for toxicants that act at particular receptors.

Toxin should be used to refer to a biological product. Since the NICEATM/ECVAM validation study tested pure chemicals, the term *toxicant* should be used.

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

The Draft BRD indicated that the RC millimole regression cannot be used with mixtures and unknown substances because the equation requires molecular weight information for the mole units. The new regression formula (developed in Section 6) based on gram units should be described in this section, too. The new regression formula would be applicable to mixtures and unknown substances.

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

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

The Draft BRD should clarify the extent to which Borenfreund and Puerner (1985) relied on morphology to determine the maximal tolerated dose.

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*

This section well delineated the differences between the cell types.

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

This section of the Draft BRD appropriately identified problems concerning substances with specific toxicity mechanisms, those that were insoluble or volatile, the presence of serum, lysosomal sequestration, and red color. It should be noted that other colored compounds may present a problem as well.

2.0 Test Method Protocol Components of the 3T3 and NHK *In Vitro* NRU Test Methods

The information presented in Section 2 of the Draft BRD appeared to be sufficient. There was a great deal of detail regarding the equipment, methods, and procedures required for implementation of the proposed 3T3 and NHK NRU test methods.

The Guidance Document (ICCVAM 2001b) recommendations were good. This section should explain why it is important to have an exposure period of at least the duration of one cell cycle.

2.1 Overview of the 3T3 and NHK NRU Test Methods

This section of the Draft BRD noted the similarities and differences of the 3T3 and NHK NRU cytotoxicity test methods. The similarities included preparation of reference substances

and the positive control, cell culture environmental conditions, determination of test substance solubility, 96-well plate configuration, 48 hour exposures, microscopic evaluation, NRU measurement as % of control with concentration in μg/mL, and data analysis. The 3T3 and NHK NRU differed in conditions for cell propagation, cell growth media, and application of reference substances (volume). The Panel noted that the IC_{50} values obtained during the study are only valid under the conditions used in the conduct of the test methods.

2.1.1 *The 3T3 NRU Test Method*

The Panel noted that the serum for the 3T3 NRU test method was not heat-inactivated. Serum that is not heat-inactivated can contain enzymes (i.e., esterases) that transform certain chemicals. The Draft BRD should explain the rationale for using serum that is not heat-inactivated. Of the 21 substances deleted from the accuracy analyses (Table 6.3 of the Draft BRD), one Panel member noted that eight substances (atropine, carbamazepine, dichlorvos, disulfoton, fenpropathrin, parathion, physostigmine, procainamide) had structures that could have been biotransformed by serum enzymes.

The Draft BRD should also discuss the rationale for the restriction of the use of the 3T3 cells to less than 18 passages after thawing.

2.1.2 *The NHK NRU Test Method*

Keratinocytes were not subcultured beyond the second passage, which is not unusual for primary cells. The Draft BRD should acknowledge that the use of different lots of NHK cells by an investigator might increase variability.

2.1.3 *Measurement of NRU for both 3T3 and NHK Test Methods*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2 Descriptions and Rationales of the 3T3 and NHK NRU Test Methods

The Draft BRD mentioned that there were problems concerning the growth of both the 3T3 and NHK cells. Since the growth rate can be very important for the results of the cytotoxicity test methods, the Draft BRD should report the doubling times after seeding the cells in 96-well plates and during exposure.

2.2.1 *Materials, Equipment, and Supplies*

Materials and equipment were listed in this section. There was no information regarding the maximum absorbance required of the plate reader; this must be provided as many spectrophotometers following Beer's Law can only read a maximum optical density (OD) of ~ 3.

2.2.2 *Reference Substance Concentrations/Dose Selection*

A commercial medium (keratinocyte basal medium [KBM®] supplied by Clonetics®) was used for culturing the NHK cells. There was no specific information on the composition of this medium. The exact composition of the medium should be specified, especially, whether sera are included, and, if so, the types and concentrations. Without this information, it is

impossible to judge whether differences in medium composition may contribute to the differing results of the test methods for several of the test substances.

2.2.3 *NRU Endpoints Measured*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.4 *Duration of Reference Substance Exposure*

The 48-hour duration of exposure was justified in this section. The differences between *in vitro* cytotoxicity at 24- and 72-hour exposures were noted. As part of future research, it might be of interest to extend the duration of exposure to 96 hours to parallel the 4-day exposure used in animal test protocols. On the other hand, a time course may be important. Recovery and cell growth would suggest that an agent's IC_{50} could change at 72 or 96 hours relative to that at 48 hours. If recovery occurs, then lethality would require a higher dose. Perhaps two time points as used by Elmore (2001, 2002) would be useful. These studies used three days and five days for exposure and noted differences in the IC_{50} values. These time points were chosen to facilitate detection of growth inhibition. Increasing toxicity at five days suggested the agent was more toxic while decreasing toxicity suggested recovery of the cells.

2.2.5 *Known Limits of Use*

This section of the Draft BRD contained caveats on solubility, volatility, and pharmacokinetics, noting that the latter was not addressed. The organ-specific section contained a 5-step *in vitro* test method. The value of including this organ-specific section was unclear since it did not refer to the use of organ-specific cells. The organ-specific section was more concerned with metabolism, energy production, and disruption of epithelial barriers.

Another limitation of use of the *in vitro* test methods is for substances that etch plastics and those that film out (i.e., form a film on the medium surface or plastic well wall). Substances that etch plastics can be detected by looking for the presence of etched rings in the 96-well plates after exposure. Some substances that film out in medium may etch plastic. Additionally, substances that film out decrease the concentration delivered to the cells. Such substances can be identified by the changes produced in the meniscus of the medium or by the presence of a film where the surface of the medium was in the well.

2.2.6 *Nature of Response Assessed*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.7 *Appropriate Vehicle, Positive, and Negative Controls*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.8 *Acceptable Ranges of Control Responses*

The Draft BRD should explain why vehicle control (VC) ODs were lower during Phase II and Phase III testing. Higher viability appeared to correlate with high absorbance. The VC OD ranges of each laboratory should be described so that the stability of cell growth conditions in each laboratory can be evaluated.

The doubling time of each cell type (for each laboratory) should be described in this section.

2.2.9 *Nature of Experimental Data Collected*

Since the Study Director decided whether to remove outliers at 99% level, the Study Director must be an expert in theory and practice of cell culture.

2.2.10 *Type of Media for Data Storage*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.11 *Measures of Variability*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.12 *Methods for Analyzing NRU Data*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.13 *Decision Criteria for Classification of Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.14 *Information and Data Included in the Test Report*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.3 Basis for Selection of the *In Vitro* NRU Cytotoxicity Test Methods

The selection of NRU cytotoxicity test methods was derived from the Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity (ICCVAM 2001a). Workshop participants evaluated several *in vitro* initiatives to evaluate the prediction of systemic toxicity from *in vitro* toxicity. Workshop participants concluded that there were no differences between species sources or between continuous cell lines and primary cells.

2.3.1 *Guidance Document Rationale for Selection of In Vitro NRU Cytotoxicity Test Methods*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.3.2 *Guidance Document Rationale for Selection of Cell Types*

ICCVAM wanted rodent cells used in a cytotoxicity test method because LD_{50} data is obtained with rodents. Cell lines rather than primary cultures would hasten generation of an *in vitro* database. Highly differentiated cells were not used and neither were metabolically active cells such as liver.

2.4 Proprietary Components of the *In Vitro* NRU Cytotoxicity Test Methods

Proprietary cells and media were used for the NHK NRU method (Clonetics®).

2.5 Basis for Number of Replicate and Repeat Experiments for the 3T3 and NHK NRU Test Methods

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.6 Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols

The Panel recommended that the OD of the positive control be included in Table 2-2 of the Draft BRD. The VC OD range was eventually deleted as a test acceptance criterion.

The Panel asked whether something other than mechanism of action contributed to the unusual concentration-response curves for aminopterin and colchicine. The Draft BRD should identify those substances for which the IC_{50} was calculated using only one point between 0 and 100% when a substance had a steep concentration-response curve. The Panel preferred that there be three points between 10 and 90% viability

2.6.1 *Phase Ia: Laboratory Evaluation Phase*

The ring of dead NHK cells was produced by the use of the plate inversion technique for removing the cell culture medium prior to refeeding the cells. Such a technique leaves residual media around the edges of each well. The ring of dead cells can be avoided by aspirating the medium from the wells prior to refeeding. Aspiration also obviates the need to prepare chemicals as a 2X dilution. A 1X chemical solution (or vehicle control) can be added to the cells immediately after aspiration to avoid drying of the cells.

2.6.2 *Phase Ib: Laboratory Evaluation Phase*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.6.3 *Phase II: Laboratory Qualification Phase*

The approach for handling of volatile agents by covering the 96-well plates with plastic film was appropriate. The Panel recommended that oil not be used to cover the culture media surface because agents that bind to lipids can bind to the oil, which reduces their effective concentration.

Prism® software calculations for IC_{50} using Hillslope and midpoints may under- or overestimate the IC_{50} depending on the inclusion of nontoxic concentrations for which viability is >100%, highest test concentrations that produce less than complete toxicity (i.e., viability >0%), or concentration-response curves for which the lowest nontoxic concentration produced <100% viability. The Panel was not satisfied with the current explanation for the IC_{50} calculation.

2.6.4 *Phase III: Laboratory Testing Phase*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.7 Differences in 3T3 and NHK NRU Test Method Protocols and the *Guidance Document* Standard Protocols

2.7.1 *Optimization of the Guidance Document Protocols Prior to Initiation of the Study*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.7.2 *Optimization of the Guidance Document Protocols During the Study*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.8 Overview of the Solubility Protocol

A complex flow chart for the solvent selection for each test substance was provided.

2.9 Components of the Solubility Protocol

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.9.1 *Medium, Supplies, and Equipment Required*
The Panel suggested that the visual solubility determination be performed using a microscope.

2.9.2 *Data Collection*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate

2.9.3 *Variability in Solubility Measurement*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.9.4 *Solubility and the 3T3 and NHK NRU Test Methods*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate

2.9.5 *Methods for Analyzing Solubility Data*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate

2.10 Basis of the Solubility Protocol

The Panel had no comments on this section, although the comments on the protocol itself are addressed below.

2.10.1 *Initial Solubility Protocol Development*
The Draft BRD noted that sometimes BioReliance and the cytotoxicity testing laboratories did not get the same solubility results and additional explanation as to why this occurred would be useful. However, as a whole, solubility was not a major issue.

2.10.2 *Basis for Modification of the Phase II Protocol*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.11 Summary

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.0 Reference Substances Used for Validation of the 3T3 And NHK NRU Test Methods

3.1 Rationale for the Reference Substances Selected for Testing

The selection of test chemicals, the determination of reference *in vivo* data, as well as test method standardization and validation appeared to be well described, and generally of high quality. A wide range of substances, belonging to many chemical classes, with varying physical properties, and different mechanisms of toxicity were included. The list included pharmaceuticals, pesticides, solvents and a number of metal-containing molecules; however, there were no polycyclic aromatic hydrocarbons, catalysts, simple aldehydes, ketones, biocides, cosmetic ingredients, mixtures/formulations, plant toxins, or other natural compounds. The molecular structures were not provided and should be.

The adequacy of the range of reference substances and their mechanisms of oral toxicity was difficult to judge because there is often very limited knowledge about their mechanisms of action. The overall poor characterization of modes or mechanisms of action of acute oral toxicity *in vivo* makes it difficult to strategically select reference substances for broad acute toxicity validation of *in vitro* methods. However, since the NRU methods are expected to detect basal cytotoxicity, the selected substances should be sufficient to evaluate reliability and accuracy. Specifically, the Draft BRD provided little information about the 72 reference substances to indicate that specific modes of action of acute systemic toxicity had been robustly explored.

The standardized methodology for acute toxicity protocols (i.e., the traditional LD_{50} or the reduced UDP procedure), which include only the most rudimentary collection of endpoints, makes no attempt to characterize even the simplest modes of action of a test substance. As such, the overall poor characterization of these reference substances for modes or mechanisms of action of acute oral toxicity *in vivo* made it difficult to strategically select reference substances for broad acute toxicity validation of *in vitro* methods.

Within this context, there may be some limited value in adding data from additional substances to improve precision. Inclusion of substances at the extremes of the GHS toxicity categories may be helpful.

3.1.1 *Reference Substance Selection Criteria*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.1.2 *Candidate Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.1.3 *Selection of Reference Substances for Testing*

The selection of reference substances for evaluating the reliability and the accuracy of the NRU cytotoxicity test methods was well planned and executed, arriving at a broad and fairly complete selection of model chemicals. However, many test substances in the regulatory testing realm are mixtures. It would have been useful to include some mixtures similar to common pesticide or household product formulations.

Also regarding the selection of reference substances used to determine the accuracy of the 3T3 and NHK test methods, there was an attempt to maintain the same proportion of "outliers" as was present in the RC. However, the total percentage of RC outliers in the set of reference substances (38%) was greater than the total percentage of outliers in the RC (27%). This should be highlighted and addressed as a potential confounder. Conversely, there was some concern that the potential for bias may exist if chemicals were pre-selected based on best fit to a regression line plotting cytotoxicity versus *in vivo* LD_{50} to evaluate *in vitro* test methods to estimate the acute oral LD_{50}. This bias likely predisposed the results to overprediction of the value of the NRU test methods for predicting random source chemicals. This potential bias needs to be discussed.

3.2 Rationale for the Number of Reference Substances Selected

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3 Characteristics of the Selected Reference Substances

3.3.1 *Source Databases Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.2 *Chemical Classes Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.3 *Product/Use Classes Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.4 *Toxicological Characteristics of the Selected Reference Substances*

Several confounding factors were addressed in the selection or evaluation of the reference substances (e.g., the octanol:water partition coefficient and the surface-active potential). These should be characterized and this information should be incorporated into the assessment.

Surface active molecules, in particular those that can partition at the oil-water interface, can significantly influence absorption, toxicity, and interactions with other molecules, and may enhance or diminish the predictive capacity of an *in vitro* test method. Test substance concentration and inherent toxic potential may be heavily influenced by molecular charge and surface activity.

Another example of a physical-chemical feature that can represent a confounding factor is given by the cationic amphiphilic molecules that contain a hydrophobic ring structure and a hydrophilic side chain with a charged cationic amine group. This chemical structure enables the substance to penetrate the cell membranes very rapidly and to interfere with phospholipid metabolism, causing phospholipidosis. This issue needs to be addressed.

3.3.5 *Selection of Reference Substances for Testing in Validation Study Phases Ib and II*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.6 *Unsuitable and Challenging Reference Substances*

The cytotoxicity endpoint for the test method is based on uptake of neutral red into lysosomes. The Draft BRD did not mention whether any of the reference substances cause lysosomal swelling, which could cause artifacts.

3.4 Reference Substance Procurement, Coding, and Distribution

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.5 Reference Substances Recommended by the *Guidance Document* (ICCVAM 2001b)

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.6 Summary

To the extent possible, characterization of the metabolic profiles of the reference substances should be added.

4.0 *In Vivo* Rodent Toxicity Reference Values Used to Assess the Accuracy of the 3T3 and NHK NRU Test Methods

This section described the problems that arise in finding and using rodent LD_{50} values taken from the published literature. These problems have been well known for decades (e.g., a review by Morrison et al. 1968) and little has improved since then as indicated by the lack of data collected under Good Laboratory Practice (GLP) guidelines. Given the shortcomings of the existing data, the information provided was adequate and revisions are unlikely to lead to any significant improvement.

The mechanisms of oral toxicity of the reference substances were difficult to determine because LD_{50} values are so rarely accompanied by more detailed information concerning the actual lesions observed and the reason for the animals' deaths. The overall poor characterization of modes or mechanisms of acute toxicity resulted in some difficulty in developing more sophisticated comparisons of *in vitro* and *in vivo* data.

4.1 Methods Used to Determine *In Vivo* Rodent Toxicity Reference Values

4.1.1 *Identification of Candidate In Vivo Rodent Toxicity Reference Data*

The selection of reference *in vivo* data was well described. A wide range of databases was searched and a comprehensive set of *in vivo* LD_{50} identified. In general, the actual data did not conform to modern standards of toxicity testing, hence their quality would be difficult to determine (99% - 452 of 459 LD_{50} values would have to be eliminated if a GLP requirement were to be mandated).

4.1.2 *Criteria Used to Select Candidate In Vivo Rodent Toxicity Data for Determination of Reference Values*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

4.2 Final *In Vivo* Rodent Toxicity Reference Values

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

4.3 Relevant Toxicity Information for Humans

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

4.4 Accuracy and Reliability of the *In Vivo* Rodent Toxicity Reference Values

Because many of the 72 reference substances had multiple LD_{50} values in the literature, these values had to be transformed to a single reference value for each chemical. The mean maximum:minimum values for those chemicals that had multiple LD_{50} values showed a tendency to decline as the toxicity decreased (See Table 4.4 of the Draft BRD). This may simply reflect the fact that inherent biological variability has a greater impact at low LD_{50} values than at high.

4.5 Summary

There was a general consensus that adequate data have been generated to draw conclusions about the accuracy and validity of the methods. The majority of the most relevant *in vivo* data from the available literature were collected to compare the two *in vitro* tests with *in vivo* acute toxicity in rodents.

5.0 3T3 and NHK NRU Test Method Data and Results

In general, the results section adequately presented the data and results. The statistical methods adopted provide a good quality analysis. However, several outcomes (indicated in the following subsections) were not adequately addressed.

5.1 3T3 and NHK NRU Test Method Protocols

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.2 Data Obtained to Evaluate Accuracy and Reliability

5.2.1 *Positive Control (PC) Data*

The Draft BRD should explain the considerably higher sensitivity of NHK cells to the positive control (sodium lauryl sulfate [SLS]).

5.2.2 *Reference Substance Data*

Consistently, carbon tetrachloride could not be tested in the 3T3 and NHK NRU test methods. The reason that this chemical could not be tested should be addressed. Several additional reference substances could not be adequately tested by one or two of the three laboratories, although they had used the same cell types and harmonized protocols. The reason(s) for these differences between the laboratories should be discussed.

5.3 Statistical Approaches to the Evaluation of 3T3 and NHK NRU Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.4 Summary of Results

Further discussion exploring the biological significance of and possible reasons for the differences in sensitivity and selectivity between the two cell lines is needed; this may be useful for selecting the appropriate cell line(s) for future use.

The significance of the steepness of the concentration-response curve was unclear from the data. The IC_{50} alone does not address this issue. While IC_{20} and IC_{80} (or at least a dose below and above the IC_{50}) were collected for most of the reference substances, they were not used in the analysis. The slope of the concentration-response curve should be included along with the IC_{50} data as additional information about the concentration-response characteristics.

The Draft BRD should include an explanation as to why 3T3 IC_{50} values for numerous reference substances were orders of magnitude different from those determined in the NHK test method. Was this due to cell-specific cytotoxicity? Or was it a consequence of differences in cell culture medium (i.e., presence or absence of serum)?

Table 5-4 in the Draft BRD was highly confusing. The column labeled "Difference (Orders of Magnitude)" contained the calculated ratios of the 3T3/NHK mean IC_{50} values. However, the column contained several mistakes. For example, potassium cyanide, with IC_{50} values of 34.6 vs. 29.0 μg/mL (ratio=1.2), has a *difference* of *1 order of magnitude* while parathion, 37.4 vs. 30.3 (ratio=1.2), has a *difference* of *0*. There were several more such cases (e.g., phenol, carbamazepine, nicotine). A more useful column to compare materials across the two NRU test methods would show the relative difference from the positive control. Since Table 5-5 uses some of the same data as Table 5-4, it must also be revised.
Noted in the summary but not discussed in Section 5.4 were the results in Table 5-4 showing that the IC_{50} values for aminopterin and digoxin differed by five orders of magnitude when tested in 3T3 versus NHK cells. Aminopterin and digoxin are established substrates for organic anionic transporters (OATs). Such transporters are very important for *in vivo* toxicity responses in terms of the ability of molecules to be absorbed, reach target tissues, accumulate, be excreted or secreted. Expression, induction, interference and binding to OATs can strongly influence the *in vivo* effects of a compound. Single nucleotide polymorphisms, which can strongly affect normal function, have been identified in human OATs. The differential susceptibility of the two studied cell lines could be explained by differential functioning of OATs between the cell types, but that was not examined or discussed. At least one publication indicated that NHK cells have at least five different OAT subclass members, with one shown to bind digoxin but not be constitutively expressed in the NHK, which could explain their sensitivity to this chemical. This issue needs to be addressed.

The summary indicated that the IC_{50} values were commonly (92%) within one order of magnitude of each other. A more descriptive and helpful summary would include the fraction

that was within specific IC_{50} ranges. For example, "for nine substances ratios between 3T3 IC_{50} values and NHK IC_{50} values were ≥10 or 0.1, respectively."

5.5 Coded Reference Substances and GLP Guidelines

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.6 Study Timeline and NICEATM/ECVAM Study Participatory Laboratories

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.7 Availability of Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.8 Solubility Test Results

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.9 Summary

One approach for comparing data generated on the same substance in different laboratories would be to normalize the data using the relative IC_{50} ratios between the reference substances and the positive control (at the level of the individual laboratory). This approach should be considered.

6.0 Accuracy of the 3T3 and NHK NRU Test Methods

This section adequately summarized the accuracy of the studies. The performance and limitations of the two NRU basal cytotoxicity tests were well defined. The overall accuracy for the prediction of Globally Harmonized System (GHS; UN 2005) acute oral hazard category was modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued. Further performance at the extremes of LD_{50} should be considered. Although some analysis of accuracy was conducted related to physical-chemical properties (e.g., solubility) and absorption, distribution, metabolism, and excretion (ADME) (e.g., biotransformation), and other factors (e.g., surface active properties, protein binding, receptor mediation) should be assessed to refine the test methods or draw greater precision by using a modular approach to define the types of materials suitable for the test methods.

Although there was not a significant difference between rat and mouse LD_{50} data (because of the variability of the data), separation was useful because it decreased the biological variability associated with species differences.

6.1 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Acute Oral Systemic Toxicity

Graphs should be added to compare the responses of the 58 RC substances to the same substances when tested using the 3T3 and NHK NRU test methods.

6.2 Improving the Prediction of *In Vivo* Rodent LD_{50} Values from *In Vitro* NRU IC_{50} Data

6.2.1 *The RC Rat-Only Regression in Millimolar Units*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.2.2 *The RC Rat-Only Regression in Weight Units*

Optimization of the IC_{50}-LD_{50} regression to allow for testing of mixtures was undertaken, yet no mixtures were used in fitting the regression curve. Since the test methods have limitations in accurately predicting the toxicity of materials with known or uncertain mechanisms, the testing of mixtures seems highly controversial.

6.2.3 *The RC Rat-Only Regression in Weight Units Excluding Substances with Specific Mechanisms of Toxicity*

It is true that many of the reference substances with underpredicted toxicity had mechanisms of toxicity that could not be expected to be detected in the 3T3 and NHK cell cultures; however, the Draft BRD incorrectly identified the mechanisms inactive in the cell cultures. The Draft BRD indicated that neurotoxic and cardiotoxic mechanisms, interference with energy utilization, and alkylation of macromolecules would not be active in the cell cultures. Interference with energy metabolism and alkylation of proteins and deoxyribonucleic acid (DNA) actually represent important mechanisms of cytotoxic action, which, in principle, should be detected by cytotoxicity assays such as the 3T3 and NHK NRU test methods. The rationale for excluding the 50 substances with *specific mechanisms of action* appears very questionable. Indeed, Table 6-2 of the Draft BRD shows that the linear regression between rodent LD_{50} values and IC_{50} values was not improved by the exclusion of these substances (R^2=0.353).

In addition, errors were made in the exclusion process based on the rules cited in the Draft BRD. For example, triethylene melamine and busulfan are both alkylating agents, but were not excluded. Paraquat and potassium cyanide were excluded based on interference with energy utilization. However, arsenic trioxide, which can uncouple oxidative phosphorylation, should have been excluded, but was not. Paraquat and potassium cyanide exert their acute systemic toxicity by means of cytotoxic action and should not have been excluded. If using a modular approach based upon proposed mechanisms (e.g., all substances interfering with energy metabolism), then hexachlorophene (a potent uncoupler of mitochondrial

phosphorylation), digoxin (a cardiac glycoside), or propanolol (a β-blocker) should have been included.

The Panel recommended against excluding reference substances based on mechanism given the numerous mechanisms of induction of cytotoxicity, the poor mechanistic understanding of the acute toxicity of many of these materials, and the incomplete knowledge of the appropriateness of the models for the individual modes/mechanisms of action.

6.3 Accuracy of the 3T3 and NHK NRU Test Methods for Toxicity Category Predictions

There was general consensus that adequate data were generated to draw conclusions about the accuracy and validity of the methods. The statistical approaches adopted to analyze data enable accurate and scientifically robust analyses of the two methods with regards to all their aspects.

The evaluation of the accuracy of the NRU basal cytotoxicity test methods for estimating GHS acute oral toxicity category was very extensive and detailed, and it identified areas of concern relative to specific chemical classes, chemicals with known mechanisms of toxicity and particular properties such as solubility, volatility, and so on. The evaluation of concordance of the observed and predicted GHS toxicity categories for each substance was performed correctly. Although a modular approach for using the model may be more reliable, the database was probably too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Since a mode of action is unlikely to be known about a random source material, it is also unlikely that a modular approach based upon mechanism will often be a viable option. A better approach would be a modular approach to validation based on chemical class, implying similar mode of action. Thus, the justification for the exclusion of 21 substances with specific modes of toxicity was not appropriate. The 26% accuracy for prediction of GHS class without removal of the 21 substances was poor, but better than a random selection using the 72 chemicals (1/6 accuracy).

Corrosivity was an exclusionary criterion intended to be applied to the selection of reference substances (see Section 3 of the Draft BRD). However, corrosive materials as a class were not subsequently deleted from the data when the regression curves were made. Corrosive chemicals are excluded from testing in *in vivo* acute toxicity tests because testing such chemicals *in vivo* is not appropriate, but using data for such chemicals in these analyses is acceptable.

For those classes of substances found to be appropriate for the assay, the NRU-based test methods may also be useful in a development context. During industry screening of new materials, a tool such as this may be useful to rank compounds belonging to the same chemical class (e.g., early lead optimization phase of drug development).

6.3.1 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Millimole Regression*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.2 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.3 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.4 *Summary of the Regressions Evaluated*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.4 Strengths and Limitations of the *In Vitro* NRU Test Methods for *In Vivo* Toxicity Prediction

Use of metabolically competent systems was recommended as one approach to improve the accuracy of *in vitro* predictions of acute toxicity; this should be explored in the future. The use of metabolizing systems is a general requirement for all *in vitro* tests for the prediction of genetic and carcinogenic potential and is considered necessary and scientifically justified. However, the contribution of metabolism of the reference substances was likely misstated, given the incomplete understanding of the acute toxicity of many of them. The substances listed in Table 3-7 of the Draft BRD, which were noted in the analysis of discordant substances, were highly variable in structure and purported mechanism. Of this set of substances, several (e.g., phthalates, valproic acid) may have active metabolites that contribute to their chronic toxicologic effects but which play little or no role in their acute toxicologic effects. Conversely, one may speculate that there may be substances *not* included in Table 3-7 of the Draft BRD for which active metabolism was an important component of its acute effects. Therefore, a more robust analysis of the contribution of metabolism to the accuracy of the models is recommended by incorporating a metabolic system into the *in vitro* assays.

As a future task, the properties of the cell lines (e.g., metabolism, receptors, transporters) that are important for basal cytotoxicity should be better characterized. Identified important properties could be used as performance standards.

6.5 Salient Issues of Data Interpretation

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.6 Comparison to Established Performance Standards

It would be informative to show comparisons of the RC LD_{50} values for the selected reference substances used in this study versus the individual laboratory responses for each test instead of the data shown in Figures 6-6 to 6-8 of the Draft BRD, which compares the *in vitro* responses to the overall RC millimole regression data.

While the mean IC_{50} values from one laboratory were generally higher than the rest, comparison to regressions with animal data (Appendix J) suggest there are no major differences between the laboratories in their ability to predict LD_{50} values. In fact the responses in Figures 6-6 to 6-8 look similar. When the *in vitro* response data from all laboratories with the agents selected from the RC are compared to the same agents for the RC, they provide a better correlation with the LD_{50} than did the overall RC data. Given this observation coupled with the variability in the data from animal studies, the data from the *in vitro* test methods would suggest that, as long as the appropriate controls (VC and PC) are used, the data from valid assays should be fairly predictive of animal response. It would be informative to show comparisons of the regression lines using the RC data for the 11 agents shown versus the individual laboratory responses for each test method instead of the data shown in Figures 6-6 to 6-8, which compares the *in vitro* responses to the overall RC millimole regression.

6.7 Summary

Protein binding should be taken into account in the data analyses. This parameter could be eventually taken into account in an additional data analysis (i.e., to the extent possible, consider the free fraction in serum corresponding to the LD_{50} dose). The Hill function slope data and LD_{50} slope data should be compared.

7.0 Reliability of the 3T3 and NHK NRU Test Methods

In general, the analyses in Section 7 adequately addressed the issues regarding both intra- and inter-laboratory reproducibility for the 3T3 and NHK NRU test methods. It was a little bit surprising, however, that some laboratories failed to obtain IC_{50} results for some of the reference substances. The Draft BRD should include an explanation or at least a discussion of these discrepancies, which may relate to the solvent protocol (discussed later). The compounds failing to yield IC_{50} values were mostly solvents (carbon tetrachloride, methanol, xylene, and 1,1,1-trichloroethane). Solvents are an important class of industrial substances for which Toxic Substances Control Act (TSCA) applies. The Draft BRD should offer an explanation if possible. Additional IC_{50} data are available for three of these substances: methanol (1000 mM), 1,1,1-trichloroethane (5.6 mM), and carbon tetrachloride (4.8 mM) using 3T3 cells after 24 hours of exposure (Gülden et al. 2005).

7.1 Substances Used to Determine the Reliability of the 3T3 and NHK NRU Test Methods

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

7.2 Reproducibility Analyses for the 3T3 and NHK NRU Test Methods

Additional consideration as to the underlying reasons for the variability between the laboratories would be helpful. The issue of intra- and inter-laboratory reproducibility due to variations in laboratory practices was addressed during the study and the findings indicated that the data from the two laboratories with GLP compliant procedures were in closer agreement and tended to show less variability and lower error rates than the other laboratory (which had an error rate of 93% for Phases 1a and 1b). Following a common training session for all laboratories, the interlaboratory variability decreased. This indicates the need for training in basic methodology and emphasis on protocol compliance. Everyone participating in such studies should be adequately trained in the basics of cell and tissue culture and sound scientific methods.

In order to increase the transparency of the comparison of the results from the different laboratories, an additional analysis of the IC_{50} data could be added: for each substance and NRU test method, the ratio between the highest and the lowest mean IC_{50} from the laboratories should be calculated. Those reference substances having ratios ≥ 3.0 should be presented in a separate table together with their calculated ratios and the names of the laboratories that delivered the corresponding IC_{50} values. From the Panel's analysis, it appeared that 17 substances for the 3T3 NRU test method and 11 substances for the NHK NRU test method had ratios ≥ 3.0. Extreme cases were cupric sulfate with a ratio of 22 (3T3 NRU test method) and digoxin with a ratio of 107 (NHK NRU test method). Furthermore, it became apparent that even for a simple compound such as sodium chloride, the results from different laboratories deviated by a factor of more than 3.0 for the NHK NRU test method.

It would be helpful to include a figure in the Draft BRD depicting all IC_{50} values for each test substance from all laboratories. Graphing of IC_{50} values plus-or-minus (±) the standard deviation (SD) and rat LD_{50} values ± SD should provide a better comparison of variation in the two sets of values.

It might also be helpful to look at ratios of the maximum IC_{50} values to the minimum IC_{50} values to see how they compare *vs.* rodent LD_{50} values. Given the variability in animal data where LD_{50} values (when more than one LD_{50} was available) could differ from 4 to 14 fold, the determination of a *precise* IC_{50} in each of the test methods to facilitate the selection of a starting dose does not seem necessary. Although the comparison of intra- and interlaboratory reproducibility for the purpose of validating the initial performance was appropriate, the use of multiple, costly test methods to identify *precise* IC_{50} values to establish initial doses for determining LD_{50} values seems counterproductive on the basis of cost and would limit acceptance of such methods.

For some of the reference substances, there was only one point and possibly even no points between 0 and 100% viability. These substances should be identified in the BRD.

NHK NRU IC_{50} data had a better correlation with human LC_{50} values (R^2=0.62) than did rodent 3T3 NRU IC_{50} data (R^2=0.51), as reported by Casati et al. (2005) at the 5th World Congress in Berlin in 2005. The correlation of NHK NRU IC_{50} data with human LC_{50} values (R^2=0.62) was also better than the correlation of rodent LD_{50} data with human LC_{50} values (R^2=0.56) (Casati et al. 2005). Discussion of this relationship should be considered for inclusion in the BRD.

7.2.1 *ANOVA Results for the 3T3 and NHK NRU Test Methods*

The Panel questioned the utility of the ANOVA for addressing the issue of intra- and inter-laboratory reproducibility. Depending upon the sample size and intralaboratory variation, a significant difference could correspond to a very small variation between laboratories or a non-significant difference could correspond to a very large difference between laboratories. Examples include parathion and procainamide. Parathion had reported IC_{50} values of 22.7, 141, and 22 µg/mL (p=0.014, not significant), and procainamide had reported IC_{50} values of 400, 431, and 497 µg/mL (p=0.007, significant). As a consequence, procainamide with satisfying, low interlaboratory reproducibility was included in Table 7-4 (because the ANOVA indicated significant laboratory differences) while parathion was not. There were more such examples that make the utility of the ANOVA questionable.

Based on the ANOVA analysis performed, FAL reported significantly different results from the two other laboratories for 20 substances (3T3 NRU test method). For 18 of these substances FAL reported the highest values. This phenomenon should be explained.

The statistically significant differences among the laboratories for 26 of the reference substances in the 3T3 NRU was worth noting, especially since it was greater than 1/3 of the agents tested. Volatility and/or presence of a precipitate were only noted for nine agents.

7.2.2 *CV Results for the 3T3 and NHK NRU Test Methods*

This section adequately elucidated associations between intra- or interlaboratory reproducibility and chemical classes, chemical properties, and potency categories. The result was that there were no clear associations between any of these parameters and CV values. What was evident, however, was that the reproducibility of both methods depends on the laboratory performing the measurements. A discussion of the possible reasons for this laboratory-specific reproducibility would be helpful.

7.2.3 *Comparison of Laboratory-Specific Linear Regression Analyses for the Prediction of In Vivo Rodent LD_{50} Values from In Vitro NRU IC_{50} Values*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

7.2.4 *Laboratory Concordance for the Prediction of GHS Acute Oral Toxicity Category*
The most important information given here was how often the data generated by the different laboratories would produce different starting doses for the ATC or UDP.

7.3 Historical Positive Control Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

7.4 Laboratory Concordance for Solvent Selection

Concern was raised about the differences in solvent selection between laboratories as compared to the BioReliance solvent information. For whatever reason, the variability between laboratories in the selection of solvent pointed out a possible flaw in the solvent determination protocol. This should be evaluated for future studies.

7.5 Summary

Irrespective of the statistical method used (ANOVA or calculation of the ratio between maximum and minimum IC_{50}), there were many more reference substances with deviating results between laboratories in the 3T3 NRU test method than in the NHK NRU test method. This should be explained.

8.0 3T3 and NHK NRU Test Method Data Quality

Section 8 adequately addressed the purpose of this section. No additional data are needed.

8.1 Adherence to Good Laboratory Practice Guidelines

8.1.1 *Guidelines Followed for In Vitro NRU Cytotoxicity Testing*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.1.2 *Quality Assurance (QA) for In Vitro NRU Cytotoxicity Test Data*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.1.3 *Guidelines Followed for In Vivo Rodent Oral LD_{50} Data Collection*
The use of the NRU test relied on the relationship between rat LD_{50} data and the observed IC_{50}. This relationship required reliable LD_{50} measurements for the RC substances used to construct the regression line. Since the LD_{50} values reported by the Registry of Toxic Effects for Chemicals Substances (RTECS®) were the most toxic found in the literature, one is unsure to what extent these LD_{50} estimates can be considered the *gold-standard*. These estimates may be appropriate for risk assessment but these extreme values can be unreliable and could lead to a misleading model of the desired linear relationship.

For comparative purposes with the IC_{50} values, LD_{50} values should reflect the variation observed. In most cases, a range of values should be shown. Such a range should reflect reasonable data with outliers omitted. If no range is shown, then a mean value (when available) plus-or-minus (±) SD should be used for the LD_{50}. The variability in animal data is usually much greater than that found *in vitro*. Therefore, comparing IC_{50} ± SD and Rat LD_{50} ± SD or data range should provide a better comparison. The Panel recommended that these data be shown in the report possibly in a bar graph similar to those in Figure 5-1. Based on the current data, it was not anticipated to have a major effect of the predictive potential of the two *in vitro* test methods. However, it could be important for future studies with other substances. The positive control response limits for a definitive test in Phase III was IC_{50} ± 2.5 SD. If the positive control showed this amount of variation, then why should the reference substances be expected to show any less? The test methods were not designed to predict hazard class but to predict starting animal dose in the acute LD_{50} tests.

8.2 Results of Data Quality Audits

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.3 Impact of Deviations from GLPs/Non-compliance

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.4 Availability of Laboratory Notebooks

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.5 Summary

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.0 Other Scientific Reports and Reviews of *In Vitro* Cytotoxicity Test Methods and the Ability of These Test Methods to Predict Acute Systemic Toxicity

In general, reports on other *in vitro* test methods using NRU were useful in providing insights into the correlation as well as the disparities between *in vitro* IC_{50} and *in vivo* LD_{50}. This was particularly true for the previously reported attempts to compare *in vitro* toxicity to *in vivo* lethality. However, it was less clear that the comparisons between eye irritation and NRU *in vitro* test methods were of use in interpreting the data used to compare *in vitro* IC_{50} to *in vivo* LD_{50}. While the mode of exposure is much more comparable between the *in vitro* test methods and the eye irritation (i.e., the test substance is applied directly to the target cell population), the endpoint is dissimilar. Furthermore, direct exposure of the target cells often

cannot adequately predict systemic effects, especially for some classes of substances that act through a known mechanism that does not relate to basal cytotoxicity.

Care was taken in the NICEATM/ECVAM study to cover a range of potencies and mode of action was also considered. It would be useful to compare the range of *in vivo* toxicities and modes of action represented in the other studies reported in Section 9 with the present NICEATM/ECVAM study.

9.1 Relevant Studies

9.1.1 *Correlation of In Vitro NRU Cytotoxicity Results with Rodent Lethality*

Additional discussion from the published literature about the advantages and limitations of using various supplemental metabolizing systems in cell culture for cytotoxicity testing could be included. For the Peloux *et al.* (1992) study, it may be worth including a discussion about the high correlation and whether the relatively good predictive value was a result of the route of exposure (i.e., intravenous [iv] and intraperitoneal [ip]). It should be clarified that the goodness of correlation for the *in vivo/in vitro* values for the different routes of exposure was iv>ip>oral and reflected different kinetic variables.

The results of the workshop presented in Seibert et al. 1996 should be included.

9.1.2 *Use of Cytotoxicity Data to Reduce the Use of Animals in Acute Toxicity Testing*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.1.3 *Other Evaluations of 3T3 or NHK NRU Test Methods*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.2 Independent Scientific Reviews

9.2.1 *Use of In Vitro Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing*

Clarification about the percentage reduction of animal use as referenced in the ICCVAM 2001a report should be included in Section 9 with the present ICCVAM study (i.e., what is the likely basis for the difference between then and now).

9.2.2 *Validation of 3T3 NRU for Phototoxicity*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.3 Studies Using *In Vitro* Cytotoxicity Test Methods with Established Performance Standards

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.4 Summary

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement)

The extent to which the NRU-based methods could contribute to a reduction in animal use was clearly discussed. The statistical analyses were clearly presented and the conclusions are clear. However, the Panel indicated that the extent to which the NRU test methods will reduce animal use for *in vivo* testing was not adequately characterized and discussed. The calculated savings (8-21%) of animals was only valid if several assumptions were accepted. For example, 21 of the 72 reference substances were excluded from the calculations because of their assumed specific modes of action. The best way to evaluate a possible reduction in animal use by using *in vitro* cytotoxicity to set the starting dose of an unknown substance is to assume that nothing is known about the mechanism(s) of toxicity of that test chemical. Therefore, all 72 reference substances should be included in the calculation of animal savings, regardless of their mode of action.

The use of the NRU cytotoxicity test methods are warranted not only if the number of animals used in the studies is reduced but also if the stress resulting from chemical exposure is minimized. The decision to use the NRU test to determine the starting dose for the ATC method or UDP is justified by the reduction in the number of animals required as indicated in the simulation studies.

The simulation studies compared the numbers of animals used with the starting dose indicated by the NRU basal cytotoxicity test method with the numbers of animals used with the default starting dose. Although the reduction in animals was not that great on a percentage basis, the testing of 4000 chemicals coming on the market in a year, could save 4000 rats at a rate of one rat per chemical. The Panel indicated, however, that a requirement to use the NRU test to determine the starting dose could lead to an increase in the number of animals required particularly if other data were available to provide a more accurate starting dose.

More information on the doses at which the reductions in expected animal numbers were found should be provided in the Draft BRD. Presumably, for the most toxic substances, the savings were at higher doses (as with the NRU test, the starting dose was below the default) and for the least toxic substances the savings were at the lower doses. The former are more important than the latter. For the most toxic substances, the largest savings in animal numbers was provided by the RC millimole regression. This was in contrast to the overall animal savings, which was smallest when this prediction is used. If the aim was to prevent animal suffering rather than to reduce animal numbers, then it appeared that the RC millimole regression was preferable.

10.1 Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Acute Systemic Toxicity Assays

This section should clarify that the NRU methods are to be used in a weight-of-evidence approach to determining the starting dose for acute oral systemic toxicity assays. Concern was expressed that underprediction of the toxicity by the cytotoxicity tests might lead to increased animal suffering. Although the accuracy for predicting the exact GHS category appears to be low, the data demonstrates that there is a reduction in animal use versus starting at the default starting dose if no other information is available (e.g., no toxicity information from chemicals with similar structure, etc.).

10.2 Reduction and Refinement of Animal Use for the UDP

Based on existing data, where molecular weight information was available for a relatively pure test substance, the millimolar regression should be used; in the absence of such data, the mg/kg regression should be used.

10.3 Reduction and Refinement of Animal Use for the ATC

The Panel found the discussion and evaluation of this section to be appropriate.

10.4 Summary

The Panel found the discussion and evaluation of this section to be appropriate.

The possibility of using the NRU test methods to determine the starting dose for the fixed dose procedure (FDP) acute toxicity test should be evaluated.

Animal savings should take into account, to the extent possible, the prevalence of chemicals in each GHS category.

11.0 Practical Considerations

Section 11 contained evaluations of potential expense to be incurred upon approval and required implementation of these procedures to aid in choosing the starting dose for a UDP or other type of rat oral toxicity study. However, a cost-benefit analysis was absent. In order to reduce the animal usage per acute oral toxicity study by approximately 1-2 rats, the estimated cost to sponsors increased by $1000-2000 for the preliminary *in vitro* study. This is not cost-effective. Obviously, additional time would be required also to complete the oral toxicity evaluation. Furthermore, although it was said that defining a starting dose to more closely coincide with the actual LD_{50} of a test substance improves the ultimate LD_{50} estimate, many regulatory tests are limit tests for which a preliminary *in vitro* test would offer no benefit.

11.1 Transferability of the 3T3 and NHK NRU Test Methods

It appears that transferability was not as easy as was stated; minor protocol differences can have profound effects. Adequate training must be conducted prior to the initiation of the study, and a demonstration of proficiency in running the test must be demonstrated before testing unknowns.

11.1.1 *Facilities and Major Fixed Equipment*

A dedicated cell culture laboratory should be added to the list of needs.

11.1.2 *Availability of Other Necessary Equipment and Supplies*

A single source for NHK medium was noted to be a problem in the NICEATM/ECVAM validation study.

Although the Draft BRD indicated that laboratories could isolate keratinocytes from donated cultures, this could increase intralaboratory variation. The Panel agreed that the recommendation for a commercial source is better.

The Draft BRD should indicate that it is necessary to confirm that cells are free from contamination (e.g., bacteria, mycoplasma).

11.2 3T3 and NHK NRU Test Method Training Considerations

11.2.1 *Required Training and Expertise*

This section noted that good cell culture practices are needed. The Panel recommended removing statements about the need for training in cloning, transfection, expression cloning, immortalization, and virus propagation since these techniques are not necessary for cytotoxicity testing.

11.2.2 *Training Requirements to Demonstrate Proficiency*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

11.3 Test Method Cost Considerations

11.3.1 *3T3 and NHK NRU Test Methods*

The Panel indicated that the costs quoted may be more than a little bit low. The Draft BRD noted that it was possible that there wouldn't be cost savings using NRU testing first, if only a few rats were used. Additionally, the NHK NRU test could be almost cost-prohibitive if 5 x $380 vials are needed per 5 x 96-well plates.

The costs of performing NRU testing were charges from commercial laboratories. A rough calculation for the cost of equipment and time need to perform each test might help individual laboratories understand the cost and time of performing the test methods.

11.3.2 *In Vivo Rodent Acute Oral Toxicity Testing*

Since the NRU test methods are to be used for reduction of animal use rather than replacement, it is appropriate to describe the number of animals that might be reduced in this section.

11.4 Time Considerations for the 3T3 and NHK NRU Test Methods

Since it takes some time to screen the NHK NRU assay medium, it should be described in this section.

11.5 Summary

The commentaries in Section 11 appeared to be appropriate. It was difficult to compare the value of the *in vitro* NRU test method ($1120-$1850) per test substance to achieve an IC_{50} versus an animal test ($750-$3750) to achieve an LD_{50}. If the *in vitro* test can save at least a single animal in the execution of the ATC or UDP test, this evaluation was worth the effort.

VALIDATION STATUS OF THE NRU TEST METHODS

The Panel agreed that the applicable validation criteria have been adequately addressed for using these *in vitro* test methods in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. However, the Panel was aware that validation of the two NRU test methods was carried out not only to determine if they could be used to set starting doses for *in vivo* acute toxicity studies, but also to determine the extent to which these tests could be a useful step in an *in vitro* tiered testing strategy for acute toxicity. The Panel agreed the validation study showed the two NRU test methods evaluated could not be used as a stand-alone replacement for the *in vivo* tests even considering the variability of the latter. The Panel encouraged future work to develop a tiered testing strategy that includes basal cytotoxicity as part of the overall strategy.

[This Page Intentionally Left Blank]

DRAFT ICCVAM RECOMMENDATIONS FOR *IN VITRO* ACUTE TOXICITY TEST METHODS

[This Page Intentionally Left Blank]

1.0 Draft ICCVAM Recommendations for *In Vitro* Acute Toxicity Test Methods

1.1 Recommended Test Method Uses

1. The 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of hazard classification (see Section 6 of the *In Vitro* Acute Toxicity Test Methods BRD).

    - The Panel agreed with this statement in that neither of the two basal cytotoxicity tests can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification.
    - In the Draft BRD, the rat *in vivo* data did not conform to current GLP standards.

2. For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral *in vivo* toxicity protocols (i.e., the Up-and-Down Procedure [UDP] and Acute Toxic Class [ATC]).

    - The Panel agreed that the *in vitro* test methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols.
    - Given the test methods' limited predictive capacity, however, it was unclear whether they will provide substantial weight in that decision.
    - The overall accuracy was modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued.

3. Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education (National Research Council 1996), and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002)[3], *in vitro* basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral *in vivo* toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed**.**

    - The Panel agreed**.**

[3] National Research Council. 1996. Guide for the Care and Use of Laboratory Animals. Washington, DC: National Academy Press.
PHS. 2002. Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals.

4. Substances with specific toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic, cardiotoxic, interfere with energy utilization, or alkylate proteins and other macromolecules) will likely be underpredicted by these *in vitro* basal cytotoxicity test methods. Therefore, until such time as a more predictive testing approach is developed, the results from basal cytotoxicity testing with such substances may not be appropriate.

    - The Panel disagreed with elements of this statement; specific toxic mechanisms that are not expected to be active in 3T3 and NHK cells, such as "interference with energy utilization and alkylation of proteins and other macromolecules", are mechanisms of cytotoxic action and should be detectable with 3T3 and NHK cells.

5. The regression formula used to determine starting doses should be the revised Registry of Cytotoxicity (RC) regression line [with IC_{50} values in μg/mL and LD_{50} values in mg/kg] developed with the RC chemicals using rat LD_{50} data only and excluding chemicals with mechanisms of action that are not expected to be active in *in vitro* basal cytotoxicity test methods.

    - The Panel did not agree with this statement.
    - There was consensus among the Panel that the data contained within the Draft BRD or the open literature were not sufficient to justify the exclusion of reference substances based on mechanism.
    - It was not justified to (retrospectively) exclude substances because of assumed modes of toxic action *in vivo* and/or possible involvement of biotransformation reactions.

6. The performance of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

    - The Panel agreed with this statement although the reliability of the test methods in this study was not quite satisfying (e.g., inter-laboratory reproducibility), the reproducibility of these methods (e.g., intra-laboratory reproducibility) was modest, and the accuracy of these methods was poor.

7. Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU cytotoxicity test method is recommended for general use.

    - Some Panel members agreed in a general sense, however, cautioned that one model be preferred over the other, based upon specific knowledge

regarding known mechanisms of action (e.g., the rationale for the disparate results observed with aminopterin and digoxin). Other Panel members agreed with this statement because the use of continuous cell lines is more efficient, especially since the overall animal savings were relatively low.

- One Panel member noted that NHK NRU IC_{50} data have shown a better correlation with human LC_{50} values (R^2=0.62) than do rodent 3T3 NRU IC_{50} data (R^2=0.51) and better than rodent LD_{50} data correlates with human LC_{50} values (R^2=0.56) as reported by Casati et al. (2005) at the 5th World Congress in Berlin. It is important to remember that hazard assessment relates to the safety of humans, not rats.
- Based on costs of commercial keratinocytes, the NHK NRU test method may be cost-prohibitive.
- The proprietary nature of the composition of the NHK culture medium made it impossible to assess the role differences in media composition may have had on the results.

1.2 Draft Recommended Test Method Limitations

- Colored substances (besides red substances) may absorb light in the optical density range of the NRU test methods, which could affect the accuracy of the results.
- The Draft BRD indicated that optimization to allow for testing of mixtures was being undertaken, yet no mixtures were used in fitting the regression curve. Given the limitations of the test methods in accurately predicting materials of known or uncertain mechanisms, the testing of mixtures seems highly controversial.

1.3 Draft Recommended Future Studies

1. Additional data should be collected using the 3T3 and/or the NHK NRU test methods to evaluate their usefulness for predicting the *in vivo* acute oral toxicity of chemical mixtures.

    - The Panel generally agreed that this is a good recommendation, although collecting data could be difficult and doing a correlation with *in vivo* data would be even more difficult. It may be useful to suggest that such data only be collected with the 3T3 NRU test method, and that it would be necessary to clarify the reasons for the interlaboratory variations for future use of the method.

2. Additional high quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to supplement the high quality validation database started by this study. Periodic evaluations of the expanded database should be conducted to further

characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

- The Panel agreed this could be valuable under certain conditions, especially if NRU data were collected as acute toxicity testing is conducted.
- However, no reviewer wanted *in vivo* testing conducted solely to collect data to assess the usefulness of the NRU test method, particularly given that the savings in animal numbers that arise from the use of the NRU test method to determine the starting dose for the ATC method or UDP are fairly modest.

3. Additional efforts should be conducted to identify additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification; specifically, studies should be conducted to investigate the potential use of *in vitro* cell-based test methods that incorporate mechanisms of action and evaluations of ADME (absorption, distribution, metabolism, excretion) to provide improved estimates of acute toxicity hazard categories.

   - The Panel agreed with this statement and added that there should be additional effort towards development of alternative methods to adequately predict the *in vivo* acute toxicity of chemicals for the purposes of hazard classification.
   - An additional statement to include could be, "and the development of methods to extrapolate from *in vitro* toxic concentrations to equivalent doses *in vivo*."

4. The *in vivo* database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral systemic toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR]).

   - The Panel agreed with this recommendation.

5. Standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included in future *in vivo* rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* methods.

   - The Panel agreed with this recommendation; this is really important and could further the development of non-animal alternatives in the future.
   - To facilitate comparisons and model development, future studies should incorporate high quality animal data for required testing of new substances, blood levels from animals (LC_{50}) (where possible), and high quality *in vitro* data for the same substances.

- To aid in this process, the Panel recommended that an expert group be convened to identify appropriate *in vivo* endpoints.
- The Panel recommended also that ICCVAM consider convening a working group to explore mechanisms of action of acute toxicity, and approaches to acquiring additional information on acute toxic mechanisms when conducting the required *in vivo* acute toxicity testing.
- Although a modular approach may be more reliable, the database was likely too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Since a mode of action is unlikely to be known about a random source material, it is also unlikely that a modular approach based upon mechanism will be a viable option. A better approach to validation is one based on chemical class, implying similar mode of action.

6. An expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test method development and validation studies.

- The Panel agreed with this recommendation; there should be a concerted effort to collect proprietary data.

[This Page Intentionally Left Blank]

APPENDIX A: DRAFT PERFORMANCE STANDARDS FOR *IN VITRO* ACUTE TOXICITY METHODS

[This Page Intentionally Left Blank]

1.0 Purpose and Background of Performance Standards

The available data from this study appeared to support the validity of the recommended performance standards for the test methods. The usefulness and limitations were well covered, and if validated, the methods may be a worthwhile option. However, there may be some cause for concern if use of the methods is made compulsory for regulatory purposes.

1.1 Introduction

The Panel found the discussion and evaluation of this section to be appropriate.

1.2 Elements of ICCVAM Performance Standards

The Panel found the discussion and evaluation of this section to be appropriate.

1.3 ICCVAM Process for the Development of Performance Standards

The Panel found the discussion and evaluation of this section to be appropriate.

1.4 ICCVAM Development of Recommended Performance Standards for *In Vitro* Acute Toxicity Test Methods

The Panel found the discussion and evaluation of this section to be appropriate.

2.0 *In Vitro* Acute Toxicity Test Methods

The Panel found the discussion and evaluation of this section to be appropriate.

2.1 Background

The Panel found the discussion and evaluation of this section to be appropriate.

2.2 Principles of *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

The Panel found the discussion and evaluation of this section to be appropriate.

2.3 Essential Test Method Components for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity (Lethality) Tests

A discussion is needed in this section regarding whether or not the NRU test methods are recommended for use with unknown substances and mixtures. The recommendations made in Section 2.3.2 (Application of the Test Substances), Section 2.3.3 (Control Substances), and Section 2.3.4 (Viability Measurements) were acceptable.

2.4 Reference Substances for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

The significance of the secondary chemical subset to be used for *investigational purposes* should be better elucidated.

2.5 Accuracy and Reliability

The Panel found the discussion and evaluation of this section to be appropriate.

APPENDIX B:
DRAFT RECOMMENDED TEST METHOD PROTOCOLS

[This Page Intentionally Left Blank]

1.0 Draft Recommended Test Method Protocols

The protocols were generally quite detailed and laboratory technicians should be able follow the procedures. The Panel recommended the following clarifications be added to the 3T3 and NHK NRU test method protocols:

1.1 Protocol Recommendations

- The rationale for testing the positive control on separate plates rather than on the test plates should be provided.
- The number of definitive tests that should be performed for a test substance should be specified.
- The range of linearity of the microplate reader should be confirmed (as per in-house SOPs) for the recommended optical density (OD_{540}) and stated.
- Maximum absorbance values needed by a spectrophotometric plate reader should be provided for application to the NRU test methods.
- The test method protocols should be streamlined. (Undefined is how this should be accomplished.)
- Guidance for using methods other than the Hill function to determine IC_{50} values should be provided.
- The lowest acceptable test substance dilution factor (i.e., 1.21) should be reduced rather than accepting only one cytotoxicity point between 0 and 100% viability on a steep dose-response curve to use for determination of the IC_{50} value.
- Study directors and quality assurance units are necessary only if testing is performed under Good Laboratory Procedures (GLP), which is not usually necessary for dose-setting tests.
- The protocol for the NHK cells should include a statement about the need to avoid allowing the cell to reach confluence: under these conditions, these cells can exhibit contact-induced differentiation. Once differentiation is induced, cells lose their ability to proliferate.

1.2 Cell Culture Recommendations

- Good cell culture practices (e.g., Hartung et al. 2002) must be followed.
- Whether or not a prequalification test of new keratinocytes should be performed by the laboratory prior to actual testing should be stated.
- A recommendation that keratinocytes should be procured only through commercial sources and not by preparing primary cultures from donated tissue should be included.

1.3 Solubility Recommendations

- Additional guidance to the solubility step-wise procedure should be added (i.e., ensure that test substance solution preparation procedures can be easily understood by laboratory technicians).

- Include a recommendation for training laboratory technicians so they better understand solvent and solubility determinations.
- Additional guidance as to the use of a microscope to assist in determining solubility of a test substance should be added.
- Test substances that may etch plastic or *film out* in medium should be identified (the importance of detecting such compounds by the laboratory technicians should be emphasized).
- The protocols should recommend the use of a solvent (e.g., dimethylsulfoxide [DMSO], ethanol) at its lowest possible concentration.
- There was concern about the differences in solvent selection between laboratories as compared to the BioReliance solvent information. The variability between laboratories in the selection of solvent points out a possible flaw in the solvent determination protocol. This should be evaluated for future studies.

REFERENCES

Autian J, Dillingham EO. 1978. Overview of general toxicity testing with emphasis on special tissue culture tests. In: *In Vitro* Toxicity Testing 1975-1976 (Berky J, Sherrod C, eds). Philadelphia:Franklin Institute Press, 23-49.

Casati S, Hoffmann S, Strickland J, Paris M, Stokes W, Tice R, Malerba I, Hartung T. 2005. Anaylsis of the correlation between *in vitro* cytotoxicity data and acute toxic effects in humans. 5th World Congress on Alternatives and Animal Use in the Life Sciences, Berlin, Germany.

Elmore E, Luc TT, Steele VE, Redpath JL. 2001. Comparative tissue-specific toxicities of 20 cancer preventive agents using cultured cells from 8 different normal human epithelia. In Vitr Mol Toxicol 14:191-207.

Elmore E, Siddiqui S, Desai N, Moyer MP, Steele VE, Redpath JL. 2002. The human epithelial cell cytotoxicity test method for determining tissue specific toxicity: method modifications. Methods Cell Sci 24:145-153.

Fentem J, Fry J, Garle M, Gülden M, Seibert H, Voss J-U, Wassermann O, Perchermeier M, Wiebel F. 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/B4-3063-4086 & B92/B4-3040/14087. FRAME Nottingham.

Gülden M, Mörchel S, Seibert H. 2005. Comparison of mammalian and fish cell line cytotoxicity: impact of endpoint and exposure duration. Aquat Toxicol 71:229-236.

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, Lewis D. 2002. Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. Altern Lab Anim 30:407-414. Available: http://ecvam.jrc.it/publication/index5007.html. [accessed 30 May 2006].

ICCVAM. 2001a. Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 29 May 2006].

ICCVAM. 2001b. Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

Morrison JK, Quinton RM, Reinert M. 1968. The purpose and value of LD_{50} determination. In: Modern Trends in Toxicology, Vol 1, (Boyland E, Goulding R, eds), Chichester:John Wiley & Sons, 1-17.

Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, Ekwall B, Garle MJ, Gómez-Lechón MJ, Gribaldo L, Gülden M, Liebsch M, Rasmussen E, Roguet, Shrivastava R, Walum E. 1996. Acute toxicity testing *in vitro* and the classification and labeling of chemicals: The report and recommendations of ECVAM Workshop 16. Altern Lab Anim 24:499-510.

UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and Geneva. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html [accessed 12 September 2005].