NIH Publication No: 07-4518

BACKGROUND REVIEW DOCUMENT

In Vitro Cytotoxicity Test Methods for Estimating Acute Oral Systemic Toxicity

Volume 1 of 2

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

THE INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS and THE NTP INTERAGENCY CENTER FOR THE EVALUATION OF ALTERNATIVE TOXICOLOGICAL METHODS

In 1997, the National Institute of Environmental Health Sciences (NIEHS), one of the National Institutes of Health (NIH), established the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to:

- Coordinate interagency technical reviews of new and revised toxicological test methods, including alternative test methods that reduce, refine, or replace the use of animals

- Coordinate cross-agency issues relating to validation, acceptance, and national and international harmonization of new, modified, and alternative toxicological test methods

On December 19, 2000, the ICCVAM Authorization Act (42 U.S.C. § 2851-2, 2851-5 [2000]) established ICCVAM as a permanent interagency committee of NIEHS under the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

ICCVAM is comprised of representatives from 15 U.S. Federal regulatory and research agencies that use, generate, or disseminate toxicological information. ICCVAM conducts technical evaluations of new, revised, and alternative methods with regulatory applicability. The Committee promotes the scientific validation and regulatory acceptance of toxicological test methods that more accurately assess the safety or hazards of chemicals and products and that refine (decrease or eliminate pain and distress), reduce, and/or replace animal use. NICEATM administers ICCVAM and provides scientific and operational support for ICCVAM-related activities. More information about ICCVAM and NICEATM can be found on the ICCVAM/NICEATM web site (http://iccvam.niehs.nih.gov) or obtained by contacting NICEATM (telephone: [919] 541-2384, e-mail: iccvam@niehs.nih.gov).

The following Federal regulatory and research agencies are ICCVAM members:

Consumer Product Safety Commission
Department of Agriculture
Department of Defense
Department of Energy
Department of Health and Human Services
- Agency for Toxic Substances and Disease Registry
- Food and Drug Administration
- National Cancer Institute
- National Institute of Environmental Health Sciences
- National Institutes of Health, Office of the Director
- National Institute of Occupational Safety and Health
- National Library of Medicine

Department of the Interior
Department of Labor
- Occupational Safety and Health Administration

Department of Transportation
Environmental Protection Agency

On the Cover
The ICCVAM/NICEATM graphic symbolizes the important role of new and alternative toxicological methods in protecting and advancing the health of people, animals, and our environment.

In Vitro Cytotoxicity Test Methods for Estimating Acute Oral Systemic Toxicity

Background Review Document

Volume 1 of 2

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

National Institute of Environmental Health Sciences (NIEHS)
P.O. Box 12233
Mail Drop: EC-17
Research Triangle Park, NC 27709

November 2006
NIH Publication No. 07-4518

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

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

Printed: 11/30/2006

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

A-CUTE-TOX	A-Cute-Tox Project (EU Research & Development Integrated Project)
ADME	Absorption, distribution, metabolism, and elimination
ANOVA	Analysis of variance
ASTDR	Agency for Toxic Substances and Disease Registry
ASTM	American Society for Testing and Materials
ATC	Acute Toxic Class method
ATCC	American Type Culture Collection
ATWG	Acute Toxicity Working Group
BBB	Blood:brain barrier
BPE	Bovine pituitary extract
BRD	Background Review Document
^{o}C	Degrees Celsius
CAS	Chemical Abstracts Service
CASRN	Chemical Abstracts Service Registry Number
CCOHS	Canadian Centre for Occupational Health and Safety (CCOHS)
CDER	U.S. FDA Center for Drug Evaluation and Research
CESARS	Chemical Evaluation Search and Retrieval System
CFU	Colony forming units
CHRIS	Chemical Hazard Response
CI	Confidence interval
CICADS	Concise International Chemical Assessment Documents
CIS	ILO Occupational Safety and Health Information Centre
CNS	Central nervous system
COLIPA	The European Cosmetic Toiletry and Perfumery Association
CPSC	U.S. Consumer Product Safety Commission
CSF	Colony stimulating factor
CTFA	Cosmetic, Toiletries and Fragrance Association
CV	Coefficient of variation
DART®/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information Center
DEA	U.S. Drug Enforcement Administration
DHHS	U.S. Department of Health and Human Services
DIMDI	Deutsches Institut fur Medizinische Dokumentation und Information (The German Institute for Medical Documentation and Information)
DNA	Deoxyribose nucleic acid
DMEM	Dulbecco's Modification of Eagle's Medium
DMSO	Dimethyl sulfoxide
D-PBS	Dulbecco's phosphate buffered saline
DOD	U.S. Department of Defense
DOT	U.S. Department of Transportation
EC	European Commission
EC_{50}	Concentration of a substance that produces 50% of the maximum possible response for that substance

ECBC	U.S. Army Edgewood Chemical Biological Center
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EC/HO	European Commission/British Home Office
ECVAM	European Centre for the Validation of Alternative Methods
EDIT	Evaluation-guided development of new *in vitro* tests
EHC	Environmental Health Criteria
EHS	EPA's Extremely Hazardous Substance list
EPA	U.S. Environmental Protection Agency
ERG	Emergency Response Guidebook
ETOH	Ethanol (Ethyl alcohol)
EU	European Union
EXTONET	The Extension Toxicology Network
FAL	FRAME Alternatives Laboratory
FAO	UN Food and Agriculture Organization
FB1	Fumonisin B1
FDA	U.S. Food and Drug Administration
FDP	Fixed Dose Procedure
FIFRA	U.S. Federal Insecticide, Fungicide, and Rodenticide Act
FR	Federal Register
FRAME	Fund for the Replacement of Animals in Medical Experiments
GABA	Gamma amino butyric acid
GCCP	Good cell culture practices
GHS	Globally Harmonized System (of Classification and Labeling of Chemicals)
GLP	Good Laboratory Practices
gm	Grams
HBSS	Hanks' balanced salt solution
HPV	High Production Volume
hr	Hour(s)
HSDB	Hazardous Substances Data Bank
HSG	Health and Safety Guides
HTD	Highest tolerated dose
IARC	International Agency for Research on Cancer
IC_{20}	Concentration producing 20% inhibition of the endpoint measured
IC_{50}	Concentration producing 50% inhibition of the endpoint measured
IC_{80}	Concentration producing 80% inhibition of the endpoint measured
ICCVAM	Interagency Coordinating Committee for the Validation of Alternative Methods
ICSC	International Chemical Safety Cards
ID	Insufficient data
ID_{50}	Index of cytotoxicity; dose producing a 50% reduction in protein value
IIVS	Institute for *In Vitro* Sciences
ILO	International Labour Organisation
i.m.	Intramuscular
INVITOXX	*In Vitro* Techniques in Toxicology (ERGATT FRAME ECVAM Data bank)

IOM	Institute of Medicine
i.p.	Intraperitoneal
IPCS	International Programme on Chemical Safety
IRAG	Interagency Regulatory Alternatives Group
IRPTC	International Register of Potentially Toxic Chemicals
ISO	International Standards Organization
IUCLID	International Uniform Chemical Information Database
i.v.	Intravenous
JECFA	Joint Expert Committee on Food Additives
JMPR	Joint Meeting on Pesticide Residues
KBM®	Keratinocyte basal medium
kg	Kilogram
K_{ow}	Octanol-water partition coefficient
L	Liter
LC	Lethal blood concentration
LD_{50}	Dose that produces lethality in 50% of test animals
LDH	Lactate dehydrogenase
MAS	Maximum average Draize score
MEIC	Multicentre Evaluation of *In Vitro* Cytotoxicity
MeSH®	Medical Subject Heading
µL	Microliters
µm	Micrometers
µM	Micromoles
mg	Milligram
MIT	Metabolic inhibition test
mL	Milliliter
mM	Millimolar
MMAS	Modified maximum average score
mmol	Millimoles
MPE	Mean photo effect
MSDS	Material Safety Data Sheets
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
N	Number (of substances)
NA	Not applicable
NADH	Nicotine adenine dinucleotide (reduced)
NC	Not calculated
NCS	Newborn calf serum
NCTR	U.S. FDA National Center for Toxicological Research
n.d.	Not detectable
NHK	Normal human epidermal keratinocytes
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIEHS	U.S. National Institute of Environmental Health Sciences
NIH	U.S. National Institutes of Health
NIOSH	U.S. National Institute for Occupational Safety and Health
NLM	National Library of Medicine

NR	Neutral red
NRU	Neutral red uptake
NTP	U.S. National Toxicology Program
OAT	Organic anionic transporters
OD	Optical density
OD_{540}	Optical density (absorbance) at a wavelength of 540 nm
OECD	Organisation for Economic Co-operation and Development
OHM/TADS	EPA Oil and Hazardous Materials/Technical Assistance Data System
OPP	U.S. EPA Office of Pesticide Programs
OPPTS	EPA Office of Prevention, Pesticides and Toxic Substances
ORD	U.S. EPA Office of Research and Development
OSHA	U.S. Occupational Safety and Health Administration
OTA	Ochratoxin A
PBS	Phosphate buffered saline
PC	Positive control
PDS	Pesticide Data Sheets
pg	Picogram
PG	Packing group
PIF	Photoinhibition factor
PIMS	Poisons Information Monographs
pK	Acid/base dissociation constant
PLS	Partial Least Squares (analysis)
PPIS	EPA Pesticide Product Information System
PPT	Precipitate
QA	Quality assurance
QC	Quality control
R^2	Coefficient of determination
r_s	Spearman correlation coefficient
RC	Registry of Cytotoxicity
REACH	Registration, evaluation, authorisation and restriction of chemicals
RTECS®	Registry of Toxic Effects of Chemical Substances
RTK NET	The Right-to-Know Network
SD	Standard deviation
SIDS	OECD Screening Information Data Sets
SIS	Scientific Information Service
SLS	Sodium lauryl sulfate
SMT	Study management team
SOP	Standard operating procedure
3T3	BALB/c mouse fibroblasts, clone A31 (ATCC # CCL-163)
TESS	Toxic Exposure Surveillance System
TG	Test guideline
TRI	U.S. EPA Toxics Release Inventory
TSCA	Toxic Substances Control Act
UDP	Up-and-Down Procedure
UN	United Nations

UNEP	United Nations Environment Programme
USP	U.S. Pharmacopoeia
UV	Ultraviolet (light)
VC	Vehicle control
WHO	World Health Organization
XTT	2,3-Bis(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide
ZEBET	Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch (German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments)

ACKNOWLEDGMENTS

The following individuals are acknowledged for their contributions to the in vitro acute toxicity test method review process.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Working Group (ATWG)

Consumer Product Safety Commission (CPSC)

Kailash Gupta, D.V.M., Ph.D.
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D. (ATWG Chair, ICCVAM Vice Chair)

Department of Energy (DOE)

Po-Yung Lu, Ph.D.

Environmental Protection Agency (EPA)

Karen Hamernik, Ph.D.
Masih Hashim, Ph.D.
Marianne Lewis
Elizabeth Margosches, Ph.D.
Deborah McCall
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration (FDA)

Leonard Schechtman, Ph.D. (ICCVAM Chair)
Kenneth Hastings, Ph.D.
Abigail Jacobs, Ph.D.
David Morse, Ph.D.
Thomas Umbreit, Ph.D.

National Institute for Occupational Safety & Health (NIOSH)

Steven Reynolds, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

Rajendra Chhabra, Ph.D., D.A.B.T.
William Stokes, D.V.M., D.A.C.L.A.M. (ICCVAM Executive Director; NICEATM Director)
Raymond Tice, Ph.D. (NICEATM Deputy Director)

European Centre for the Validation of Alternative Methods (ECVAM) Liaisons

Silvia Casati, Ph.D.
Pilar Prieto, Ph.D.

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

Dave Allen, Ph.D.
ILS, Inc.

Bradley Blackard, M.S.P.H.
ILS, Inc.

Sue Brenzel
ILS, Inc.

Thomas Burns, M.S.
ILS, Inc.

Patricia Ceger, M.S.
ILS, Inc.

Jeffery Charles, Ph.D., M.B.A., D.A.B.T.
ILS, Inc.

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

William Stokes, D.V.M., D.A.C.L.A.M. (Director)
NIEHS

Judy Strickland, Ph.D., D.A.B.T.
ILS, Inc.

Raymond Tice, Ph.D. (Deputy Director)
NIEHS

Jim Truax, M.A.
ILS, Inc.

Participants in the *In Vitro* Cytotoxicity Validation Study

NICEATM gratefully acknowledges the generous contributions of the individuals who directly participated in the NICEATM/ECVAM Validation Study. Their time and efforts are greatly appreciated.

BioReliance Corp.
(Chemical Distribution)
Martin Wenk, Ph.D. – Principal Investigator

Institute for *In Vitro* Sciences (IIVS)
(Lead Laboratory – Protocols)
Hans Raabe, M.S. – Study Director
Greg Mun – Laboratory Manager
Angela Sizemore – Research Technician
Gregory O. Moyer – Research Technician
John Harbell, Ph.D. – Scientific Director

U.S. Army Edgewood Chemical Biological Center (ECBC)
(Testing Laboratory)
Cheng Cao, Ph.D. – Study Director
Janna Madren-Whalley – Research Technician
Chundakkadu Krishna, Ph.D. – Research Technician
James J. Valdes, Ph.D. – Scientific Advisor

FRAME Alternatives Laboratory (FAL)
University of Nottingham, UK
(Lead Laboratory – Software)
Richard Clothier, Ph.D. – Study Director
Nicola Bourne – Research Technician
Monika Owen – Research Technician
Rachel Budworth – Research Technician

Constella Group (Statistical Analyses)
Patrick Crockett, Ph.D.
Eric Harvey, Ph.D.
Wendell Jones, Ph.D.
Robert Lee, M.S.
Jessica L. Matthews, M.S.
Michael Riggs, Ph.D.
Janine Wilcox
Nicole Williams

Statistical Consultant
Joseph Haseman, Ph.D.

NIEHS
Grace Kissling, Ph.D. – Contract Project Officer
Molly Vallant - Contract Project Officer

Study Management Team (SMT)
NICEATM
William Stokes, D.V.M., D.A.C.L.A.M. (NIEHS) – Director, NICEATM
Raymond Tice, Ph.D. (NIEHS) – Deputy Director, NICEATM – Advisor
Judy Strickland, Ph.D., D.A.B.T. (ILS, Inc.) – Project Coordinator
Michael Paris (ILS, Inc.) – Assistant Project Coordinator
Jeffrey Charles, Ph.D., D.A.B.T. (ILS, Inc.) – Advisor

ECVAM
Thomas Hartung, Ph.D., M.D. – Head of Unit from 2002
Silvia Casati, Ph.D. – Task Leader
Michael Balls, D. Phil.– Head of Unit until June 2002

[This Page Intentionally Left Blank]

PREFACE

The Institute of Medicine estimates that more than 4 million poisonings occur annually in the United States (Institute of Medicine [IOM] 2004). In 2001, 30,800 deaths placed poisoning as the second leading cause of injury-related death behind automobile accidents (42,433 deaths) (IOM 2004). In order to ensure that all potentially hazardous substances have proper warning labels, regulatory agencies require determination of acute toxicity hazard potential of substances and products. This determination for oral acute toxicity hazard is currently made using a test that requires laboratory rats. Historically, lethality estimated by the LD_{50} (i.e., the dose of a test substance that produces death in 50% of the animals tested) has been a primary toxicological endpoint in acute toxicity tests.

The conventional LD_{50} acute oral toxicity *in vivo* test method has been modified in various ways to reduce and refine[1] animal use in toxicity testing (OECD 2001a, c, d, e; EPA 2002a). Most recently, the LD_{50} was replaced, for hazard classification testing purposes, with the UDP, based on an Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) technical evaluation and formal ICCVAM recommendations (ICCVAM 2000, 2001c). This method now reduces animal use by over 70% compared to the previous method.

In 1999, at the request of the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances, ICCVAM reviewed the validation status of *in vitro* methods for estimating acute oral toxicity. This request was based on studies published in recent years that showed a correlation between *in vitro* and *in vivo* acute toxicity. *In vitro* cytotoxicity methods have been evaluated as another means to reduce and refine the use of animals and these methods may be helpful in predicting *in vivo* acute toxicity. Since moving the starting dose closer to the LD_{50} reduces the number of animals necessary for the acute oral systemic toxicity test, the use of *in vitro* cytotoxicity assays to predict a starting dose close to the LD_{50} may reduce animal use.

In October of 2000, the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity sponsored by the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS) and the EPA was convened in Arlington, VA. The Organizing Committee invited 33 expert scientists from academia, industry, and government agencies to participate in the Workshop. Invited scientific experts and ICCVAM agency scientists were assigned to one of four Breakout Groups and prepared recommendations on the following:

- *In Vitro* Screening Methods for Assessing Acute Toxicity
- *In Vitro* Methods for Toxicokinetic Determinations
- *In Vitro* Methods for Predicting Organ Specific Toxicity
- Chemical Data Sets for Validation of *In Vitro* Acute Toxicity Test Methods

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

Workshop participants concluded that none of the proposed *in vitro* methods had been formally evaluated for reliability and relevance, and that their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing had not been adequately assessed. However, an *in vitro* approach proposed by the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) was recommended for rapid adoption so that data could be generated to establish its usefulness with a large number of chemicals (ICCVAM 2001a). In addition, a separate *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was prepared to provide sample cytotoxicity protocols and instructions for using *in vitro* data to predict starting doses for acute *in vivo* systemic toxicity tests.

ICCVAM, which is charged with coordinating the technical evaluations of new, revised, and alternative test methods with regulatory applicability (ICCVAM Authorization Act of 2000, Public Law 106-545; available: http://iccvam.niehs.nih.gov/about/PL106545.pdf), agreed that *in vitro* basal cytotoxicity test methods should have a high priority for evaluation. The NTP Center for the Evaluation of Alternative Toxicological Methods (NICEATM) collaborated with the European Centre for the Validation of Alternative Methods (ECVAM), a component of the European Commission's Joint Research Centre, to further characterize the usefulness of *in vitro* cytotoxicity assays as predictors of starting doses for acute oral lethality assays. NICEATM and ECVAM designed a multi-laboratory validation study to evaluate the performance of two standardized *in vitro* basal cytotoxicity test methods using 72 reference substances with the ZEBET approach of using the Registry of Cytotoxicity (RC) regression model. Based on the procedures described in the *Guidance Document* (ICCVAM 2001b), the validation study used two mammalian cell types (i.e., BALB/c 3T3 mouse fibroblasts [3T3] and primary normal human epidermal keratinocytes [NHK]) for *in vitro* basal cytotoxicity test methods with a neutral red uptake (NRU) cell viability endpoint to predict starting doses for acute oral systemic toxicity test methods. The inclusion of human cells in the validation study also implements another workshop recommendation, that of evaluating whether cytotoxicity in human or rodent cells can be used to predict human acute toxicity.

The objectives identified for the validation study were to:

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

- Develop high quality *in vivo* acute oral lethality and *in vitro* NRU cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of *in vivo* acute oral lethality

Scientists assembled for the ICCVAM-sponsored scientific peer review panel meeting ("Panel") on May 23, 2006 independently assessed the usefulness and limitations of the *in vitro* basal cytotoxicity test methods to predict starting doses for acute oral systemic toxicity test methods. The Background Review Document (BRD) on the two *in vitro* NRU test methods prepared by NICEATM and provided to the peer review panel and the public contains:

1. Comprehensive summaries of the data generated in the validation study
2. An analysis of the accuracy and reliability of the test method protocols
3. Related information characterizing the potential animal savings produced by using the *in vitro* basal cytotoxicity test methods as adjuncts to specific acute systemic toxicity test methods

The Panel also evaluated draft test method performance standards, protocols, and draft ICCVAM recommendations for test method uses and future studies. The public was invited to provide comments on the BRD and other documents and to attend the Panel meeting. Prior to the Panel meeting, public comments provided about the documents were provided to the Panel for their consideration. The BRD can be obtained from the ICCVAM/NICEATM Web site (http://iccvam.niehs.nih.gov) or by contacting NICEATM.

Following the conclusion of the Panel meeting, the ICCVAM and its Acute Toxicity Working Group (ATWG) considered the Panel report, the performance standards for the use of *in vitro* basal cytotoxicity test methods to predict starting doses for acute systemic toxicity test methods, and any public comments in preparation of its final test method recommendations for these *in vitro* basal cytotoxicity test methods. These recommendations will be made available to the public and provided to the U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (Public Law 106-545).

On behalf of the ICCVAM, we gratefully acknowledge the many contributions of all who participated in the *in vitro* cytotoxicity validation study and those who assisted in the preparation of the documents evaluated at the peer review meeting. We extend a special thanks to the participating laboratory Study Directors and scientists who worked diligently to provided critical data and information. We also thank the ECVAM scientists who participated in the management of the validation study and who provided valuable information, comments, and opinions throughout the study. The efforts of the ATWG members were instrumental in assuring a complete and informative BRD. The efforts of the NICEATM staff in coordinating the validation study, providing timely distribution of

information, and preparing the various documents are acknowledged and appreciated. We especially acknowledge Dr. Judy Strickland and Mr. Michael Paris for their coordination of the validation study and preparation of the BRD and other documents.

William S. Stokes, D.V.M. D.A.C.L.A.M.
RADM, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Leonard Schechtman, Ph.D.
Deputy Director for Washington Operations
National Center for Toxicological Research
U.S. Food and Drug Administration
Chairman, ICCVAM

EXECUTIVE SUMMARY

This Background Review Document (BRD) reports the results of a validation study, organized and managed by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM), conducted to characterize two *in vitro* basal cytotoxicity tests for determining starting doses for rodent acute oral toxicity assays. In conducting this validation study, the protocols for two *in vitro* neutral red uptake (NRU) assays using BALB/c mouse fibroblast 3T3 cells (3T3) and normal human epidermal keratinocytes (NHK) were standardized and optimized, and the LD_{50} values for the reference substances were refined. The accuracy and reliability of the two *in vitro* NRU test methods were determined using 72 reference substances of various toxicities. Computer simulations were used to estimate the potential reduction in animal usage that could be accomplished by the use of either of these *in vitro* test systems. One outcome of this effort has been the generation of high quality *in vivo* lethality and *in vitro* cytotoxicity reference databases that will be useful in the development of other *in vitro* toxicity tests.

The validation study showed that the 3T3 and NHK NRU test methods are not sufficiently accurate as stand-alone methods to correctly predict rodent acute oral toxicity. However, based on computer simulations for the reference substances tested in this study, the use of either of these two *in vitro* basal cytotoxicity test methods for the selection of starting doses for rodent acute oral toxicity testing has the potential to reduce the number of animals used per test and, in some cases, the number of substance-induced animal deaths.

Introduction and Rationale

Although *in vitro* basal cytotoxicity test methods are not currently regarded as suitable replacements for rodent acute oral toxicity tests (Spielmann et al. 1999; ICCVAM 2001a), such methods have been examined as a possible approach to reduce and refine[2] the use of animals for such testing. An international Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) was initiated in 1983 to evaluate the relationship between *in vitro* cytotoxicity and acute human toxicity. Tests of 50 substances in 61 *in vitro* assays by multiple laboratories led to the identification of a battery of three human cell line assays whose cytotoxicity responses were highly correlated to human lethal blood concentrations (Bondesson et al. 1989; Clemedson et al 1996, 1996a; Ekwall et al. 1998a, 1998b, 2000). The Registry of Cytotoxicity (RC), initially published in 1998, is a database of 347 substances that currently consists of acute oral toxicity data from rats and mice and *in vitro* cytotoxicity data from studies using various mammalian cell types with a number of different toxic endpoints (Halle 1998, 2003). A regression formula, the RC millimole regression, constructed from these data was proposed by ZEBET, the German National Centre for the Documentation and Evaluation of Alternative Methods to Animal Experiments, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral toxicity tests (Halle 1998, 2003; Spielmann et al. 1999).

---

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

These, and other, initiatives to use *in vitro* cytotoxicity test methods to reduce animal use in acute toxicity testing were evaluated at the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, in October 2000 ("Workshop 2000"; ICCVAM 2001a). This workshop was organized by the U.S. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and NICEATM. Pursuant to this workshop, ICCVAM recommended (ICCVAM 2001a) further evaluation of the use of *in vitro* cytotoxicity data as one of the approaches that could be used to estimate the starting doses for rodent acute oral toxicity studies. The recommendations are based on preliminary information suggesting that this approach could reduce the number of animals used in such studies (i.e., reduction), minimize the number of animals that receive lethal doses (i.e., refinement), and avoid underestimating hazard. To assist in the adoption and implementation of the ZEBET approach, the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b) was prepared by ICCVAM with the assistance of the workshop participants.

In its recommendations for further evaluations, ICCVAM concurred with the Workshop 2000 recommendation that near-term validation studies should focus on two standard basal cytotoxicity assays: one using a human cell NHK system and one using a rodent cell (3T3) system. Historical data for *in vitro* cytotoxicity testing using mouse 3T3 cells are available (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996), as are historical data for *in vitro* basal cytotoxicity testing using NHK cells (e.g., Gettings et al. 1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994).

NICEATM, in partnership ECVAM, designed an international, multi-laboratory validation study to evaluate the reduction or refinement in animal use that could result from using cytotoxicity data from the 3T3 and NHK NRU test methods to estimate starting doses for two rodent acute oral toxicity test methods, the Up-and-Down Procedure (UDP; OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001d). The NRU protocols, as presented in the *Guidance Document,* were the initial basis of the NICEATM/ECVAM validation study protocols. These protocols were originally derived from the BALB/c 3T3 Cytotoxicity Test, INVITTOX Protocol No. 46 (available at the FRAME-sponsored INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]), the 3T3 cell studies by Borenfreund and Puerner (1984, 1985) and the rat epidermal keratinocyte study of Heimann and Rice (1983). A detailed description of the 3T3 and NHK NRU test method protocols used in the NICEATM/ECVAM validation study is provided in **Section 2**.

Protocol Components

Many protocol components used in the validation study are similar for the 3T3 and NHK cells. The following procedures are common to both cell types:

- Testing was performed in four phases (Ia, Ib, II, and III)
- Preparation of reference substances and positive control
- Cell culture environment conditions
- Determination of test substance solubility
- Configuration of 96-well plates for testing samples
- 48-hour exposure to test substance
- Range finder and definitive testing

- Microscopic evaluation of cell cultures for toxicity
- Measurement of NRU
- Data analysis

The main differences in the test methods for the two cell types are:

- The conditions of propagation of the cells in culture
- The cell growth medium components
- The volumes of reference substance added to the 96-well plate

Three laboratories participated in testing the 72 reference substances in both cell types:

- ECBC: The U.S. Army Edgewood Chemical Biological Center (Edgewood, MD)
- FAL: Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory (Nottingham, UK)
- IIVS: The Institute for *In Vitro* Sciences (Gaithersburg, MD)

BioReliance Corporation (Rockville, MD) procured and distributed the coded reference substances and performed solubility tests prior to their distribution to the testing laboratories, but did not perform any of the *in vitro* tests.

Validation Reference Substances

The 72 reference substances were selected to represent: (1) the complete range of *in vivo* acute oral toxicity (encompassing all five GHS acute oral toxicity categories as well as lower toxicities [GHS; UN 2005]); (2) the types of substances regulated by various regulatory authorities; and (3) substances with human toxicity data and/or human exposure potential. To ensure that the complete range of toxicity was covered, 12 substances were selected for each of the five acute oral toxicity categories, with an additional 12 substances with lower toxicities (i.e., LD_{50} >5000 mg/kg). A discussion of the characteristics and sources of the reference substances can be found in **Section 3**. The selected reference substances had the following characteristics:

- 58 (81%) of the 72 substances were also included in the RC, and 38% (22/58) of these were outliers with respect to the RC millimole regression.
- 27 (35%) of the substances were pharmaceuticals, 17 (22%) were pesticides, 8 (10%) were solvents, and 5 (6%) were food additives. The remaining substances were used for a variety of manufacturing and consumer products. The number of assigned uses (77) is greater than the number of selected substances because some of the substances have more than one use.
- 57 (79%) were organic compounds and 15 (21%) were inorganic; well-represented classes of organic compounds included heterocyclics, carboxylic acids, and alcohols.
- 22 (31%) substances were known, or expected to have, toxicologically active metabolites.
- Many of the selected substances had multiple target organs/effects; including neurological, liver, kidney, and cardiovascular effects.

Table ES-1 reports the number of substances that were tested and the number of substances used for the various analyses performed.

Table ES-1 Datasets Used for Validation Study Analyses[1]

Use	3T3 NRU[1]	NHK NRU[1]	Characteristics of Dataset
Testing	72	72	Substances tested
Comparison of laboratory IC_{50}-LD_{50} regressions to one another	47	51	RC substances with IC_{50} values from all laboratories and reference rat oral LD_{50} values
Comparison of combined-laboratory IC_{50}-LD_{50} regressions to a regression calculated with RC data	47	47	RC substances with IC_{50} values for both test methods from all laboratories and rat oral reference LD_{50} values
Prediction of GHS accuracy using IC_{50} values in IC_{50}-LD_{50} regressions; prediction of starting doses for acute oral toxicity test (UDP and ATC) simulations	67	68	Substances with IC_{50} values from at least one laboratory
Reproducibility of acceptable rat oral LD_{50} values	NA	NA	62 substances with more than one acceptable rat oral LD_{50} value
Reproducibility of IC_{50} values	64	68	Substances with IC_{50} values from all laboratories

Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake; NA=Not applicable.
[1]Number of substances.

Rodent Acute Oral Toxicity Reference Data

Because the 3T3 and NHK NRU test methods are intended to be used as adjuncts to rodent acute oral toxicity test methods, the LD_{50} values from rodent acute oral toxicity tests are the most appropriate reference data for evaluating the *in vitro* IC_{50} values (i.e., the test chemical concentration that reduces cell viability by 50%). Rodent acute oral LD_{50} reference data for the 72 reference substances were obtained from the literature. It was not possible to limit the data to studies conducted under Good Laboratory Practice (GLP) guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003) because only 2% of the published data retrieved were from such studies. Although mouse toxicity data were initially considered for inclusion in the database, the accuracy analyses were restricted to rat data. A total of 459 acute rodent oral LD_{50} values were identified for the reference substances. Reference LD_{50} values for each substance were identified by excluding studies with the following characteristics:

- Feral rats
- Rats <4 weeks of age
- Anesthetized rats
- Test substance administered in food or capsule
- LD_{50} reported as a range or an inequality

For substances with multiple LD_{50} values (i.e., from different sources), the rodent reference LD_{50} values for use in the validation study were determined by calculating a geometric mean of the available values for each reference substance. The reference LD_{50} values for 19 (26%) of the 72 substances varied sufficiently from the initial LD_{50} values that came from the RC

database and other summary sources, that the substances were reclassified into different GHS categories.

The reliability of the calculated rat acute oral LD_{50} reference values was assessed by comparison to other evaluations of the performance of rodent acute oral toxicity tests. For the 62 reference substances that had more than one LD_{50} value, the maximum:minimum ratios ranged from 1.1 to 25.9, with most below an order of magnitude.

Test Method Accuracy

Although the 3T3 and NHK NRU test methods are not intended to be used as replacements for rodent acute oral toxicity tests, they were evaluated for their ability to correctly predict the reference LD_{50} values (i.e., accuracy[3]). The rationale for evaluating the accuracy of LD_{50} predictions is that the current acute oral toxicity test methods (i.e., UDP, ATC, and Fixed Dose Procedure [FDP; OECD 2001c]) call for starting doses to be placed as close as possible and just below the true LD_{50}. When the starting dose is close to the true LD_{50} for a test substance, fewer animals are needed. When the starting dose is below the true LD_{50}, there is reduced pain and suffering because doses tend to be lower, and the test outcome bias is more conservative (i.e., higher toxicity). Regression models developed using IC_{50} and LD_{50} values were used to derive estimated LD_{50} values from 3T3 or NHK NRU IC_{50} values.

A number of different analyses were performed in an attempt to improve the estimation of the rat acute oral LD_{50}. IC_{50}-LD_{50} regressions (in millimole units) were calculated for each *in vitro* cytotoxicity test method and participating laboratory using the 3T3 and NHK IC_{50} values. Because the regressions for each NRU test method among laboratories were not significantly different from one another (for each NRU test method, $p > 0.5$), the regression for each NRU test method was based on data pooled across the laboratories. This combined-laboratory regression was then compared to the RC data using a regression based on RC IC_{50} and LD_{50} data for the 47 substances common to the validation study and the RC, with rat acute oral LD_{50} reference values, and with both 3T3 and NHK IC_{50} values produced by all three participating laboratories. The statistical comparison of slope and intercept (simultaneously) using an F test showed that neither the 3T3 regression nor the NHK regression was significantly different from the RC regression for the 47 substances ($p = 0.61$ and 0.76 respectively). These outcomes support use of the RC millimole regression.

Reference substances that fit the RC millimole regression poorly (i.e., outliers) were evaluated to determine whether there were relationships between their outlier status and their physical or chemical characteristics. Because the IC_{50}-LD_{50} regressions for the 3T3 and NHK NRU test methods yielded results that were not different from the RC regression for 47 substances, the RC millimole regression was preferred for analysis of outliers because it was based on a much larger data set and because it had established acceptance limits (Halle 1998, 2003). Certain chemical structural classes, boiling points, molecular weights, and log K_{OW} values were related with outliers, but solubility in the 3T3 or NHK medium and the cells' lack of xenobiotic metabolic capability did not correlate with outlier status. Because these *in*

[3] Accuracy is the agreement between a test method result and an accepted reference value (ICCVAM 2003).

vitro NRU test methods are based upon basal cytotoxicity, the mechanism of toxicity was also considered as a characteristic to explain the presence of outliers. Twenty-two reference substances were neurotoxic or cardiotoxic and were not expected to be active in the 3T3 and NHK cell cultures. Of these 22 substances, 13 (59%) were outliers (i.e., they fit the RC millimole regression poorly) using the 3T3 NRU and 12 (55%) were outliers using the NHK NRU. These substances represented 13/28 (46%) and 12/31 (39%) of the outliers for the 3T3 and the NHK NRU test methods, respectively. More information on the outlier analysis is presented in **Section 6.2**.

The potential variation produced by combining the LD_{50} values of two rodent species in the RC millimole regression was eliminated by developing a regression based solely on RC substances with rat LD_{50} data (i.e., the RC rat-only millimole regression). The RC rat-only data were also converted to a weight basis for an additional regression, the RC rat-only weight regression, for applicability to mixtures or to substances for which molecular weight is unknown.

The accuracy of the *in vitro* NRU test methods when used with each of the IC_{50}-LD_{50} regressions was characterized by determining the proportion of reference substances for which their GHS categories (based on rat acute oral LD_{50} data) were correctly predicted. The accuracy of the RC rat-only millimole regression was 31% (21/67 reference substances) and 29% (20/68 reference substances) with the 3T3 and the NHK NRU test methods, respectively. The accuracy of the RC rat-only weight regression was similar, 31% with the 3T3 NRU test method (21/67 reference substances) and 31% with the NHK NRU test method (21/68 reference substances). The poor accuracy is due, in part, to the skewness of the reference substance set with respect to the fit of the reference substances to the regressions and to the differences between cell cultures and whole animal exposures. Each regression showed a general trend to underpredict the toxicity of the most toxic chemicals, and to overpredict the toxicity of the least toxic chemicals. A detailed discussion of the accuracy analyses is presented in **Section 6.4.**

Test Method Reliability

Reproducibility is the consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or among different laboratories (interlaboratory reproducibility) using the same protocol and test samples. Reproducibility was evaluated using results from the 64 reference substances tested in 3T3 cells and the 68 substances tested in NHK cells that yielded replicate IC_{50} values in all three laboratories. Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC_{50} data was assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC_{50}-LD_{50} regressions, and comparison of maximum:minimum mean laboratory IC_{50} values. Reproducibility was generally better with the NHK NRU test method.

Although ANOVA results for the positive control (sodium lauryl sulfate [SLS]) IC_{50} values from the 3T3 NRU test method indicated that there were significant differences among laboratories ($p = 0.006$) but not between study phases within laboratories ($p > 0.01$), the data show (see **Figure 7-5**) that laboratory means and standard deviations from each testing phase overlap , and that the IC_{50} was stable between testing phases. The interlaboratory CV values

for the various study phases ranged from 2 to 16%. ANOVA results for the SLS IC_{50} from the NHK NRU test method showed significant differences among laboratories ($p < 0.001$) and among study phases within laboratories ($p \leq 0.001$). The use of a different cell culture method at FAL was responsible for SLS IC_{50} differences among the laboratories in Phases Ia and Ib. After harmonization of culture methods across laboratories, the laboratory means and standard deviations were similar for Phases II and III (see **Figure 7-5**). Interlaboratory CV values for the NHK NRU for Phases Ia and Ib, were 39% and 21%, respectively. Interlaboratory CV values for Phases II and III were 31% and 8%, respectively. The linear regression analyses of the SLS IC_{50} over time (within each laboratory) for both NRU test methods indicated that IC_{50} values generated over the 2.5-year duration of the study were stable.

For the reference substances, the similarity among the laboratories' LD_{50} predictions (via regression) from IC_{50} values (see **Figure 7-1**) was considered significant with respect to the reproducibility analyses because these *in vitro* NRU test methods are proposed for use in determining starting doses for acute oral toxicity tests using the predicted LD_{50}. ANOVA showed significant laboratory differences for 23 substances with the 3T3 NRU test method (see **Table 7-4**) and six substances with the NHK NRU test method (see **Table 7-6**). Mean intralaboratory CV values were 26% for both NRU test methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs. 47%). An analysis to determine the relationship, if any, between reference substance attributes and interlaboratory CV indicated that chemical class, physical form, solubility, and volatility had little effect. The CV seemed to be related instead to the GHS hazard category, the IC_{50}, and boiling point (see **Section 7.2.3**). However, the usefulness of these relationships is not known. Mean interlaboratory CV values were larger for substances in the most toxic GHS hazard categories than for substances in the other toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for substances in the $LD_{50} \leq 5$ mg/kg (72%) and $5 < LD_{50} \leq 50$ mg/kg (78%) categories were larger than the mean overall interlaboratory CV (47%) with the 3T3 NRU test method. When the NHK NRU test method was used, the mean interlaboratory CV was 37% for substances with $LD_{50} \leq 5$ mg/kg, and 41% for substances with $5 < LD_{50} \leq 50$ mg/kg, and the mean overall interlaboratory CV was 28%. A Spearman correlation analysis indicated that the IC_{50} was inversely correlated to interlaboratory CV for both the 3T3 ($p = 0.015$) and NHK ($p = 0.014$) NRU test methods, and that boiling point was positively correlated to interlaboratory CV ($p = 0.007$) for the 3T3 but not the NHK ($p = 0.809$) NRU test method.

The maximum:minimum mean laboratory IC_{50} ratios for the 3T3 NRU test method ranged from 1.1 to 21.6, with 37 of 64 (58%) reference substances having ratios less than 2.5. The maximum:minimum mean laboratory IC_{50} ratios for the NHK NRU test method ranged from 1.0 to 107.6, with 58 of 68 (85%) reference substances having ratios less than 2.5.

Data Quality

The laboratories reported no significant deviations from the protocols, and deviations that did occur were acknowledged and addressed by the Study Directors. Tests that had deviations affecting the data were rejected by the Study Directors and repeated. The computation of test method and data collection errors showed that the non-GLP laboratory consistently had the

highest error rate and the lowest intralaboratory reproducibility for IC_{50} results; however, the laboratory's GHS acute oral toxicity category predictions were comparable to that for the other laboratories.

An electronic copy of all data for the validation study can be obtained from NICEATM upon request by mail, fax, or e-mail to Dr. William S. Stokes, NICEATM, NIEHS, P. O. Box 12233, MD EC-17, Research Triangle Park, NC, 27709, (phone) 919-541-2384, (fax) 919-541-0947, (e-mail) niceatm@niehs.nih.gov.

Other Scientific Reports and Reviews

3T3 and NHK NRU methods have been evaluated for purposes other than the prediction of starting doses for acute toxicity studies (e.g., ocular irritancy; human lethal blood concentrations, *in vivo* phototoxicity). *In vitro* NRU cytotoxicity test methods using various cell types have been evaluated for their correlation with rodent lethality endpoints (e.g., rat/mouse intravenous[i.v.], intraperitoneal [i.p.], and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993) showed good correlations (r =0.88) of *in vitro* cytotoxicity with rodent i.p./i.v. and i.v. toxicity data, respectively. A 3T3 NRU test method has been validated by ECVAM for the identification of *in vivo* phototoxic potential.

No *in vitro* test methods have been validated for the prediction of acute oral toxicity. Estimations of animal savings using *in vitro* cytotoxicity data to estimate starting doses for the UDP did not use actual *in vitro* cytotoxicity data. Instead, animal savings were estimated by assuming that the *in vivo* starting dose equals the true LD_{50}, which is an approach that assumes that cytotoxicity data can perfectly predict *in vivo* lethality. These theoretical predictions of animal savings in the UDP ranged from 25-40% (ICCVAM 2001a), as compared with the average animal savings of 5.3-7.8% predicted using computer simulation modeling of the UDP for the reference substances tested in the NICEATM/ECVAM study. Halle et al. (1997) used the *in vitro* cytotoxicity data in the RC to determine that an animal savings of 32% can be attained for the ATC method by using the LD_{50} predicted by the RC millimole regression as the starting dose. For the reference substances tested in the NICEATM/ECVAM validation study, most of which were a poor fit to the RC millimole regression, the average animal savings for the ATC, as determined by computer simulation modeling, was 4.8-10.2%.

Animal Welfare Considerations: Reduction, Refinement, and Replacement

Computer models were used to simulate testing of the reference substances using to the UDP and ATC test methods. In principle, animal savings with the FDP could be estimated even though death is not the primary endpoint, but the validation study did not include this analysis. The number of animals that would be used, and the number of animals that would survive or die during the UDP or ATC procedure, were determined for the default starting doses and compared with those when starting dose was based on LD_{50} values determined from IC_{50} values for each reference substance using the RC rat-only regressions.

Computer simulation of UDP testing showed that, for the reference substances used in this validation study, using the 3T3 or NHK NRU test methods and the RC rat-only millimole regression to identify the starting dose resulted in the use of fewer animals per test by an

average of 5.3% (0.50 animals) to 6.6% (0.53 animals), depending upon the assumed mortality-response slope and *in vitro* NRU test method used. The RC rat-only weight regression predicted mean animal savings of 6.0% (0.56 animals) to 7.8% (0.62 animals). When substances were grouped by GHS acute oral toxicity category, there were no animal savings for substances in the 50 $<LD_{50} \leq$300 mg/kg category because the default starting dose is in this range. The greatest animal savings were observed for substances with 2000 $< LD_{50} \leq$5000 mg/kg and LD_{50} >5000 mg/kg because the limit test, which would be used for such substances, uses fewer animals that the main test. Animal savings for these toxicity categories using the RC rat-only millimole regression ranged from 11.3% (1.21 animals) to 20.3% (1.58 animals) per test. Use of the RC rat-only weight regression produced animal savings of 12.8% (1.38 animals) to 21.0% (1.63 animals) per test. Although the use of the 3T3 and NHK NRU test methods to estimate starting doses for the simulated UDP decreased the numbers of animals used per test, it did not change the numbers of animals that died.

Computer simulation of ATC testing showed that, for the reference substances used in this validation study, using the 3T3 or NHK NRU test methods and the RC rat-only millimole regression to identify the starting dose resulted in a savings of 4.8% (0.51 animals) to 7.3% (0.80 animals) per test, depending upon the assumed mortality-response slope and the *in vitro* NRU test method used. The use of the RC rat-only weight regression produced animal savings of 8.6% (0.91 animals) to 10.2% (1.09 animals). When substances were grouped by GHS acute oral toxicity category, there were no animal savings for substances in the 300 $< LD_{50} \leq$2000 mg/kg category because this category contains the default starting dose for the ATC method. Animal savings were highest for substances with 5 $< LD_{50} \leq$50 mg/kg and LD_{50} >5000 mg/kg. The mean animal savings for both *in vitro* NRU test methods for substances with 5 $< LD_{50} \leq$50 mg/kg ranged from 9.8% (1.15 animals) to 11.4% (1.33 animals) per test for the RC rat-only millimole regression. The greatest reduction in animal use would be for substances with LD_{50} >5000 mg/kg because the limit test used fewer animals than the main test. Animal savings for these substances ranged from 17.1%, (2.03 animals) to 22.2% (2.66 animals) per test for the RC rat-only millimole regression. When the RC rat-only weight regression was used, the mean animal savings with both *in vitro* NRU test methods for substances with 5 $< LD_{50} \leq$50 mg/kg ranged from 10.8% (1.25 animals) to 13.0% (1.51 animals) per test. Mean animal savings for substances with LD_{50} >5000 mg/kg ranged from 24.8% (2.94 animals) to 27.7% (3.33 animals) per test. The use of IC_{50} values to estimate starting doses for the ATC tests refined animal use by producing fewer animal deaths by approximately 0.5 to 0.6 animals per test.

Simulations for the UDP and ATC method showed that the use of cytotoxicity results to estimate starting doses did not significantly alter the GHS categorizations compared with the categories determined using default starting doses. This concordance was 97 to 99% for the 3T3 and NHK NRU test methods.

Practical Considerations

Practical issues with respect to the implementation of these *in vitro* NRU test methods include the need for, and availability of, appropriate cell culture equipment, training and expertise, cost, and time expenditure. The ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002) encourages the establishment of laboratory practices and

principles that will reduce uncertainty in the development and application of *in vitro* test methods.

All equipment and supplies are readily available, and the *in vitro* NRU test methods are easily transferable to laboratories experienced with mammalian cell culture techniques. Much of the training and expertise needed to perform the 3T3 and NHK NRU test methods are common to people with mammalian cell culture experience. Additional technical training would not be intensive because these methods are similar in general performance to other *in vitro* mammalian cell culture assays. GLP training should be provided to laboratory personnel (including study directors and principal investigators) to ensure proper adherence to test protocols and data documentation and verification procedures.

Prices for commercial *in vitro* NRU cytotoxicity testing to determine the IC_{50} for one substance ranged from $1120 to $1850. It is not clear if the price of an *in vivo* test would be reduced if it were preceded by an *in vitro* cytotoxicity test to set the starting dose. Thus, use of these test methods may not reduce the overall cost of rodent acute oral toxicity testing and may increase the cost, but their use has the potential to reduce the number of animals and the time needed for a study. The greatest savings in time and animals will occur if the IC_{50} data determine that the rodent acute oral toxicity limit test should be performed, rather than the main test. Based on the cost and technical procedures associated with cell culture maintenance, the 3T3 NRU test method is less expensive and less complicated to conduct than the NHK NRU test method.

[This Page Intentionally Left Blank]

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

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

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

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

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

This section provides:

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A partial least squares (PLS) analysis indicated that the IC_{50} data generated from as many as 61 test methods predicted the three sets of LC data well (R^2=0.77, 0.76, and 0.83, Q^2=0.74, 0.72, and 0.81, respectively, where R^2 is the determination coefficient and Q^2 is the predicted variance according to cross-validation in the PLS model used). A two component PLS model using rat and mouse oral LD_{50} values less accurately predicted human LC values (R^2=0.65, Q^2=0.64). These results suggested that *in vitro* basal cytotoxicity assays might be more effective in estimating human acute oral lethality than rodent acute oral toxicity test methods.

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

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

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

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

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

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

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

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

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

---

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

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

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

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

1.1.3 The Registry of Cytotoxicity (RC)

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

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

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

The following cytotoxicity endpoints were accepted:

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

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

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

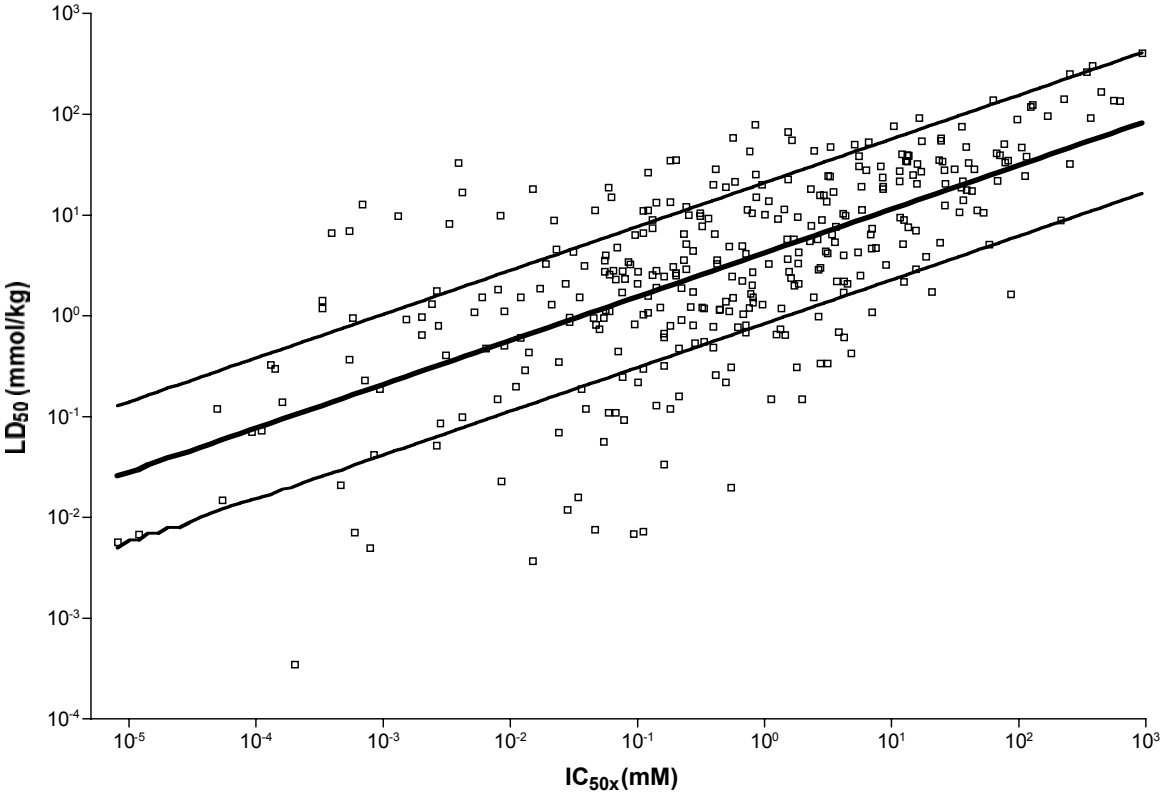
To develop a model for the prediction of acute oral LD_{50} values from IC_{50x} values, Halle (1998, 2003) calculated a linear regression from pairs of the log-transformed IC_{50x} values (in mM) and log transformed rodent oral LD_{50} values (in mmol/kg) (see **Figure 1-1**). Molar concentrations were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

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

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

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

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



Abbreviations: RC=Registry of Cytotoxicity; IC_{50x}=Geometric mean (of multiple endpoints and cell types) test substance concentration that reduces cell viability by 50%; LD_{50}=Dose producing death in 50% of the animals tested.
The heavy line shows the fit of the data to a linear regression model, log (LD_{50}) = 0.435 x log (IC_{50x}) + 0.625; r=0.67. The thinner lines show the empirical prediction interval (± log 5, or ±0.699) that is based on the anticipated precision for the prediction of LD_{50} values from cytotoxicity data (Halle 1998, 2003).

1.1.4 The ZEBET Initiative to Reduce Animal Use

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The specific objectives for this validation study were to:

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

1.1.6.1 *Study Design*

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

Figure 1-2 NICEATM/ECVAM Validation Study Phases

Phase Ia: Laboratory Evaluation

Development of a positive control database for each laboratory

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

⇓

Phase Ib: Laboratory Evaluation

Limited substance testing to demonstrate the reliability of the protocol

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

⇓

Phase II: Laboratory Qualification

Evaluation of protocol refinements

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

⇓

Phase III: Laboratory Testing Phase

Test of optimized protocols

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

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

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

1.2.1 Current Regulatory Testing Requirements for Acute Oral Toxicity

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

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

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

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

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

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

Table 1-2 Regulatory Classification Systems for Acute Oral Toxicity

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

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

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

1.2.1.1 *Test Methods for Assessing Acute Oral Toxicity*

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

Ekwall (1983) and others (e.g., Grisham and Smith 1984) concluded that, because the actions of substances that produce injury and death are ultimately exerted at the cellular level, *in vitro* cytotoxicity assays might be useful for the prediction of acute lethality potency, as well. Considerable research has been undertaken to develop and evaluate *in vitro* tests for use as screens and as potential replacements for rodent LD_{50} tests, and numerous groups have reported good agreement between *in vitro* cytotoxicity and animal lethality (see reviews by Phillips et al. 1990; Garle et al. 1994; Guzzie 1994). However, none of the proposed *in vitro* models have been evaluated in any formal studies for reliability and relevance, and their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing data have not been assessed.

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

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

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

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

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

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

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

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

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

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

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

[This Page Intentionally Left Blank]

2.0 TEST METHOD PROTOCOL COMPONENTS OF THE 3T3 AND NHK *IN VITRO* NRU TEST METHODS

The *Guidance Document* (ICCVAM 2001b) recommended that the following be incorporated into any *in vitro* cytotoxicity protocol used to predict rodent acute oral lethality:

- A cell line (or primary cells) that divides rapidly (e.g., with a doubling time of <24 hours)
- An initial seeding density that allows for exponential cell growth throughout the exposure period
- An exposure period that spans at least one cell cycle
- Appropriate positive control (PC) and vehicle control (VC) substances for which toxicity and lack of toxicity, respectively, has been well characterized by the performing laboratory
- Solvents that are used only at concentrations that do not cause significant toxicity to the cell system over the entire period of the assay
- A well-established, quantifiable cytotoxicity endpoint that has good interlaboratory reproducibility
- Tests that are compatible with at least 96-well plates and equipment (e.g., spectrophotometric microplate reader) that allow a quick and precise measurement of the endpoint of interest
- Use of a progression factor in the concentration-response experiment that yields graded effects between 0% and 100% cytotoxicity

Section 2.1 provides the basis for the selection of the *in vitro* 3T3 and NHK NRU test methods. **Section 2.2** provides descriptions of the NRU protocols applicable to this validation study. **Section 2.3** provides details for performing the 3T3 and NHK NRU test methods and explains the rationales for the various test method components, and **Section 2.4** describes any 3T3 and NHK NRU test method proprietary aspects. **Section 2.5** discusses the basis for the replicate and repeat tests conducted during validation of these two test methods. **Section 2.6** details the modifications and revisions made during the first two phases of the validation study which contributed to the development of the final protocol used in Phase III. **Section 2.7** describes the differences between the protocols used in this study and the protocols outlined in the *Guidance Document*. **Sections 2.8**, **2.9**, and **2.10** provide details on the solubility protocol evaluated during the validation study and used to identify the appropriate solvent for dissolving the reference substances.

The 3T3 and NHK NRU test method protocols were provided to the three laboratories that participated in the validation study (see **Section 5.6.3** for additional laboratory information). These were:

- The U.S. Army Edgewood Chemical Biological Center (ECBC)
- The FRAME Alternatives Laboratory (FAL)
- The Institute for *In Vitro* Sciences (IIVS)

A fourth laboratory (BioReliance Corporation, Rockville, MD) was used to procure and distribute the coded reference substances, and to perform solubility tests on the validation study reference substances prior to their distribution to the participating laboratories.

2.1 Basis for Selection of the *In Vitro* NRU Cytotoxicity Test Method

As stated in **Section 1**, in agreement with the recommendations of the Workshop 2000 participants (ICCVAM 2001a), ICCVAM made the following recommendations and forwarded them to U.S. Federal agencies along with the Workshop 2000 Report (ICCVAM 2001a) and *Guidance Document* (ICCVAM 2001b).

> "ICCVAM concurs with the Workshop recommendation that near-term validation studies should focus on two standard cytotoxicity assays: one using a human cell system and one using a rodent cell system. Since the murine BALB/c 3T3 cytotoxicity assay has been evaluated for only a limited number of chemical classes, there is merit in determining its usefulness with a broader array of chemical classes. Cell lines established from the rat rather than the mouse might also be considered, as most acute oral toxicity testing is conducted in this species. Human cell lines should also be considered since one of the aims of toxicity testing is to make predictions of potential toxicity in humans. Future validation studies should therefore compare rodent and human *in vitro* data with one another, with rodent *in vivo* data, and with human *in vivo* data. Correlations between *in vitro* and *in vivo* data might help in selecting cytotoxicity assays for further evaluation". (ICCVAM 2001a)

Based on this recommendation and the *Guidance Document* recommendation, NICEATM and ECVAM selected the 3T3 and NHK NRU basal cytotoxicity test methods for validation.

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

The *Guidance Document* (ICCVAM 2001b) provided the basic approach for the use of *in vitro* NRU basal cytotoxicity test methods to predict starting doses for rodent acute oral toxicity assays using the RC millimole regression. The 3T3 and NHK NRU test method protocols used in the validation study were derived from those proposed in the *Guidance Document*.

2.1.2 *Guidance Document* Rationale for Selection of Cell Types

The Workshop 2000 participants (ICCVAM 2001a) concluded that there were no significant differences between the basal cytotoxicity results obtained using permanent mammalian cell lines, primary human cells (e.g., NHK cells), or the IC_{50x} approach of Halle and Spielmann (Halle 1998, 2003; Spielmann et al. 1999; Halle and Spielmann 1992). Further, the *Guidance Document* recommended that *in vitro* basal cytotoxicity test methods not use hepatocytes (or related metabolically competent cells) or other types of highly differentiated cells because they may not give the best prediction of acute lethality for the large variety of chemicals likely to be tested (ICCVAM 2001b). However, it was recognized that, ultimately, simple predictive systems (*in vitro* or *in silico*) would be needed for early identification of those substances likely to be metabolized to more toxic or less toxic species than the parent chemical as well as those that were likely to exhibit cell-specific toxicity (e.g., Fentem et al. 1993; Seibert et al. 1996; Curren et al. 1998; Ekwall et al. 1999).

Established rodent cell lines were recommended for validation because (ICCVAM 2001b):

- It was assumed that such cells would give the best prediction of rat and mouse acute oral lethality (i.e., like correlates with like).
- The use of a readily available, easy to culture, immortalized cell line for *in vitro* cytotoxicity testing would accelerate the development of a database that can be used to analyze the usefulness of this approach.

Human cells also offer potential advantages. As determined in the MEIC project, the *in vitro* test methods with the best predictivity for peak acute LC values in humans generally used human cell lines (Ekwall et al. 1998b). Thus, a long-term advantage of using human cells is that *in vitro* human cell cytotoxicity data can be added to human toxicity databases to facilitate the development of test methods that may better predict acute oral human lethality.

3T3, an immortalized mouse fibroblast cell line, and NHK, primary human cells, were selected as representative rodent and human cells, respectively, for the NICEATM/ECVAM validation study. Historical data for the 3T3 NRU test were available from a variety of studies, including controlled and blinded validation studies, indicating the reliability of this test method (Gettings et al. 1991, 1992, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996; Balls et al. 1995; Brantom et al. 1997). NHK cells have also been used in validation studies for basal cytotoxicity test methods with good results (Willshaw et al. 1994; Sina et al. 1995; Gettings et al. 1996; Harbell et al. 1997).

2.2 Overview of the 3T3 and NHK NRU Test Methods

The *Guidance Document* (ICCVAM 2001b) includes a proposed 3T3 NRU test method protocol based on the 3T3 Cytotoxicity Test (INVITTOX Protocol No. 46; available from the FRAME-sponsored INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]), which in turn was based on the Borenfreund and Puerner (1985) protocol, as elaborated on in Spielmann et al. (1991, 1996). This protocol was updated based on experience obtained during the validation of the 3T3 NRU Phototoxicity Test (INVITTOX Protocol No. 78; also available at the FRAME INVITTOX database). The RC millimole regression for prediction of acute oral rat and mouse toxicity (Halle 1998, 2003; Spielmann et al. 1999) was included as the prediction model (ICCVAM 2001b; see **Section 1.1.2**).

The NHK NRU protocol provided in the *Guidance Document* was based on the protocol used by IIVS, which was based on a NRU protocol of Borenfreund and Puerner (1984) and a rat epidermal keratinocytes protocol (Heimann and Rice 1983). Formulations for the media and solutions, and general NHK cell culture techniques, correspond to Clonetics® products from the CAMBREX Corporation.

The protocol components for the 3T3 and NHK NRU test methods used in this validation study are similar (see **Figure 2-1**). The nature of the NRU response is described in **Section 1.3.1**. **Figure 2-1** provides an overview to the major steps for performance of the *in vitro* NRU test methods. The following procedures are common to both cell types:

- Preparation of substances and the PC
- Cell culture environmental conditions
- Determination of test substance solubility
- 96-well plate configuration for testing samples

- Range finder and definitive tests
- Microscopic evaluation of cell cultures for toxicity based on morphological alterations
- Procedures for measurement of NRU
- Data analysis procedures

The main protocol differences between the two cell lines are:

- The conditions of propagation of the cells in culture (e.g., time needed to reach appropriate confluence)
- The growth media components
- The volumes of substances applied to the 96-well plates
- The number of cell divisions undergone by each cell line during exposure to a test substance

2.2.1 The 3T3 NRU Test Method

2.2.1.1 *Initiating and Subculturing 3T3 Cells*

Each laboratory initially prepared a large pool of 3T3 cells (described further in **Section 2.3.1.1**), cryogenically preserved multiple ampules of these cells in liquid nitrogen, and periodically removed an ampule when needed. Although the NRU protocols used for each study phase provided cell culture density guidelines for subculturing the cells, each laboratory refined the final seeding density to achieve optimal growth.

Cryopreserved 3T3 cells were thawed, resuspended in a culture medium containing Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with non-heat-inactivated 10% newborn calf serum (NCS), transferred into tissue culture flasks (25 or 75-80 cm^2), and incubated at 37 °C ±1 °C, 90% ±5% humidity, and 5.0% ±1% CO_2/air. When cells reached 50 to 80% confluence (as estimated from a visual inspection of cell density), they were removed from the flask by trypsinization. A single-cell suspension was added to new flasks for propagation and the cells were passaged/subcultured at least two times[1] before seeding into 96-well plates for testing. This study did not evaluate the potential effects that cell passage number may have on the performance of the 3T3 NRU test method.

[1] 3T3 cells were maintained in culture for approximately two months (approximately 18 passages) and used for the NRU test. The *Guidance Document* (ICCVAM 2001b) did not provide a rationale for using 18 passages as the limit, but it was probably recommended to maintain homogeneity of the 3T3 cell population (i.e., decrease the potential of the population to drift genetically). The more passages the cells undergo, the more likely their response to chemical stress may change.

Figure 2-1 Major Steps in the Performance of the NRU Test Methods

(1) Cells (3T3 or NHK) are seeded into 96-well plates to form a sub-confluent monolayer; plates are incubated at 37 ºC (24 hours for 3T3 cells; 48-72 hours for NHK cells)

⇓

(2) Culture medium is removed (3T3 cells only)

⇓

(3) Reference substances in the appropriate solvents are added to the cells; cells are exposed for 48 hours at 37 ºC over a range of eight (8) concentrations

⇓

(4) Cells are evaluated microscopically for toxicity based on morphological appearance

⇓

(5) Treatment medium is removed; cells are washed once with Dulbecco's Phosphate Buffered Saline (D-PBS); Neutral Red (NR) dye medium is added (3T3 cells: 25 μg/mL NR dye; NHK cells: 33 μg/mL NR dye); plates are incubated for 3 hours at 37 ºC

⇓

(6) NR medium is discarded; cells are washed once with D-PBS; NR desorbing fixative is added to the wells

⇓

(7) Plates are shaken for 20 minutes at room temperature

⇓

(8) NR absorption is measured at optical density (OD) 540 ±10 nm

⇓

(9) NRU is calculated as a percent of vehicle control values to define IC_{20}, IC_{50}, and IC_{80} concentrations (μg/mL)[2]

Abbreviations: NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; IC_{20}, IC_{50}, IC_{80}=Substance concentration that reduces cell viability by 20, 50, and 80%, respectively.

[2] IC_{50} values are used for estimating the LD_{50} value of a reference substance. The IC_{20} and IC_{80} values were determined for possible use in estimating human lethal concentrations in blood.

2.2.1.2 *Preparation of Cells for 96-well Plate Assays*

After subculturing the cells, 100 µL of the cell suspension ($2.0 – 3.0x10^3$ cells/well) were placed in the appropriate wells and 100 µL of cell-free culture medium were dispensed into the 36 peripheral wells (blanks). The peripheral wells were in rows 1 and 8 and columns 1 and 12 (See **Figure 1** in **Appendix B1** or **B2**). Peripheral wells were used only for blanks because they may be subjected to more evaporation than interior wells. The *Guidance Document* authors (and the SMT and Study Directors) concluded that such conditions would ultimately affect cell growth in these wells. One plate was prepared for each reference substance. The cells were incubated for 24 ±2 hours at 37 ºC and checked visually to be sure that approximately a 50% confluent monolayer was present at the time of substance application.

2.2.1.3 *Reference Substance Application*

After the appropriate incubation period to achieve a half-confluent monolayer, the medium was removed and 50 µL of culture medium with 10% NCS were added to each well. Then, 50 µL treatment medium containing the appropriate substance concentrations were added for a final concentration of 5% NCS. The cells were then incubated at 37 ºC for 48 ±0.5 hours. At the end of the incubation period, the cells were microscopically evaluated for changes in morphology and their appearance was documented (as per Visual Observation Codes in the protocol) prior to measurement of NRU.

2.2.2 The NHK NRU Test Method

2.2.2.1 *Initiating and Subculturing NHK Cells*

Cryopreserved NHK cells (ampules of cryopreserved cells were obtained from CAMBREX Corporation and stored in liquid nitrogen until needed) were thawed, resuspended in serum-free keratinocyte complete growth medium (see **Section 2.3.1.4** for components of the medium), transferred into tissue culture flasks (25 cm^2 without fibronectin-collagen coating), and incubated at 37 °C ±1 °C, 90% ±5% humidity, and 5.0% ±1% CO_2/air. When the cells reached 50 to 80% confluence (as estimated from a visual inspection of cell density), they were removed from the flask by trypsinization and prepared for subculturing into the 96-well plates. Care was taken to prevent the keratinocyte cultures from becoming 100% confluent as this may lead to cell differentiation, which would alter the intrinsic sensitivity of these cells to cytotoxic substances. To minimize potential sources of experimental variability, the laboratories used the same lot of Clonetics® cells throughout the validation study, the same brand of growth medium and supplements (and concentrations of supplements), and cells were not used beyond their second passage. The protocols for each study phase provided cell culture density guidelines, but each laboratory refined the final seeding densities to achieve appropriate growth.

2.2.2.2 *Preparation of Cells for 96-well Plate Assays*

After subculturing the cells, 125 µL of the cell suspension ($2.0 – 2.5x10^3$ cells/well) were placed in the appropriate wells and 125 µL of cell-free culture medium were dispensed into the peripheral wells (blanks). One plate per reference substance was prepared. The cells were incubated at 37 ºC for 48-72 hours and checked to be sure that cultures were at 20 to 50% confluence at the start of exposure to the reference substance.

2.2.2.3 *Reference Substance Application*

To add the reference substances, 125 µL of culture medium containing the appropriate reference substance concentrations were added to the existing 125 µL of culture medium in the test wells. The cells were then incubated at 37 ºC for 48 ±0.5 hours. At the end of the exposure period, the cells were microscopically evaluated for changes in morphology and their appearance was documented (as per Visual Observation Codes in the protocol [see **Appendices B1** and **B2**]) prior to measurement of their NRU.

2.2.3 Measurement of NRU in the 3T3 and NHK Test Methods

The treatment medium was removed from the 96-well plates, the cells were rinsed with phosphate buffered saline (PBS), and 250 µL NR dye medium was added to the wells (25 µg NR/mL for 3T3 cells; 33 µg NR/mL for NHK cells). The plates were then incubated (37 °C ±1 °C, 90% ±5% humidity, and 5.0% ±1% CO_2/air) for three hours. After incubation, the NR medium was removed, the cells were rinsed with PBS, and 100 µL of the desorb solution were applied. The plates were shaken on a microtiter plate shaker for 20 to 45 minutes to extract NR from the cells and to form a homogeneous solution. The optical density (OD) of the resulting colored solution was measured (within 60 minutes of adding the desorb solution) at 540 nm ±10 nm (OD_{540}) in a spectrophotometric microtiter plate reader, using the blank wells as reference. Data from the plate reader were transferred to a Microsoft® EXCEL® (Microsoft Corporation, Redmond, WA, USA) spreadsheet template (hereafter know as EXCEL® template) designed by the SMT and the testing laboratories for statistical analyses.

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

The protocols used in Phases I, II, and III of the validation study (**Appendices B** and **C**) are modifications of the protocols reported in the *Guidance Document* (ICCVAM 2001b). The participating laboratories provided comments and recommendations during the development of these protocols. The following information is specific to the protocols used in this validation study.

2.3.1 Materials, Equipment, and Supplies

2.3.1.1 *3T3 Cells*

The CCL-163, 3T3 BALB/c mouse fibroblast, cell line, clone 31 from the American Type Culture Collection (ATCC), Manassas, VA, USA, was used. The 3T3 cells, an immortalized mouse fibroblast cell line, were procured from the ATCC by IIVS at passage 64. IIVS cultured the cells to expand their number and cryogenically preserved them as a pool at passage number 69. ECBC and FAL received frozen ampules of cells at passage number 69 from IIVS, propagated the cells, and cryopreserved multiple ampules of cells at a slightly higher passage number to establish their working cell banks for use throughout the study. Each laboratory determined the doubling time for the 3T3 cell line prior to NRU testing in Phase Ia as required by the protocol in **Appendix C1**. The following doubling times were reported: 18.6 hours by ECBC; 17 hours by FAL; and 17 hours by IIVS. No other doubling time measurements were made. The extent of cell confluence was monitored during the study to identify when the cultures were in exponential growth.

2.3.1.2 *NHK Cells*

A single lot of pooled donor, primary neonatal foreskin keratinocyte (NHK) cells (Clonetics® # CC-2507; lot # 1F0490N) from CAMBREX Bio Science Walkersville, Inc., Walkersville, MD, USA, was used throughout the validation study. Keratinocytes from other sources would be acceptable if they meet the growth requirements identified in the protocols. Each laboratory determined the doubling time for the NHK cells prior to testing in Phase Ia (as required by the protocol in **Appendix C2**). The following doubling times were reported: 21 hours by ECBC; 10 hours by FAL; and 15.8 hours by IIVS. No other doubling time measurements were made. The extent of cell confluence was monitored during the study to identify when the cultures were in exponential growth.

2.3.1.3 *Tissue Culture Materials and Supplies*

The 3T3 and NHK NRU test methods require general tissue culture materials and supplies (see **Appendices B1** and **B2** [protocols] for formulations, and concentrations of solutions and media). Both test methods used the same materials for solubility testing (**Section 2.8.1**). Freshney (2000) provides information on all aspects of cell culture, including materials, supplies, and equipment needed. The following materials were needed for both test methods:

- Trypsin (0.05%)
- PBS
- Hanks' Balanced Salt Solution (HBSS) without Ca^{2+} and Mg^{2+}
- NR dye
- Glacial acetic acid
- Dimethyl sulfoxide (DMSO) [analytical grade]
- Ethanol (ETOH) [100% non-denatured for test substance preparation]
- Distilled water

2.3.1.4 *Cell Culture Materials*

Laboratory items needed include the following:

- Sterile, disposable tissue culture plasticware (e.g., 25 cm^2,75-80 cm^2 flasks; multiwell/microtiter [96-well] plates; petri dishes) [Note: The laboratories in this study used tissue culture plasticware from various suppliers.]
- Cryogenic ampules
- Pipettes, pipettors, pipette tips
- Multichannel solution reservoirs
- Centrifuge tubes
- Microporous sterilization filters
- General plastic containers
- Glass tubes (for preparation of substance dilutions)

2.3.1.5 *Equipment*

Performance of the NRU tests requires a laboratory equipped with a designated cell culture area. Essential equipment for cell culture work and the NRU test methods include:

- Incubator (37 ºC ±1 ºC, 90% ±5% humidity, 5.0% ±1% CO_2/air)
- Laminar flow clean bench/cabinet (standard: "biological hazard")
- Water bath (37 ºC ±1 ºC)
- Inverted phase contrast microscope (with 10X to 40X objectives)
- Centrifuge (capable of 220 x g)
- Laboratory balance (capable of measuring to 10 mg)

- Spectrophotometer for reading 96-well plates (i.e., microtiter plate reader) equipped with 540 nm ±10 nm filter
- Shaker for microtiter plates
- Cell counter or hemocytometer
- Pipetting aid (e.g., vacuum pipettor unit)
- Pipettes, pipettors (multi-channel and single channel, multichannel repeater pipette)
- Waterbath sonicator
- Refrigerator
- Freezer (to at least -70 ºC)
- Cryostorage container (and liquid nitrogen supply)
- Magnetic stirrer
- Antistatic bar ionizer
- Personal computer
- Osmometer
- pH meter

2.3.1.6 *Culture Medium*

For 3T3 Cells

DMEM containing high glucose (4.5 gm/L) and supplemented with NCS, L-glutamine, penicillin, and streptomycin was used for the 3T3 cells. Heat-inactivated serum was not used in this study. Heat-inactivation of serum is often used to destroy heat-labile components such as complement factors, and microbial contaminants such as mycoplasma (Hyclone® 1996; Mediatech, Inc. 2006). However, some heat-labile complement factors can also be inactivated by the standard cell culture practice of warming serum-containing medium to 37 °C prior to use, and mycoplasma can be eliminated by filtering the medium (e.g., using 0.1 µm pore-size rated filters). Heating serum to 56 °C (heat-inactivation temperature) can destroy other heat-labile components such as growth factors, vitamins, amino acids, and hormones. Loss of these components can diminish the capacity of the serum to promote attachment of cells to culture vessel surfaces and to support cell growth. An additional confounding factor is that the procedure for heat-inactivation is highly precise, and deviation from the basic protocol can create additional issues such as protein denaturation and serum turbidity.

For NHK Cells

Although the contents of the NHK basal culture medium are proprietary, the formulation is based on a commercially available, non-proprietary basal medium (MCDB 153 medium formulation [Tsao et al. 1982]; e.g., MCDB 153 medium - SIGMA-ALDRICH product number #M 7403 http://www.sigmaaldrich.com/sigma/datasheet/m7403dat.pdf). The laboratories recommended this medium for use with the CAMBREX Clonetics® NHK cells because they all had access to this supplier. Other media are acceptable for NHK NRU testing if the performance standards prescribed in the media prequalification protocol are met (see **Appendix B4** and **Section 2.6.3.5**).

The serum-free culture medium used for NHK cells was Clonetics® keratinocyte basal medium (KBM®) supplemented with KBM® SingleQuots® (epidermal growth factor, insulin, hydrocortisone, antimicrobial agents, bovine pituitary extract [BPE]) and Calcium SingleQuots®

(calcium) [all from CAMBREX Corporation] to make keratinocyte complete growth medium. Although the keratinocyte complete growth medium is a defined serum-free medium, it contains BPE collected from bovine pituitary glands. BPE contains growth factors and hormones, and is added to serum-free medium as a mitogenic supplement. Variability in the composition of the BPE could be a factor in cell growth kinetics. However, it is suggested that the undefined BPE components could be replaced with defined growth supplements, such as insulin, epidermal growth factor, and fibroblast growth factor, without adversely affecting the cellular proliferation rates and general physiology of human keratinocytes (Life Technologies, Inc. 1997).

2.3.2 Reference Substance Concentrations/Dose Selection

Each laboratory weighed and dissolved the reference substances on the same day as the start of the exposure period. The highest concentration of dissolved reference substance was identified using the solubility protocol and designated as the 2X stock solution. All reference substance dilutions for an assay were serially derived from this stock solution (see *Guidance Document* for serial dilution methods).

2.3.2.1 *Range Finder Test*

A range finder 3T3 or NHK NRU test was performed to determine the concentrations of a reference substance to be used for the definitive (concentration-response) test (see **Section 2.3.2.2**). The range finder test used eight concentrations of the reference substance prepared by diluting the stock solution using log intervals to cover a large concentration range (e.g., 1:10, 1:100, 1:1000, etc.; up to eight orders of magnitude). The highest concentrations applied to the cells were 10 mg/mL for substances dissolved in culture medium and 1 mg/mL in medium for substances dissolved in DMSO, unless precluded by solubility. ETOH was not used as a solvent for any of the substances in the validation study (see **Sections 2.8, 2.9,** and **2.10**).

If the range finder test did not produce cytotoxicity, then a second range finder test was conducted at higher concentrations (e.g., the highest concentration would be >10 mg/mL if in medium, >1 mg/mL if in DMSO) unless precluded by solubility. If the substance being tested was insoluble or poorly soluble, then more stringent solubility procedures were employed to increase the stock concentration (to the maximum concentration specified in **Appendices B1** and **B2**). If the range finder test produced a biphasic dose-response curve[3] for NR uptake, the concentrations selected for the definitive tests covered the response range that included the lowest concentration that reduced viability by 50% (see **Section 2.6.3.2**).

2.3.2.2 *Definitive Test*

The concentration-response determination is referred to as the definitive test because it is used to determine the IC_{50} value of the substance being tested. The concentration closest to the calculated IC_{50} value in the range finder test served as the midpoint of the eight concentrations tested in the definitive test. In the absence of other information (e.g., knowledge of the slope of the toxic response), the recommended dilution factor was 1.47 ($\sqrt[6]{10}$), which divides a log interval into six equidistant steps (e.g., 10, 14.7, 21.5, 31.6, 46.4, 68.1, 100). The *Guidance Document* considered a progression factor of 1.21 ($\sqrt[12]{10}$) to be the

[3] A biphasic dose-response curve is a dose-response in which cytotoxicity increases (as dose increases), plateaus, and then increases again.

smallest factor practically achievable, and this was the lowest required concentration interval. The PC was tested similarly to the reference substances in the definitive test and the same recommended dilution factors were used (dilution factor at the discretion of the Study Director).

A definitive test was considered successful if it met all of the test acceptance criteria outlined in the NRU protocols. Definitive tests were repeated as per the protocols if the test failed to meet all of the test acceptance criteria. **Section 2.5** addresses the basis for replicate testing.

If minimal or no cytotoxicity was observed in the range finder test, the maximum concentration for the definitive test was determined as follows:

- For Substances Prepared in NHK or 3T3 Medium: A review of the RC chemicals used in this validation study showed that, among water-soluble chemicals, glycerol had the highest reported IC_{50} value (57 mg/mL). To capture this value, and that of other relatively non-toxic chemicals, the highest concentration of a substance applied to the cells in the definitive test was either 100 mg/mL (using 200 mg/mL 2X stock) or the maximum soluble dose if the substance was not soluble at that concentration.
- For Substances Prepared in DMSO: Based on the maximum concentration of DMSO that could be added to culture medium without causing cytotoxicity (i.e., 0.5%), the highest concentration of a substance that could be applied to the cells in the definitive test was 2.5 mg/mL. In the event that the reference substance was not soluble at this concentration, the highest soluble concentration was used.

2.3.3 NRU Endpoints Measured

2.3.3.1 *NRU and Measurement*

After cells were exposed to the reference substance or the controls (PC; VC) for 48 hours, they were washed and incubated with the NR dye at 37 ºC for an additional three hours. The dye was eluted from the cells using a desorb solution and the OD of the resulting solutions were measured using a spectrophotometric microtiter plate reader. Because NR is absorbed by healthy cells, the amount of dye eluted, as measured by the spectrophotometer, is proportional to NRU and thus to the number of live cells present at culture termination. The OD data from the spectrophotometer were recorded on the EXCEL® template. Relative cell viability for each reference substance and the PC was determined using six replicate wells (six wells [minimum of four scorable] in the 96-well plate) per concentration. Cells treated with the VC were considered to have 100% cell viability (i.e., the mean OD of the VC wells = 100% viability). Cell viability in other test wells was computed in reference to the mean VC OD value (i.e., [well OD/mean VC OD] x 100 = % viability).

2.3.3.2 *Determination of IC_{50}, IC_{20}, and IC_{80} Values*

IC_{50} values were determined from the concentration-response curve using a Hill function, which is a four parameter logistic mathematical model relating the concentration of a substance to the response (typically following a sigmoidal shape). Modifications to the Hill function used in later phases of the study are described in **Section 2.6.3**.

Data from the EXCEL® template were transferred to a template designed by the SMT for GraphPad PRISM® 3.0, a commercially available statistical software (GraphPad Software, Inc., San Diego, CA, USA – hereafter known as PRISM® template). The PRISM® template used the Hill function to calculate the IC_{50}, IC_{20}, and IC_{80} concentrations, reported as μg/mL of reference substance in solution. IC_{20} and IC_{80} data were collected for potential use in designing a prediction model for estimating human lethal blood concentrations.

2.3.4 Duration of Reference Substance Exposure

The SMT and laboratory representatives reevaluated the reference substance exposure duration recommended in the *Guidance Document* (ICCVAM 2001b) before initiating the study. The *Guidance Document* recommended an exposure of 24 hours for the 3T3 cells and 48 hours for the NHK cells. However, Riddell et al. (1986) showed large differences in cytotoxicity for 3T3 cells in response to some chemicals, depending on whether the exposure duration was 24 or 72 hours. Although the toxicity induced by substances that damage, for example, cell membranes is likely to be observed in a relatively shorter time, the toxic effects of substances that interfere with cell functions/processes specifically relating to DNA replication (e.g., protein and nucleic acid synthesis) and cell division (e.g., mitotic spindle formation) are more pronounced after longer exposure periods. This occurs because cells are affected only at certain phases of the cell cycle.

IIVS conducted studies to evaluate the effect of exposure durations of 24, 48, and 72 hours and of 48 and 72 hours on the sensitivity of 3T3 cells and NHK, respectively, to six chemicals selected from the list in Riddell (1986). Because the closest fit to the RC millimole regression occurred when a 48-hour exposure duration was used, this exposure duration was selected for use with both cell types in the validation study (Curren et al. 2003) (see **Appendix E**).

2.3.5 Known Limits of Use

2.3.5.1 *Solubility/Precipitation/Volatility*

In vitro test methods cannot be used for substances that cannot be dissolved in media, DMSO, or ETOH at a sufficiently high concentration to induce cytotoxicity in excess of 50%. Also, chemicals that are unstable or exothermic in water cannot be adequately tested with these *in vitro* test methods (as well as *in vivo* methods).

Precipitation of a test substance in the dosing solution or in the culture medium after the substance to be tested has been added can affect the concentration-response and thus reduce the accuracy of the calculated IC_{50}. Some reference substances used in the validation study had precipitates in their medium/DMSO 2X concentrations prior to dilution for application to the test wells. Precipitates were also observed for some substances in a number of test wells after addition of the media/DMSO 1X solutions (see **Section 5.8** and **Table 5-11**) to the cultures and/or at the end of the exposure period.

Volatility was detected for a number of reference substances during the range finder tests by observance of cross contamination (i.e., high cytotoxicity) in VC wells. Plate sealers were used during the definitive tests to control volatility (see **Section 2.6.3** – *Testing Volatile Reference Substances*), and could be used during the range finder tests if the Study Director suspected that the reference substance might be volatile. The use of plate sealers required

additional laboratory training, and some volatile substances were difficult to test even with the use of plate sealers. Furthermore, some test substances (e.g., organic solvents) may react chemically with the plastic in the sealers.

2.3.5.2 *Biokinetic Determinations*

The Workshop 2000 report (ICCVAM 2001a) discussed the role of chemical biokinetics *in vivo* vis-a-vis acute toxicity, as illustrated in the following quote:

> "Results obtained from *in vitro* studies in general are often not directly applicable to the *in vivo* situation. One of the most obvious differences between the situation *in vitro* and *in vivo* is the absence of processes regarding absorption, distribution, metabolism and excretion (i.e., biokinetics) that govern the exposure of the target tissue in the intact organism. The concentrations to which *in vitro* systems are exposed may not correspond to the actual situation at the target tissue after *in vivo* exposure. In addition, the occurrence of metabolic activation and/or saturation of specific metabolic pathways or absorption and elimination mechanisms may also become relevant for the toxicity of a compound *in vivo*. This may lead to misinterpretation of *in vitro* data if such information is not taken into account. Therefore, predictive studies on biological activity of compounds require the integration of data on the mechanisms of action with data on biokinetic behavior."

The 3T3 and NHK NRU test methods do not account for biokinetics.

2.3.5.3 *Organ-Specific Toxicity*

The Workshop 2000 report also addressed concerns about the *in vitro* prediction of organ-specific toxicity, and identified the organ systems for which failure after acute exposure could lead to lethality (i.e., liver, central nervous system, kidney, heart, lung, and hematopoietic system) (ICCVAM 2001a). Each organ system was reviewed individually. Although the 3T3 and NHK NRU test methods do not assess organ-specific toxicity, they may be useful in a test method battery such as that proposed by the Workshop 2000 participants (see **Section 2.3.5.4**).

2.3.5.4 *The Role of Cytotoxicity Tests in an In Vitro Battery Approach for Possible Replacement of In Vivo Acute Toxicity Testing*

A five-step *in vitro* testing scheme was proposed for a test battery that may eventually be demonstrated to be an adequate replacement for rodent acute oral toxicity test methods for regulatory purposes (ICCVAM 2001a).

- Step 1: Perform a physico-chemical characterization and biokinetic modeling.
- Step 2: Evaluate basal cytotoxicity using, for example, the 3T3 or NHK NRU test methods.
- Step 3: Evaluate the potential that metabolism will mediate the basal cytotoxicity effect.
- Step 4: Assess the test substance's effect on energy metabolism.
- Step 5: Assess the ability of the test substance to disrupt epithelial cell barrier function.

The Workshop 2000 participants suggested that implementation of the 5-step testing scheme would require the following:

- Identification of the most appropriate cell culture systems to use based on accuracy, reproducibility, cost, and availability
- Development of a standardized protocol for each test method used in each of the five steps, and validation of each test method using that protocol
- Development of prediction models for the relevant human toxic levels required by regulatory agencies
- Evaluation of the test battery using substances that are appropriate for all endpoints, and then test sufficient substances to develop a prediction model
- Validation of the entire testing scheme and the prediction model

2.3.6 Basis of the Response Assessed

Neutral red is a weakly cationic, water-soluble, supravital dye that stains living cells by readily diffusing through the cell membranes and concentrating in lysosomes. The intensity of the dye desorbed from the cells in a culture is directly proportional to the number of living cells. Cell death and/or growth inhibition decreases the amount of neutral red taken up by the culture (see **Section 1.3.1**).

2.3.7 Appropriate Positive, Vehicle, and Negative Controls

2.3.7.1 *Positive Control*

The *Guidance Document* recommended sodium lauryl sulfate (SLS; Chemical Abstracts Service Registry Number [CASRN] 151-21-3) as an appropriate PC for *in vitro* cytotoxicity test methods (ICCVAM 2001b), and historical data are available (e.g., Spielmann et al. 1991). A PC test plate was included with every 3T3 and NHK NRU test method assay and was treated the same as any reference substance assay plate.

The historical mean PC IC_{50}, standard deviation (SD), and acceptance limits, were determined separately for each laboratory (see **Table 5-3**), based on their individual historical databases (see **Figure 1-2**). The acceptable range for the PC IC_{50} was based on the statistical approach recommended in the *Guidance Document*. In Phase Ib, the IC_{50} limits accepted for the PC tests were within two SD of the historical mean PC IC_{50} value. In the Phase II studies, the IC_{50} limits for PC tests were within 2.5 SD of the historical mean value (i.e., from Phases Ia and Ib). In Phase III, the IC_{50} limits used for the PC were within 2.5 standard deviations of the mean PC IC_{50} from Phases I and II. The exception to this was the FAL NHK data, where only the Phase II data were used as the basis for establishing the acceptable PC range. The SLS data produced by FAL during Phase I was not used in subsequent historical database compilations because FAL used a modified cell culture protocol in Phase II (see **Section 2.6.2.6**).

2.3.7.2 *Vehicle Control*

The VC consisted of complete DMEM (see **Appendix B1**) for 3T3 cells and complete growth medium (Clonetics® KBM® with supplements [see **Appendix B2**]) for NHK cells when the reference substances were dissolved in culture medium. For reference substances dissolved in DMSO, the VC consisted of medium with the same amount of DMSO (0.5% [v/v]) as was applied to the 96-well test plate.

2.3.7.3 *Negative Control*

Negative control cultures (i.e., those that were not exposed to the solvent) were not used in this validation study. Neither DMSO, at the concentration used, nor the culture medium affected the performance of the 3T3 and NHK NRU test methods.

2.3.8 Acceptable Ranges of Control Responses

The *Guidance Document* established an absolute value (i.e., uncorrected for blank absorbance) range of the OD_{540} for the VC to indicate whether the cells seeded in the 96-well plate had grown with a normal doubling time during the assay. A mean $OD_{540} \geq 0.3$ was recommended as the acceptable range of VC responses and was made a test acceptance criterion for both cell types at the start of the study. However, prior to Phase II, this was rescinded as a test acceptance criterion. The protocols for Phases II and III provide a range of OD values for use as guidance in future studies with these test methods (**Table 2-1**).

Table 2-1 Measured VC OD_{540} Values[1] and Targets

Laboratory	Phase Ia	Phase Ib	Phase II	Phase III
3T3 NRU Test Method				
Target Range[2]	0.3≤ OD ≤1.1	0.30≤ OD ≤0.80	0.103≤ OD ≤0.813	0.103≤ OD ≤0.813
ECBC	0.326 – 0.457	0.214 – 0.839	0.217 – 0.730	0.191 – 0.797
FAL	0.490 – 0.780	0.247 – 0.742	0.289 – 0.768	0.126 – 1.161
IIVS	0.336 – 0.538	0.319 – 0.598	0.307 – 0.578	0.256 – 0.544
NHK NRU Test Method				
Target Range[2]	0.3≤ OD ≤1.1	0.60≤ OD ≤1.70	0.35 ≤ OD ≤ 1.50	0.205 ≤ OD ≤ 1.645
ECBC	0.863 – 2.312	0.788 – 1.282	0.139 – 1.175	0.114 – 1.344
FAL	0.484 – 1.698	0.146 – 1.706	0.110 – 1.292	0.183 – 1.347
IIVS	0.550 – 1.883	0.487 – 1.001	0.201 – 0.841	0.430 – 0.834

Abbreviations: VC=Vehicle control; OD_{540}=Optical density at 540 nM; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; NHK=Normal human epidermal keratinocytes; ECBC=Edgewood Chemical Biological Center; FAL=Fund for Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Lowest to highest OD values for tests that meet test acceptance criteria.
[2]Ranges used for all laboratories. Ranges for Phases Ia and Ib were test acceptance criteria. Ranges for Phases II and III were used as target ranges, rather than as test acceptance criteria.

In Phase III, 99.5% (914/919) of all 3T3 mean VC OD values and 97% (913/944) of all NHK mean VC OD values were within the target ranges. Most OD values outside the ranges were from range finding tests and were usually the result of volatile reference substances affecting the VC cells adjacent to the highest reference substance concentration wells.

The VC OD values had a tendency to be lower in Phases II and III as compared to Phases Ia and Ib. Protocol revisions made throughout Phases Ia, Ib, and II (as listed below) most likely contributed to the differences in the OD values. Possible explanations for changes in OD values for the 3T3 cells include:

- Some tests in Phases Ia and Ib exhibited NR crystals that caused higher OD readings.
- Cell seeding densities were revised from 2.5×10^3 cells/well to a range of $2.0 – 3.0 \times 10^3$ cells/well.

Possible explanations for changes in OD values for the NHK cells include:

- The minimum percent confluence of cells necessary before the reference substance could be applied was reduced from 30% to 20% confluence.
- Cell growth was reduced in some tests in the later study phases as a result of medium and supplement issues (e.g., certain lots of basal medium and medium supplements for NHK cells did not provide optimum growth conditions for the keratinocytes).

2.3.8.1 *Vehicle Controls as a Quality Control Tool*

To check for systematic cell seeding errors and reference substance volatility, VCs were placed both at the left side (row 2) and the right side (row 11) of the 96-well plate (see **Figure 1** in **Appendix B1**). Volatile reference substances generally affected the left side VC, which was next to the highest reference substance concentration in the 96-well plate. The test acceptance criterion for the VC was that the means for the left and the right set of VCs had to be within 15% of the mean of all VCs. This criterion, which was adopted from the protocols in the *Guidance Document* (ICCVAM 2001b), was used for reference substances and the PC in all phases of the validation study.

2.3.9 Nature of Experimental Data Collected

Each laboratory maintained a study workbook to document all aspects of the study and included the raw data for all steps of each assay (e.g., cell growth, test substance treatment, weighing and dilution of reference substances), as well as for all solubility studies.

2.3.9.1 *NRU OD Measurements*

At the conclusion of the NRU desorb step, the OD of the resulting colored solution in each well of the 96-well plates was measured at 540 ±10 nm in a spectrophotometric microtiter plate reader. Each laboratory followed its in-house Standard Operating Procedure (SOP) for use of the microplate readers. These SOPs included instructions for operation and calibration of the instruments. Critical specifications such as alignment, accuracy, reproducibility, and linearity were included as standard parameters for review and routine calibration. Raw OD data from the plate reader was electronically transferred to the EXCEL® template. The template converted the raw data from each treatment well (six wells/reference substance concentration) to derived data by subtracting the mean blank OD value (two blank wells/reference substance concentration) from each reference substance well OD. There were 12 VC wells and 20 associated blank wells. The corrected VC OD values were used to calculate the mean VC OD, which was then used to calculate relative viability (% of mean VC OD) in each test well for the reference substance or PC. The percent viability values were then transferred to the PRISM® template to calculate the IC_{20}, IC_{50}, and IC_{80} values.

2.3.9.2 *Information and Data Collected*

Originals of the raw data (i.e., the Study Workbook and computer printouts of absorbance readings from the plate reader) and copies of other raw data, such as instrument logs, were collected and archived under the direction of the Study Director according to Good Laboratory Practice (GLP)-compliant procedures.

The Study Director/technicians entered the following information into the EXCEL® template:

- Testing identification for: test facility, chemical code, study number, 96-well plate number, experiment number

- Reference substance preparation: solvent used, solvent concentration in dosing solutions, highest stock concentration, dilution factor, pH of 2X dosing solutions, medium clarity/color, presence/absence of precipitate in 2X solutions, PC concentration range
- Cell line/type: cell supplier, lot number, cryopreserved passage number, passage number in assay
- Cell culture conditions: medium, supplements, suppliers and lot numbers, serum concentrations
- Timeline: dates of cell seeding, dose application, OD_{540} determination
- Raw data: OD values from each well from the microtiter plate reader
- Test results: mean corrected OD_{540} value, Hill function R^2 value, logs of IC_{20}, IC_{50}, and IC_{80} (PRISM® template presents data as logs of the IC_x; EXCEL® converts values to μg/mL)
- Test acceptance criteria: acceptable number of values on each side of the IC_{50} (i.e., number of points >0 and ≤50% viability, and >50 and <100% viability), acceptable percent difference for the VCs, acceptable Hill function R^2 value (coefficient of determination) and calculated IC_{50} concentration for the PC
- Visual observations: protocol codes for cell culture conditions for all reference substance concentrations (i.e., relative level of cell cytotoxicity, cell morphology, presence of precipitate)

2.3.10 Data Storage Media

Raw and derived data from the NRU tests were saved in the EXCEL® template file format provided by the SMT. All EXCEL® and PRISM® files were copied and transferred to compact disks. NICEATM and the laboratories printed copies of all data sheets (stored at NICEATM and at the testing facilities), and included copies in the laboratories' final reports.

2.3.11 Measures of Variability

Each 96-well plate used in the NRU tests had three main measures of variability.

1) Each plate contained VCs on each end of the plate (columns 2 and 11) (see **Figure 1** in **Appendix B1** for plate map). The difference between the mean NRU OD for each VC column and mean of the pooled VC wells was used as a test acceptance criterion. The Study Director rejected the test if the difference was greater than 15%, which indicated cross-contamination from a volatile substance or possible cell seeding errors.
2) A mean relative viability was determined for each concentration of the substance tested along with the SD and coefficient of variation (%CV=SD/mean x 100).
3) Macros were included in the EXCEL® template to perform an outlier test (Dixon and Massey 1981) on the data for the six replicate wells for each concentration. Outliers (i.e., individual well values that exceeded the 99% confidence interval [CI] for the replicate wells) were highlighted and could be excluded from the resulting analysis to improve curve fit. The Study Director made the decision as to whether or not to remove outliers and provided a justification for the decision.

Other test-to-test measures of variability were considered in this study.

- Each set of assays for reference substances included a PC plate. If the SLS PC test did not meet test acceptance criteria, then the tests for the associated reference substances were rejected. The SMT recommended testing a manageable number of definitive test plates (e.g., 4 to 6) with each PC to limit the number of definitive NRU tests rejected for PC failure. In this validation study, 4.2% of all definitive tests performed were rejected because the PC failed (i.e., the PC IC_{50} was outside the acceptable confidence limits).
- SDs and CVs were determined for mean IC_{50} values from replicate tests. Replicate testing included three definitive tests for each reference substance, each performed on a different day.

2.3.12 Methods for Analyzing NRU Data

Relative cell viability for each reference substance concentration was calculated using the ODs of the six replicate values (minimum of four acceptable replicate wells) per test concentration. Relative cell viability was expressed as a percentage of the mean VC OD. Absolute OD data from the microtiter plate reader was transferred to the EXCEL® template for performance of these calculations. Where possible, the concentration range (eight concentrations) tested for each reference substance ranged from no effect to 100% toxicity.

The IC_{20}, IC_{50}, and IC_{80} values were determined from the concentration-response curve using the PRISM® template and applying a Hill function to the % viability data. The IC_{20} and IC_{80} values were calculated for potential use in the development of a human prediction model (reported elsewhere).

2.3.13 Decision Criteria for Classification of Reference Substances

The 3T3 and NHK NRU test methods will not be used to classify reference substances in hazard categories but rather to aid in setting the starting dose for sequential rodent acute oral toxicity test methods (i.e., the UDP and ATC) (see **Section 10** for an analysis of the estimated animal savings). The RC millimole regression procedure was used to predict a rodent LD_{50} value from an NRU IC_{50} value. **Section 6.3** addresses the accuracy of the 3T3 and NHK NRU test methods for predicting GHS hazard categories when used with IC_{50}-LD_{50} regressions, calculated using a subset of the RC data (i.e., substances with rat oral LD_{50} data).

2.3.14 Information and Data Included in the Test Report

Test and Control Substances

With the exception of the PC, the laboratories tested coded substances and had minimal information about the test substances' properties (see **Section 3.3** for the reference substance information provided to the laboratories). The following describes the test and test substance information that should be included in an NRU test method report.

- Chemical name(s) and synonyms, if known
- The CASRN, if known
- Formula weight, if known
- Purity and composition of the substance or preparation (in percentage[s] by weight)
- Physicochemical properties (e.g., physical state, volatility, pH, stability, chemical class, water solubility)

- Solubilization of the test/control substances (e.g., vortexing, sonication, warming, grinding) prior to testing, if applicable

Information Concerning the Sponsor and the Test Facility

- Name and address of the sponsor, test facilities, study director, and participating laboratory technicians
- Justification of the test method and specific protocol used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time (e.g., use of the PC data)

Criteria for an Acceptable Test

- Acceptable VC differences between each column of wells and the mean of both columns
- Acceptable concurrent PC ranges based on historical data (include the summary historical data)
- Number of toxic points on either side of the IC_{50} (i.e., number of points >0 and ≤50% viability and >50 and <100% viability)

Test Conditions

- Experiment start and completion dates
- Details of test procedures used
- Test concentration(s) used and how they were derived
- Cell type used and source of cells
- Description of modifications made to the test procedure
- Reference to historical data of the test model (e.g., solvent and PCs)
- Description of the evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., IC_{50} values for the reference substance and the PC, absolute and derived OD readings, reported in tabular form, including data from replicate repeat experiments as appropriate, and the means and standard deviations for each experiment)

Description of Other Effects Observed

- Cell morphology, precipitate, NR crystals, etc.

Discussion of the Results

Conclusion

Quality Assurance (QA) Statement for GLP-Compliant Studies

- A statement describing all inspections and other QA activities during the study, and the dates results were reported to the Study Director. This statement will also serve to confirm that the final report reflects the raw data.

During the validation study, the GLP-compliant laboratories, IIVS and ECBC, followed additional reporting requirements provided in the relevant GLP guidelines (e.g., OECD 1998; EPA 2003a, b; FDA 2003).

The SMT and laboratories developed standard forms for data collection (i.e., EXCEL® and PRISM® templates). The solubility test form was derived from a standard form provided by IIVS. The EXCEL® template was an adaptation of a template format presented in the *Guidance Document* (ICCVAM 2001b).

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

The only proprietary components used in these test methods are the NHK cells and the NHK basal culture medium obtained from CAMBREX Clonetics®. All other components are readily available through various scientific product suppliers.

Section 2.3.1.2 describes the NHK cells used in the study and provides the only commercial source. All laboratories throughout the entire study used cells from the same lot. Procedures used to verify the integrity of the NHK cells included comparison of positive control data across laboratories and observations of cell growth throughout the study. If a laboratory reported a problem with the cells, the SMT and Study Directors evaluated the testing parameters to decide if the problem was cell-oriented or if other factors influenced the problem. **Section 2.6.3.5** provides information concerning the resolution of cell-related issues and revisions made to the protocols to address such difficulties.

Section 2.10.1.1 and **Appendices B2** and **B4** provide information about the NHK growth medium, supplements, and commercial source. Problems arose with the keratinocyte growth medium during the study and resolutions and outcomes are addressed in **Sections 2.6.3.5**, **2.6.3.6**, **5.3.4**, and **11.1.2.2**.

Although this study used proprietary components for the NHK NRU test method, cells and medium from the commercial source used in the study are not required for implementation of this test method.

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

The study protocols required each laboratory to test each coded reference substance in at least one range finding test using a log dilution factor, and in at least three definitive tests on three different days using a smaller dilution factor than used in the range finding test. Assays were performed over a number of days to evaluate day-to-day variation. Laboratories tested each coded reference substance until three definitive tests met the test acceptance criteria. Additional testing was often dictated by:

- Chemical issues (low toxicity, volatility, insolubility, and precipitation)
- PC failure
- Technical difficulties such as NR crystal formation

A stopping rule for insoluble reference substances was incorporated into the protocols for Phase III to limit the number of retests (see **Appendices B1** and **B2**):

> "If the most rigorous solubility procedures have been performed and the assay cannot achieve adequate toxicity to meet the test acceptance criteria after three definitive tests, then the Study Director may end all testing for that particular chemical."

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

2.6.1 Phase Ia: Laboratory Evaluation Phase

All protocol revisions were implemented during Phase Ia unless otherwise stated.

2.6.1.1 *NR Dye Crystals*

NR dye crystals formed in the 96-well test plates when used at 50 μg/mL (OD values measured in the blanks increased from ~ 0.05 to 0.10) in both NRU test procedures. Troubleshooting efforts included incubating the NR medium overnight; centrifuging and filtering the NR medium prior to application to the 96-well plates; and reducing the concentration of NR dye. The laboratories performed tests using a reduced NR concentration of 33 μg/mL. Since there were no quantitative differences in results between tests with 50 μg/mL and tests with 33 μg/mL NR, the SMT accepted tests with both concentrations.

Protocol Revision: The NR dye concentration was reduced to 33 μg/mL for both cell types in subsequent test Phases.

2.6.1.2 *3T3 Cell Growth*

The growth rate of 3T3 cells (as determined by monolayer confluence) was slower than expected. As a result, the cells required more time in culture to obtain the proper density after seeding.

Protocol Revision: The 3T3 cells must be passaged 2-3 times after thawing before being used for the test. The protocol also emphasized attainment of the appropriate percentage of cell confluence (not more than 50% for 3T3 cells) required at the time the cells were exposed to the reference substance, rather than using the time in culture as the guide.

2.6.1.3 *NHK Cell Growth*

The NHK cells had an additional growth problem that manifested as a ring of dead/dying cells around the center of the wells. Troubleshooting efforts included evaluating various brands of 96-well plates (laboratories were not required to use the same brand of plates) and eliminating the change of medium prior to reference substance treatment. All laboratories participated in evaluating the effect of changing (i.e., refeeding) or not changing (i.e., no refeeding) the medium by performing a small study with the PC (SLS). Tests were performed: 1) after refeeding the cells with fresh medium, and 2) by adding SLS to the medium already on the cells. Control ODs were generally higher in the tests in which the medium was not replenished, but sensitivity to SLS was generally unchanged (see **Table 2-2**). FAL was experiencing difficulties in NHK cell growth at this stage of the study which may account for the difference in the refeeding and no refeeding SLS IC_{50} values. The SMT accepted tests with refeeding and those without refeeding (for Phase Ia) as long as they met the test acceptance criteria.

IIVS presented detailed information on the ring of dead cells issue (Raabe 2004). The laboratory showed that the ring of cell death coincided with the formation of a meniscus resulting from the residual medium left in the well after removal of the spent medium. The problem was resolved by eliminating the removal of medium before applying test chemical rather than requiring a standard brand of 96-well plates.

Protocol Revision: Step 2 of the NHK NRU test method was eliminated (change of medium prior to addition of reference substance). The volume of medium (with cells) was changed from 250 μL/well to 125 μL/well.

Table 2-2 Refeeding/No Refeeding Data for the NHK NRU Test Method

	ECBC		IIVS		FAL	
	Refeed	**No Refeed**	**Refeed**	**No Refeed**	**Refeed**	**No Refeed**
Number of Test Plates	4	4	6	6	2	4
Absolute OD[1] for VC	0.265 ±0.151	0.621 ±0.322	0.885 ±0.057	1.12 ±0.033	1.41 ±0.127	1.24 ±0.430
OD[1] for SLS IC_{50}	0.102 ±0.079	0.282 ±0.165	0.415 ±0.029	0.533 ±0.017	0.696 ±0.065	0.606 ±0.217
SLS IC_{50} (µg/mL)[1]	3.33 ±0.47	3.23 ±0.61	3.41 ±0.58	3.49 ±0.39	6.21 ±0.88	8.14 ±0.40

Abbreviations: NHK=Normal human epidermal keratinocyte; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; IIVS=Institute for *In Vitro* Sciences; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; VC=Vehicle control; OD=Optical density; SLS=Sodium lauryl sulfate.
Note: OD values for SLS IC_{50} were extrapolated from the concentration-response curve data
[1]Mean ± standard deviation (uncorrected for blank absorbance

FAL, in contrast to the other two laboratories, used 80 cm^2 culture flasks for culturing the thawed cells from the ampules of cryogenically-preserved pool of cells and encountered difficulties in obtaining a satisfactory number of adhering NHKs.

Protocol Revision (FAL only): Culture flasks were coated with fibronectin-collagen to promote cell adherence.

2.6.1.4 *Vehicle Control OD Limits*

In Phase I, the acceptable range of VC OD values designated in the protocols ($0.3 \leq OD \leq 1.1$) were frequently unattainable in both test methods. Despite this, the Study Directors reported that the cells were adequately responsive. The SMT withdrew the VC OD limits as a test acceptance criterion.

Protocol Revision for Phase Ib: OD ranges were provided as guidelines for each cell type based on OD data from all laboratories, a review of the concentration-response data, and the ability of each test to pass the other test acceptance criteria. Each laboratory developed its own VC OD acceptability range based on its historical data.

2.6.1.5 *Precipitate Formation*

During solubility testing, it was observed that some substances, when tested at the same concentrations, precipitated in the 3T3 medium but not in the NHK medium. When a liquid reference substance (i.e., 2-propanol) produced this effect, the precipitate was attributed to the protein in the serum in the 3T3 medium rather than insolubility.

Protocol Revision: The reference substances were dissolved in 3T3 medium without NCS to make the 2X solutions. The dissolved 2X reference substance was added to medium containing 10% NCS to reach the final 5% NCS and 1X reference substance concentrations.

2.6.1.6 *Dilution Factor*

After a range finder test was performed, the definitive tests were to be performed using a $\sqrt[6]{10}=1.47$ dilution scheme centered on the IC_{50} that was calculated from the range finder. In Phase Ia, the Study Directors, for various reasons related to the specific substance being

tested, sometimes deviated from this requirement and used other dilution factors. The SMT agreed that the dilution factor requirements should be modified to allow more flexibility in setting up tests. The SMT accepted data generated using dilution factors other than the recommended 1.47 for definitive tests if all other test acceptance criteria were met. The use of smaller dilution factors generally increased the number of concentrations in the 10% to 90% viability range, which improved the precision of the IC_{50} calculation.

Protocol Revision: The $^{6}\sqrt{10}$=1.47 dilution scheme was a suggested starting range, rather than a specific test acceptance criterion in subsequent test Phases.

2.6.1.7 *Test Acceptance Criteria*

The test acceptance criteria at the beginning of Phase Ia were:

- The IC_{50} for SLS had to be within the 95% CI of the historical PC mean established by the Test Facility (*rescinded after commencement of Phase Ia*)
- The OD_{540} of the VCs (with blank subtracted) had to be ≥0.3 and ≤1.1 (*rescinded after commencement of Phase Ia*)
- Mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test plate) must not differ by more than 15% from the mean of all VC OD values
- At least two cytotoxicity values, one on either side of the IC_{50} but between 10% and 90% viability, needed to be present (*added after commencement of Phase Ia*)
- The Hill function curve fits (R^2 >0.9 or 0.8 < R^2 <0.9) were evaluated on a case by case basis for acceptability by the SMT (*added after commencement of Phase Ia*).

2.6.2 Phase Ib: Laboratory Evaluation Phase

All protocol revisions developed during Phase Ia were implemented during Phase Ib unless otherwise stated.

2.6.2.1 *NR Crystal Formation*

FAL and ECBC routinely observed NR crystals forming in the 96-well test plates in the 3T3 NRU tests when 33 µg/mL NR was used. All laboratories tested 25 and 33 µg/mL NR concentrations and 2- and 3-hour NR incubation periods to determine which NR concentration and incubation period would provide optimal NRU measurements without crystal formation. In addition to determining whether NRU had reached a plateau at these concentrations and incubation times, the laboratories also determined whether the response to SLS differed under these conditions. Crystals were observed only at 33 µg/mL NR when present for three hours. **Figure 2-2** shows that the average OD results were similar for all NR concentrations and incubation periods tested. **Figure 2-3** shows that the SLS IC_{50} values were equivalent at the different NR concentrations and incubation periods. To minimize changes to the 3T3 protocol, the NRU concentration was lowered from 33 to 25 µg/mL, while the NR incubation period was maintained at three hours. The NR concentration and the incubation period for the NHK NRU test method remained at 33 µg/mL and three hours, respectively.

Protocol Revision for Phase II: The NR concentration for the 3T3 NRU test method was reduced to 25 µg/mL for the three-hour incubation period. Revised methods for preparation

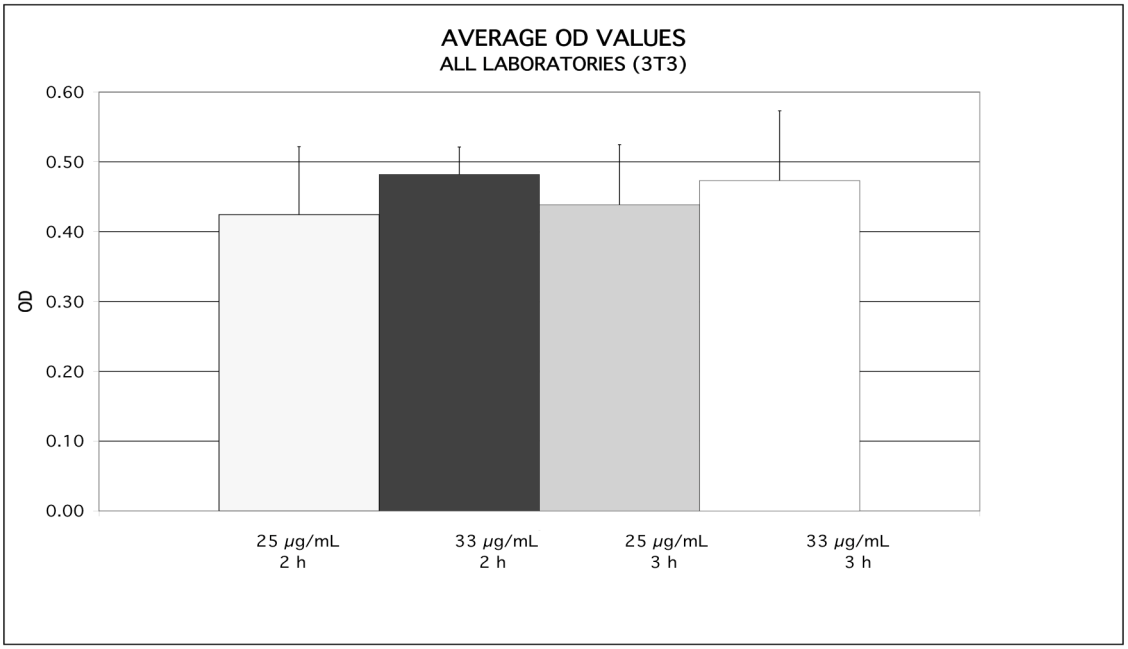
of the NR dye solution included filtration of the solution, maintenance of the solution at 37 ºC prior to application to the cells, and application of the NR solution to the cells within 15 minutes after removing it from 37 ºC. Also, cells were observed during the NR incubation period to monitor possible crystal formation.

2.6.2.2 *Heating of Reference Substance Solutions*

The laboratories had difficulty solubilizing arsenic trioxide, one of the reference substances used in Phase Ib. Heating and mechanical applications for increasing the laboratory's ability to solubilize substances into culture medium were reviewed and revised.

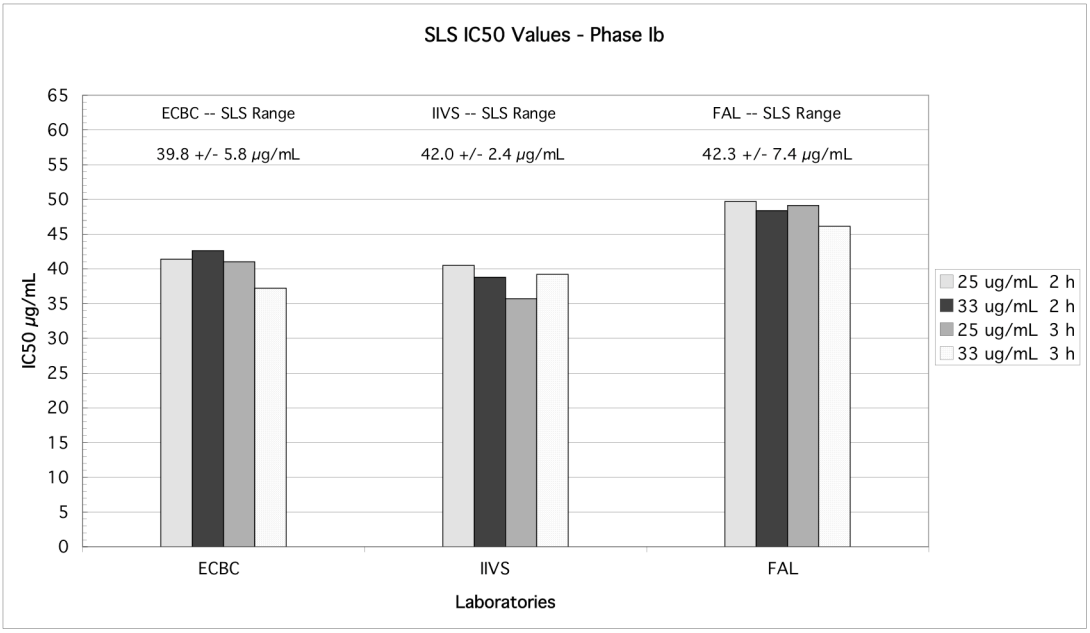
Protocol Revision for Phase II: The duration range for heating a stock solution at 37 ºC (if heating is needed) was increased from 5 to 10 minutes to 5 to 60 minutes.

Figure 2-2 3T3 NRU OD for SLS as a Function of NR Concentration and Duration



Abbreviations: OD=Optical density; NR=Neutral red; SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; h=Hours.
Note: Error bars are one standard deviation.

Figure 2-3 SLS IC_{50} Values for Each NR Concentration and Incubation Duration (3T3 NRU)



Abbreviations: SLS=Sodium lauryl sulfate; IC_{50}=Test substance concentration that reduces cell viability by 50%; NR=Neutral red; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; IIVS=Institute for *In Vitro* Sciences; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory.
Note: SLS range is mean IC_{50} value ± one standard deviation.

Protocol Revision for Phase II: The duration range for heating a stock solution at 37 ºC (if heating is needed) was increased from 5 to 10 minutes to 5 to 60 minutes.

2.6.2.3 *Growth of Untreated Cells*

VC OD values were frequently lower than specified in the Phase I acceptance criteria. Phases Ia and Ib incorporated the acceptance limits shown in **Table 2-1** for the VC, but the limits were rescinded as test acceptance criteria for Phase II because the laboratories frequently failed to meet them even though cell growth and responsiveness to SLS was adequate.

Protocol Revision for Phase II: The specified VC OD range was eliminated as a test acceptance criterion. The OD data (all laboratories combined) from the VCs for both cell types was used to calculate OD ranges that would serve as guidelines for other tests (see **Section 2.2.9).**

2.6.2.4 *Correction of Reference Substance OD Values*

Each reference substance concentration was applied to six treatment wells and to two cell-free wells (i.e., blank wells) used to generate the background OD_{540} values to adjust for potential interference with the NR dye. The mean blank well OD (absolute OD) for each reference concentration was subtracted from the reference substance concentration ODs to provide the corrected OD for each replicate well.

2.6.2.5 *Laboratory Error Rates*

The SMT determined the Phase 1b error rates (number of tests with errors/total number of tests conducted) for each laboratory (**Table 2-3**) and compiled a list of the types of errors encountered. The vast majority of errors were transcriptional and typographical errors in the data sheets provided to the SMT.

Table 2-3 Error Rates[1] in Phase Ib by Laboratory and Test

Laboratory	NRU Test Method	
	3T3	NHK
ECBC	1/9 (10%)	4/17 (23%)
FAL	42/45 (93%)	12/29 (41%)
IIVS	1/20 (5%)	1/20 (5%)

Abbreviations: NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals In Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences
Note: Most errors were transcriptional and typographical and not technical.
[1]Number of tests with errors/total number of tests (some data files had more than one error)

2.6.2.6 *Resultant Protocol Changes for Phase II*

Following the completion of Phases Ia and Ib, IIVS sponsored a weeklong laboratory training exercise for the cytotoxicity testing laboratories to help standardize the level of training among the technical staff and to identify any further 3T3 and NHK NRU protocol revisions that might be needed. Protocol revisions made because of this exercise included:

- Multi-channel repeater pipettes can be used for dispensing cells into the 96-well plates and dispensing plate rinse solutions, NR medium, and desorb solution but are not accurate enough to dispense the PC or the reference substances to the treatment wells.
- Use of 8-channel reservoirs for applying dosing solutions to the wells so that multi-channel single delivery pipettes could be used
- Use of a standardized length of time that the HBSS rinse remains on the cell monolayers in flasks during the cell subculture step
- Protection of plates from light during the shaking step for NR extraction; all laboratories will cover plates with a light-impermeable barrier (e.g., aluminum foil) during this step
- Allow plates to stand for at least five minutes after the shaking step is complete and eliminate any bubbles in media observed in the wells before measuring the OD
- Change the allowable seeding density range for 3T3 NRU test method from $2.5x10^3$ cells/well to $2 – 3x10^3$ cells/well
- Change the NHK culture flask size used at FAL for start-up of cryopreserved cells from 80 cm^2 to 25 cm^2 (the size the other laboratories had been using), and discontinue using a fibronectin-collagen coating.

2.6.2.7 *Test Acceptance Criteria*

The test acceptance criteria were revised as follows:

- The IC_{50} for SLS (PC) should be within 2 SDs (approximately 95%) of the historical mean established by each laboratory in Phase Ia.
- The mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test plate) should not differ by more than 15% from the mean of all VC OD values on that plate.
- At least one calculated cytotoxicity value should be between 10% and 50% viability, and one value between 50% and 90% viability.
- The Hill function curve fit ($R^2 > 0.9$ or $0.8 < R^2 < 0.9$) should be evaluated on a case-by-case basis for acceptability by the SMT.
- VC OD criteria were based on Phase Ia data (mean ± two SDs): 0.3 to 0.8 for the 3T3 test method, and 0.6 to 1.7 for the NHK NRU test method (requirement for use of VC OD criteria as test acceptance criteria was rescinded after commencement of Phase Ib)

2.6.3 Phase II: Laboratory Qualification Phase

All protocol and acceptance criteria revisions were implemented during Phase II unless otherwise stated.

2.6.3.1 *Testing of Volatile Reference Substances*

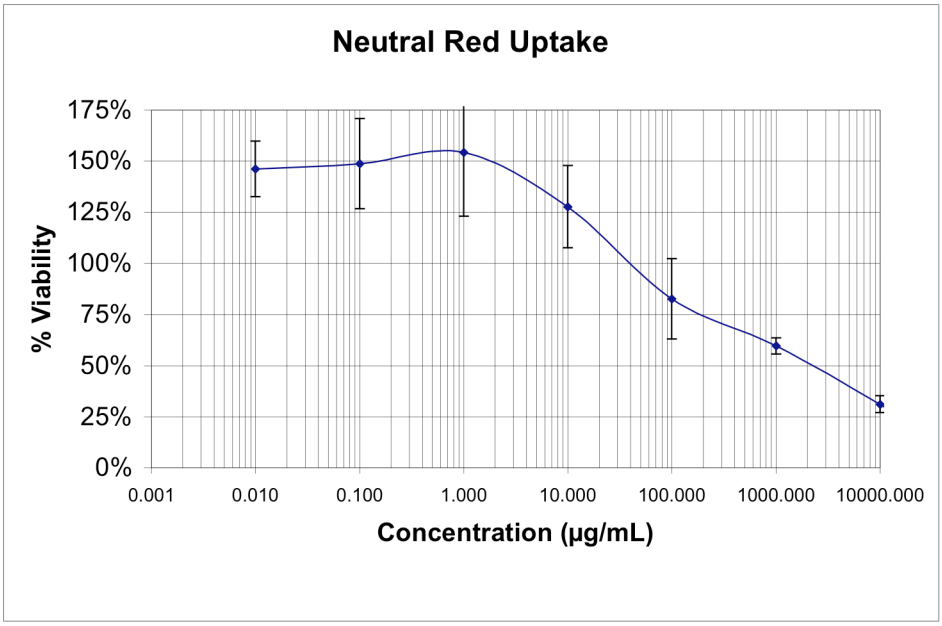
When 2-propanol was tested in 3T3 and NHK cells, vapors from the highest concentration wells contaminated the adjacent VC wells and also appeared to affect some lower concentration wells (i.e., the wells exhibited unexpectedly reduced levels of NRU). An example range finder concentration-response curve is shown in **Figure 2-4**. Such tests failed the VC criterion. When lower concentrations were used to avoid contaminating the VC wells adjacent to the highest concentration, the toxicity was inadequate to produce an IC_{50}. To address this problem, IIVS repeated their tests using film plate sealers, which isolated individual wells from one another; this was sufficient to prevent the cross-well contamination, and acceptable results were obtained. Based on these data, the SMT recommended to the other two laboratories that film plate sealers be used when testing 2-propanol.

FAL had previous experience layering mineral oil on the culture media in a well to prevent volatile substances from escaping, and provided 2-propanol test data where mineral oil had been added to each well. The data showed that the average oil vs. film IC_{50} values were not significantly different. However, there was less variability in the NRU data when using the film sealer so the SMT recommended this methodology.

A >15% difference between the mean VC OD of all VC cells and the mean OD of each VC columns on opposite ends of the test plate was used as a general indicator of substance volatility in the test if the VC adjacent to the highest test concentration had a significantly reduced OD value.

Protocol Revision: The SMT included the use of film sealers in the Phase III protocols when testing suspected volatile compounds.

Figure 2-4 Representative Concentration-Response for 2-Propanol in a 3T3 NRU Range Finder Test



96-WELL PLATE MAP

	1	2	3	4	5	6	7	8	9	10	11	12
A	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank
B	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
C	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
D	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
E	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
F	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
G	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
H	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank

CORRECTED ABSORBANCE (Sample OD_{540} - Mean Blank OD_{540})

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.000	-0.002	-0.001	-0.001	0.000	-0.003	0.001	0.002	0.002	-0.001	-0.002	-0.003
B	0.002	0.080	-0.001	0.070	0.124	0.206	0.296	0.389	0.291	0.301	0.343	0.002
C	-0.001	0.067	0.004	0.059	0.109	0.171	0.284	0.334	0.237	0.308	0.337	-0.004
D	0.003	0.058	0.003	0.056	0.110	0.163	0.243	0.271	0.246	0.251	0.283	0.002
E	0.003	0.077	0.001	0.067	0.106	0.092	0.218	0.252	0.328	0.250	0.290	0.003
F	-0.004	0.068	-0.002	0.050	0.110	0.164	0.216	0.289	0.336	0.267	0.281	-0.001
G	-0.004	0.071	0.003	0.053	0.122	0.147	0.204	0.226	0.263	0.295	0.330	-0.003
H	0.004	0.000	0.001	0.001	0.000	0.003	-0.001	-0.002	-0.002	0.001	0.001	-0.002

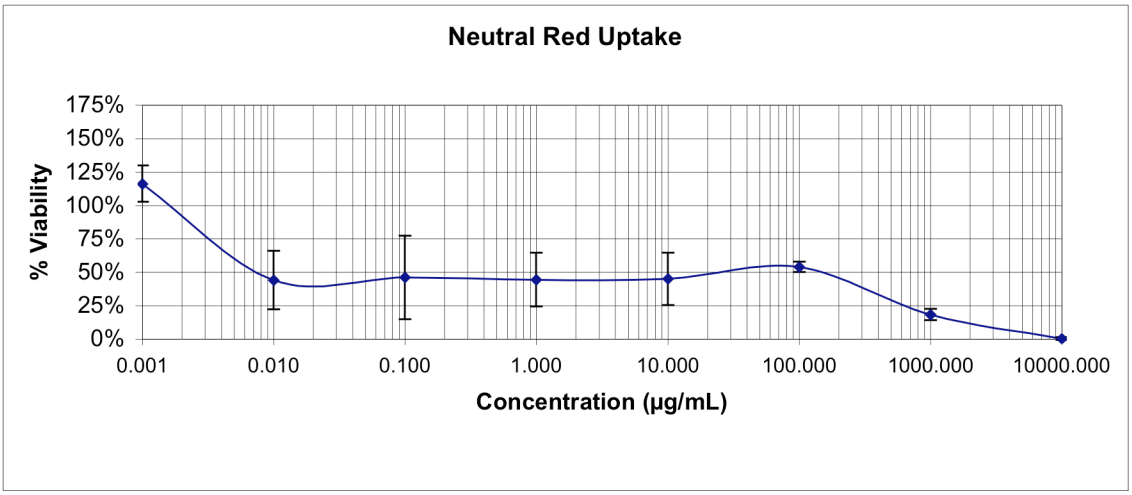
Abbreviations: 3T3=BALB/c 3T3 fibroblasts; VC=Vehicle control; C1 to C8=Test substance concentrations (C1-highest concentration, C8-lowest concentration); OD_{540}=Optical density at 540 nm; A to H=Row identification.
Note: %Difference of the two VC columns from the average VC was 63%. The mean corrected optical density (OD) for VC1, adjacent to the highest 2-propanol concentration, was 0.070, while that for VC2, adjacent to the lowest 2-propanol concentration, was 0.310. Setting the mean VC OD to 100% viability shifted the toxicity curve such that lower concentrations of 2-propanol seemed to be less toxic to the cells than the VCs (i.e., >100%).
Error bars are ±1 standard deviation.

2.6.3.2 *Atypical Concentration-Responses*

Atypical concentration-responses are defined for this study as response curves that differ from a basic sigmoidal shaped curve. Curves that show a biphasic response as well as those that exhibited a plateau-like response at toxicity levels than 100% were considered atypical.

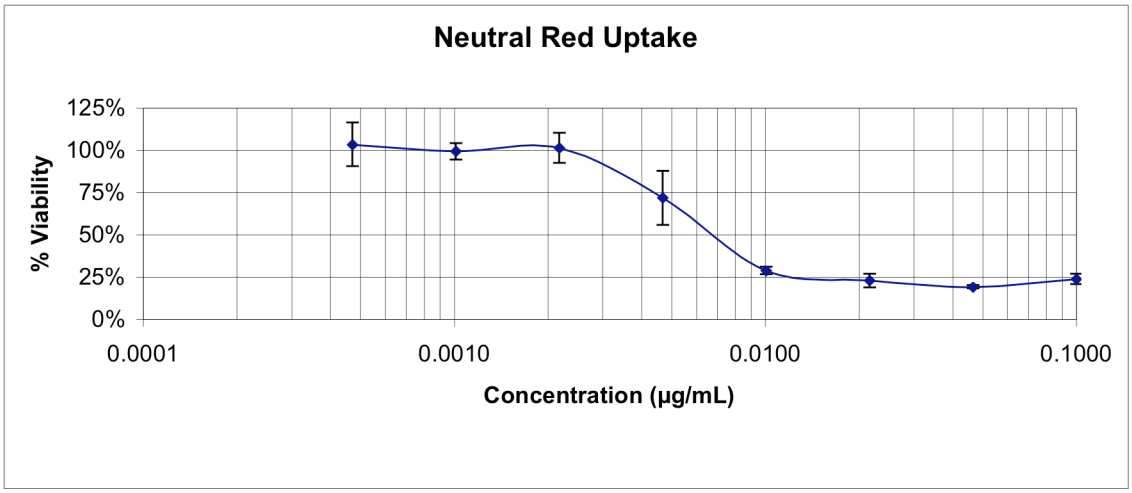
Two of the laboratories observed biphasic concentration-responses in the range finder tests for aminopterin and colchicine. When the range finder tests produced a biphasic response (see **Figure 2-5** for an example), the SMT advised the laboratories to focus the definitive tests on the lowest concentrations that produced at least a 50% loss in viability. Although doing so eliminated the biphasic response in the definitive tests, the highest tested concentrations did not reduce cell viability to 0% (see **Figure 2-6**). This effect with colchicine was very reproducible across laboratories in the NHK NRU test, but only FAL achieved this biphasic type of response with colchicine in the 3T3 NRU test. Aminopterin produced similar concentration-responses in the NHK NRU test at ECBC and FAL, but not at IIVS. In the 3T3 NRU test, only FAL obtained a biphasic response with aminopterin.

Figure 2-5 Representative Concentration-Response for Aminopterin in a NHK NRU Range Finder Test



Abbreviations: NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake.
Representative dose-response for aminopterin in a NHK range finder test. Laboratories were instructed to focus the definitive tests on the lowest concentration that produced a 50% reduction in viability in the range finder test.

Figure 2-6 Representative Concentration-Response for Aminopterin in a NHK NRU Definitive Test



Abbreviations: NHK=Normal human epidermal keratinocytes; NRU=Neutral red uptake.
Note that the maximum reduction in cell viability plateaued at about 75%

Biphasic concentration-responses are not uncommon. Calabrese (2005) states that numerous mechanistic explanations (including hormesis[4]) could account for biphasic response curves. Such concentration-responses could be because the substance acts through more than one mechanism of action (e.g., one mechanism that is active at low test substance concentrations and other mechanism[s]) that are effective at higher concentrations). Conolly and Lutz (2004) also provide examples of pharmacological and toxicological data sets of biologically based mechanisms that could explain biphasic responses. These examples include:

- Membrane receptor subtypes with opposite downstream effects
- Receptor-mediated gene expression
- Induction of DNA repair and "co-repair" of background DNA damage
- Modulation of the cell cycle

Although non-linear responses could also be due to technical error (e.g., improper dosing, unacceptable media, contamination), the responses seen in this study were reproducible, and there was no evidence to suggest that technical errors were involved. The SMT assumed that these responses were based on the chemicals' mechanisms of action. For example, colchicine binds to microtubular protein and interferes with function of mitotic spindles, which arrests cell division (NLM 2003). Aminopterin blocks the use of folic acid by the cells, inhibiting metabolism, RNA production, and protein synthesis, which is lethal during the S phase of the cell cycle by (NLM 2002). The variability of IC_{50} results for these substances among the laboratories may be due to different levels of cell confluence in the cultures at the time of treatment.

[4] Hormesis is a dose-response characterized by a compound's ability to produce an opposite effect at low doses compared with its effect at high doses (e.g., stimulatory at low doses and inhibitory at high doses).

2.6.3.3 *Hill Function*

The Hill function used in the various phases of this study was defined as follows:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC50 - \log X)\text{HillSlope}}}$$

where Y=response (i.e., % viability), X is the substance concentration producing the response, Bottom is the minimum response (0% viability, maximum toxicity), Top is the maximum response (maximum viability), EC_{50} is the substance concentration at the response midway between Top and Bottom, and HillSlope describes the slope of the curve. When Top=100% viability and Bottom=0% viability, the EC_{50} is the equal to the IC_{50}.

Responses that do not achieve 100% cytotoxicity with increasing substance concentration do not fit the Hill function well. The R^2 values from such tests often failed the acceptance criterion. To obtain a better model fit, the Bottom parameter was estimated without constraints (the previous practice was to use Bottom=0). However, when Bottom≠0, the EC_{50} reported by the Hill function was not the same as the IC_{50} because the Hill function relies on $EC_{50,}$ which is defined as the point midway between the Top and Bottom responses. Thus, the Hill function calculation using the Prism® software was rearranged to calculate the IC_{50} as follows:

$$\log IC_{50} = \log EC_{50} - \frac{\log\left(\frac{\text{Top} - \text{Bottom}}{\text{Y} - \text{Bottom}} - 1\right)}{\text{HillSlope}}$$

where IC_{50} is the concentration producing 50% toxicity, EC_{50} is the concentration producing a response midway between the Top and Bottom responses; Top being the maximum response (maximum survival), Bottom is the minimum response (0% viability, maximum toxicity), Y=50 (i.e., 50% response), and HillSlope describes the slope of the response. The X from the standard Hill function equation is replaced, in the rearranged Hill function equation, by the IC_{50}.

IIVS performed the recalculations for their NHK NRU colchicine tests and the SMT performed the necessary recalculations for the other laboratories. Tests that were recalculated by the SMT are noted in the data summaries.

Protocol Revision: The protocol was revised to state that if a range finding test produces a biphasic response, then the concentrations selected for the subsequent tests should cover the most toxic dose-response range.

2.6.3.4 *Insoluble Reference Substances*

Lithium carbonate was insoluble in 3T3 medium. Only ECBC managed to expose 3T3 cells to sufficient lithium carbonate to produce three tests that met the acceptance criteria. Precipitate was reported for two of those tests at the three highest concentrations in the wells. Because the third highest concentration, 510.2 μg/mL, was approximately the IC_{50} (average was 564 μg/mL), the true IC_{50} for lithium carbonate may actually be lower than was

calculated, and therefore the LD_{50} value would be underestimated. However, the data were reproducible and were not discarded.

Protocol Revision for Phase III: The protocol was revised to allow an increase in the stirring/rocking duration in an incubator from one to three hours if cytotoxicity in the range finder test was limited by solubility. Also, a *Stopping Rule for Insoluble Chemicals* was added (see **Section 2.5** and **Appendices B1** and **B2**) so that the laboratories would not continue repeated testing of insoluble substances in order to obtain three acceptable definitive tests.

2.6.3.5 *Inadequate Cell Growth in NHK Medium*

IIVS and FAL had several NHK NRU test failures that were attributed to poor cell growth. The SMT compiled KBM® and SingleQuot® lot numbers that the laboratories were using, along with the laboratory assessments of NHK cell growth. The information was used to identify the lots that produced adequate growth. The SMT also obtained quality assurance and quality control test results from CAMBREX Clonetics® on the lots of KBM®, but the information provided was inadequate for determining how the medium would perform in the NHK NRU test method.

Resolution: A protocol for prequalifying the medium was developed (see **Appendix B4**). For Phase III, the SMT asked IIVS to prequalify new lots of KBM® and SingleQuots® for use by all laboratories.

2.6.3.6 *Performance Standards for Media to Support NHK Growth*

A prequalification-of-medium protocol (**Appendix B4**) was developed and used by IIVS to test several different lots of medium and supplements to find combinations that maintained the typical growth characteristics of the NHK cells used in this study. The laboratories then reserved samples of the acceptable lots at CAMBREX so that testing would not be interrupted due to unavailability of adequate materials.

Test Acceptance Criteria for Prequalifying Media Using SLS

- The fit of the SLS dose-response to the Hill model should be $R^2 \geq 0.85$ (i.e., from PRISM® software).
- The difference between the mean of all VCs and (a) the left mean VC, and (b) the right mean VC should be ≤15%.
- At least one concentration should exhibit >0% and ≤50% viability and at least one should exhibit >50% and <100% viability.
- After meeting all other acceptability criteria, the SLS IC_{50} must be within the historical range (±2.5 SD) established by the laboratory.

Other Criteria for Prequalifying Media (for consideration by a Study Director)

- General observations: rate of cell proliferation; percent confluence; number of mitotic figures per field; colony formation; distribution of cells in the flask; absence or presence of contamination
- Cell morphology observations should include overall appearance (e.g., good, fair, poor), and presence of abnormal cells

- Mean corrected OD_{540} of the VCs (e.g., are the values high/low when compared to historical data)
- Cell morphology and confluence of the VC wells at the end of the 48-hour treatment
- Cell doubling time, as compared to the doubling time with the previous batches of medium

2.6.3.7 *Test Acceptance Criteria for Phase II*

- The IC_{50} for SLS (PC) should be within 2.5 SDs of the historical mean established by the laboratory (*Phases Ia and Ib*)
- Mean OD values of the left and right VCs (columns 2 and 11 in the 96-well test plate) do not differ by more than 15% from the mean of all VC well OD values. At least one calculated cytotoxicity value ≥10% and ≤50% viability and at least one value >50% and ≤90% viability
- R^2 ≥0.90. The test fails if R^2 <0.80. If the 0.80 ≤ R^2 <0.90, the SMT evaluates the model fit (Note: The Study Director makes this determination for non-validation studies.)

2.6.4 Phase III: Laboratory Testing Phase

The changes below were made in the Phase III protocols based on the data and results in Phase II.

2.6.4.1 *Required Cytotoxicity Values*

Obtaining at least one calculated cytotoxicity value >0% and ≤50% viability and at least one that is >50% and <100% viability may be difficult or unattainable for substances with steep dose responses.

Protocol Revision: The test acceptance criterion was qualified so that tests with only one concentration between 0 and 100% viability were acceptable if the smallest practical dilution factor (i.e., 1.21) was used and all other test acceptance criteria were met.

Tests for three reference substances were accepted that met this new criterion in the 3T3 NRU test method: diquat dibromide (1/9 tests); epinephrine bitartrate (2/9 tests); 1,1,1-trichloroethane (2/8 tests). No NHK tests required the use of these criteria (i.e., one point between 0% and 100% viability at the lowest dilution factor).

2.6.4.2 *Revisions to Data Analysis Procedures*

The following revisions to data analysis procedures were made in Phase III NRU protocols:

- If the Bottom parameter of the Hill function was fit to a value <0%, then the parameter was set to zero (0) for the IC calculations.
- If toxicity plateaued above 20% viability (i.e., toxicity was <80%), the IC_{80} was not determined. The IC_{20} and IC_{50} values were calculated from the range of available toxic responses.
- The requirement for substance dose-responses to fit the Hill equation with R^2 ≥0.90 was rescinded. The Hill equation was used to characterize the shape of the response rather than to establish an acceptance criterion. The PC acceptance criterion was modified to R^2 ≥0.85.

2.7 Differences Between the 3T3 and NHK NRU Protocols for the Validation Study and the *Guidance Document* Standard Protocols

As the validation study progressed through Phases I and II, the protocols provided in the *Guidance Document* (ICCVAM 2001b) were optimized to address problems that were encountered during the validation study phases. Changes to the *Guidance Document* protocols are described below.

- 3T3 cell seeding density for 96-well plates was decreased from $1x10^4$ cells/well to $2.0 – 3.0x10^3$ cells/well.
- The calcium concentration in NHK medium was changed from 0.15 mM to 0.10 mM. The test laboratories had expressed concern that cell differentiation would occur at the higher concentration and requested a lower concentration. CAMBREX Clonetics®, the supplier of the NHK cells and NHK medium used in this study, normally grows NHK cells in 0.15 mM calcium and has seen no differentiation. The supplier agreed that the cells would grow well at 0.10 mM but should not be cultured at concentrations <0.10 mM in order to avoid morphological and growth rate changes (CAMBREX technical division, personal communication).
- NHK cells were subcultured once prior to being distributed to the test wells, rather than for three passages. The laboratories expressed concern about the possibility of cell differentiation with subsequent passages in culture.
- The highest recommended final concentrations of DMSO and ETOH in the culture media were reduced from 1% to 0.5%. IIVS performed experiments with both cell types to determine the concentration necessary to avoid solvent toxicity. 3T3 cells were tested with 0.5, 1, and 2% ETOH and DMSO at 0.1, 0.2, 0.3, 0.4, 0.5, 1, and 2% concentrations. The 0.5% concentrations of both solvents were chosen as optimal because that concentration of ETOH produced no toxicity. Although 0.5% DMSO produced slight toxicity (i.e., cells were 91% viable as compared to the control cells; See **Appendix E1**), this concentration was chosen by the SMT and laboratories as an acceptable trade-off between slight toxicity and the ability to test substances at higher concentrations, and was used throughout the study for all reference substances that needed solvents other than culture medium (see Curren et al. 2003). DMSO was the preferred solvent if the test substance was not soluble in culture medium, and ETOH was not used in this study.
- The pH of the reference substance solutions was not adjusted with NaOH or HCl regardless of whether solutions became acidic or basic (optimum mammalian cell culture pH is approximately 7.4 [Freshney, 2000]) upon addition of the test substance because some of the basal cytotoxicity produced by test substances may be due to pH effects. See **Appendix F1** for pH values of the reference substances in culture medium.
- The CO_2 concentration in the incubator was reduced from 7.5% to 5.0% because the laboratories were already set up to use 5% CO_2, which is a typical optimum CO_2 concentration for mammalian cell culture.
- Washing and fixing the cells with a formaldehyde solution prior to NR elution from the cells was eliminated. Formaldehyde disposal was problematic in FAL's regulatory environment. The SMT and the laboratories agreed that the

use of formaldehyde was unnecessary because the NR desorb solution (1% glacial acetic acid, 50% ETOH, and 49% H_2O) adequately fixed the cells to the test plate (INVITTOX 1991).

- Reference substance exposure time for the 3T3 cells was extended from 24 hours to 48 hours (see **Section 2.2.4** and **Appendix E1**).
- Cell culture seeding densities for subculture were provided as guidance, rather than as strict cell number ranges. The laboratories determined adequate cell densities (see **Table 2-4**) based on their own experience with the growth of the cells in the wells, and the time needed to reach the appropriate level of confluence needed for addition of the test substance, the VC, and PC.

Table 2-4 Cell Seeding Densities[1]

Protocol	3T3 cells/cm^2 subculture to flasks	3T3 cells/well 96-well Plate	NHK cells/cm^2 subculture to flasks	NHK cells/well 96-well Plate
Guidance Document[2]	$1.25x10^4$	$2.5x10^3$	$3.5x10^3$	$2 – 2.5x10^3$
Phase Ia	$0.42 – 1.68x10^4$	$2.5x10^3$	$2.5 – 9x10^3$	$2 – 2.5x10^3$
Phase Ib	$0.42 – 1.68x10^4$	$2.5x10^3$	$2.5 – 9x10^3$	$2 – 2.5x10^3$
Phase II	$0.42 – 1.68x10^4$	$2 – 3x10^3$	$2.5 – 9x10^3$	$2 – 2.5x10^3$
Phase III	$0.42 – 1.68x10^4$	$2 – 3x10^3$	$2.5 – 9x10^3$	$2 – 2.5x10^3$

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes
[1]Cell numbers determined by Coulter Counter or hemocytometer
[2]ICCVAM (2001b)

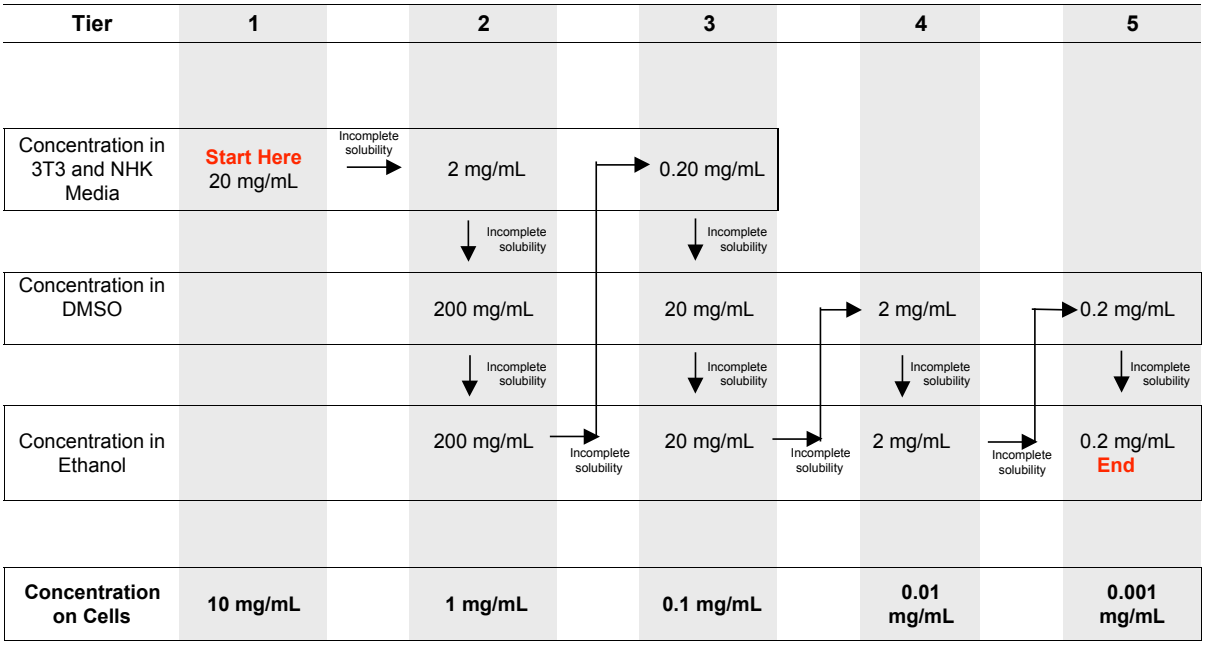
2.8 Overview of the Solubility Protocol

The SMT, with assistance from the laboratories, developed a solubility protocol to provide guidance for determining the most appropriate solvent for each test substance. The solubility protocol was based on an EPA guideline (EPA 1998) that involved testing for solubility in a particular solvent, beginning at a relatively high concentration and proceeding to successively lower concentrations by adding more solvent as necessary for dissolution. Testing stopped when, upon visual observation, the procedure produced a clear solution with no cloudiness or precipitate. The order of selection priority was culture medium, DMSO, and ETOH. Each laboratory tested the solubility of each reference substance using this protocol and provided the data to the SMT prior to initiating cytotoxicity testing. The SMT analyzed the solubility data provided by BioReliance and each testing laboratory, and designated the solvent to be used by all laboratories for each reference substance. This eliminated one potential variable in the NRU test results among laboratories.

The solubility protocol used by the *in vitro* laboratories required the sequential testing of reference substances in the various solvents at concentrations that would be equivalent to the concentration that would be applied to the cell cultures. The solubility flow chart in **Figure 2-7** shows, for example, that 2 mg/mL medium and 200 mg/mL DMSO or ETOH were equivalent concentrations because they yielded 1 mg/mL in cell culture. Medium was diluted by one-half when applied to cultures. The 0.5% [v/v] final concentrations were achieved by diluting DMSO and ETOH by 200-fold. At each concentration, the following mixing procedures were employed, as necessary, to completely dissolve the reference substance in

the sequence: vortex (1 to 2 minutes); sonication (up to 5 minutes); warming to 37 °C (5 to 60 minutes [NRU protocols allow warming to be extended to three hours if cytotoxicity in the range finder test was limited by solubility]). If the reference substance was still not dissolved, the next lower concentration, or a different solvent, was tested.

Figure 2-7 Flow Chart for Determination of Reference Substance Solubility in Medium[1], DMSO, or ETOH



Abbreviations: DMSO=Dimethyl sulfoxide; ETOH=Ethanol; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human epidermal keratinocytes.
Note: DMSO is U.S.P. analytical grade. ETOH is U.S.P. analytical grade (100% non-denatured).
[1]3T3 Medium - DMEM (Dulbecco's Modification of Eagle's Medium) with supplements; NHK medium - KBM® (Keratinocyte Basal Medium) with supplements (from CAMBREX Clonetics®).

2.9 Basis of the Solubility Protocol

The solubility protocol used by BioReliance, which tested solubility of the reference substances prior to testing by the *in vitro* laboratories, is provided in **Appendix G**. The protocol is based largely on information from the literature and Internet searches for solubility procedures, the experience of the SMT and IIVS, and solubility and IC_{50} information from the RC chemicals database (Halle 1998, 2003). The only formal solubility protocol discovered was the EPA Product Properties Test Guideline, OPPTS 830.7840 Water Solubility Column Elution Method; Shake Flask Method (EPA 1998).

2.9.1 Initial Solubility Protocol Development

BioReliance evaluated the solubility of each reference substance in cell culture media at 2000, 400, and 200 mg/mL, and if not soluble at those concentrations, in DMSO and then ETOH, at the same concentrations (initial protocol). It was apparent that these concentrations were not low enough when the laboratory was unable to achieve solubility for arsenic

trioxide. The solubility protocol was revised twice to lower the range of concentrations tested (see **Table 2-5**). An extra tier of concentrations ≤1 mg/mL was added for poorly soluble and insoluble substances. The protocol used by the laboratories was further revised to reduce the number of steps required (by testing in log units) and to test in tiers using concentrations that reflected the concentrations anticipated in the cell cultures (see **Figure 2-7**).

Table 2-5 Comparison of Concentrations Tested in the Various Solubility Protocols

Solubility Protocol Version	Concentrations Tested (mg/mL)					
	Step 1	Step 2	Step 3	Step 4	Step 5	Steps 6-10
BioReliance (1st) (4/26/02) and Phase Ia	2,000	400	200	NA	NA	NA
BioReliance (2nd) (9/17/02)	200	40	20	10	2	NA
BioReliance (3rd) (10/11/02)	200	40	20	10	2	1, 0.5, 0.25, 0.125, 0.05
Phases Ib, II, III for cytotoxicity laboratories	20 Medium	2 Medium 200 DMSO 200 ETOH	0.2 Medium 20 DMSO 20 ETOH	2 DMSO 2 ETOH	0.2 DMSO 0.2 ETOH	NA

Abbreviations: DMSO=Dimethyl sulfoxide; ETOH=Ethanol; Medium=Cell culture medium; NA=Not applicable
Note: DMSO is U.S.P. analytical grade. ETOH is U.S.P. analytical grade (100% non-denatured).

In Phases Ib and II, the SMT used the data from BioReliance to select the solvents to be used for testing the various chemicals. When it became apparent that the laboratories sometimes obtained different solubility results than those reported by BioReliance, the SMT used the cytotoxicity results from the laboratories to determine the solvents to be used for Phase III reference substances.

The final protocol provided a tiered approach for determining the 2X stock concentration for each reference substance (see **Figure 2-7**). This protocol had the advantage of reducing the number of steps for testing (compared to that used by BioReliance) (see **Appendix B3**).

2.9.2 Basis for Modification of the Phase II Protocol

All three testing laboratories found arsenic trioxide (tested in Phase Ib) less soluble (see **Table 5-10**) than was reported by BioReliance (BioReliance values: 0.25 mg/mL in 3T3 medium and 0.05 mg/mL in NHK medium). This chemical was not soluble using the procedures in the initial solubility protocol. IIVS warmed the stock solution (at least 200 μg/mL for 2X) for longer than the protocol specified (i.e., 30 to 50 minutes) but still had persistent, small, undissolved particles. ECBC obtained a clear solution (highest 2X concentration was 30 to 50 μg/mL), but found precipitated particles after the solution stood at room temperature. Sonication time was increased to 15 to 30 minutes, and heating time to approximately 30 minutes to get a finer suspension. This procedure achieved a more homogeneous mixture, resulting in more uniform serial dilutions and a more even application of the reference substance to the cells. FAL stirred the suspension (approx. 20 to 90 μg/mL) in the CO_2 incubator for 1.5 to 2 hours to get clear medium.

Protocol Revision for Phase II: The duration of the heating step was altered from 5 to 20 minutes to 5 to 60 minutes.

2.10 Components of the Solubility Protocol

2.10.1 Medium, Supplies, and Equipment Required

2.10.1.1 *Medium and Chemical Supplies*

- 3T3 culture medium: DMEM without L-glutamine and containing Hanks' salts and high glucose [4.5gm/l]; L-glutamine, 200 mM; NCS
- NHK culture medium: Keratinocyte Basal Medium without Ca^{++} (KBM®, Clonetics® CC-3104); KBM® SingleQuots® medium supplements (Clonetics® CC-4131): epidermal growth factor, insulin, hydrocortisone, bovine pituitary extract; Calcium SingleQuots® (Clonetics® CC-4202); penicillin/streptomycin solution (antimicrobial agents)
- United States Pharmacopoeia (U.S.P.) analytical grade DMSO
- U.S.P. analytical grade (100%, non-denatured) ETOH

2.10.1.2 *Equipment*

- Waterbath (37 °C)
- Sonication apparatus
- Vortex mixer
- Micropipettors
- Balance (capable of weighing 10 mg)
- pH meter

2.10.1.3 *Procedures*

The Phase III solubility protocol required the dissolving of approximately 10 mg of reference substance in approximately 0.5 mL medium (both 3T3 and NHK media were used) for a final concentration of 20 mg/mL (see **Appendix B3**). In order, the mixture was vortexed for 1 to 2 minutes, sonicated for up to 5 minutes, and warmed to 37 °C for 5 to 60 minutes, as necessary, to dissolve the substance. The endpoint for dissolution was a clear solution with no noticeable precipitate. If the reference substance was not soluble in medium at 20 mg/mL, then more medium was added to a concentration of 2 mg/mL (i.e., a total volume of approx. 5 mL) (Step 2). The mixing procedures were repeated as necessary to dissolve the reference substance. If the reference substance did not dissolve, approximately 10 mg reference substance was added to approximately 0.5 mL DMSO in an attempt to dissolve it at a concentration of 200 mg/mL (Step 3). If the reference substance was not dissolved, the same concentration was attempted in 100% ETOH (Step 4). Step 5 began in the same way, with 0.2 mg/mL medium and then progressed to 20 mg/mL DMSO, and then 20 mg/mL ETOH.

Determination of reference substance solubility was limited to visual observation of the resulting solution. If a solution appeared clear, then solubility testing ceased. If particles were visible or if the solution appeared cloudy, then more stringent mixing and/or heating procedures were employed. If necessary, the solubility procedure proceeded to the next solvent/concentration tier. The duration of the solubility test was dependent on the procedures used to achieve solubility. Some reference substances were immediately solubilized (e.g., liquids) and others required up to 60 minutes of heating and agitation or sonication.

2.10.2 Data Collection

All laboratories (including the reference substance distribution laboratory, BioReliance) used a worksheet designed to capture the solubility information for each reference substance. The endpoint for each step was a visual observation of the solution, a documented comment describing the observation, the concentration, and a conclusion of soluble or insoluble. Each worksheet contained:

- Reference substance code number and physical description
- Solvent used (3T3 medium, NHK medium, DMSO, ETOH)
- Amount of reference substance (mg) used in the initial stage
- Volume of solvent added and final volume (mL)
- Test substance concentration (µg/mL) in the solvent
- pH and color of the solution
- Mechanical procedures used (vortexing, sonication, heating), duration, and temperature
- Comments (soluble/insoluble at the particular concentration; visual observations; reactivity with solvent)

The solubility test information and data from the laboratories were transferred via email to the SMT and stored on the NICEATM server and as hard-copy printouts. Each laboratory also maintained electronic and hard-copy files of its data.

2.10.3 Variability in Solubility Measurements

Solubility determinations were not replicated because within-laboratory results were not expected to vary. Comparison of the results to determine inter-laboratory concordance for the 72 reference substances (see **Section 5.8** for results) provided a measure of variability among the laboratories and information about the reproducibility of the solubility determinations (see **Section 7.4**).

2.10.4 Solubility Issues During the Testing of the Reference Substances

Substance solutions were monitored throughout all aspects of the test procedures, and observations were documented. The lowest concentration of the substance in a 2X solution that contained observable precipitates, particles, globules, or oily droplets, was documented in the EXCEL® template. After substance exposure, all wells of the 96-well test plates were observed microscopically and scored using a visual observation code. The code addressed growth characteristics and the presence or absence of precipitates (see **Appendix B** [test method protocols] for the observation codes used). For solubility issues, the Study Directors made determinations of test acceptance based on the recommended concentration levels and the presence of precipitates, their scientific expertise, and test acceptance criteria.

2.10.5 Analysis of Solubility Data

During Phase III, the SMT used the solubility data from all laboratories to determine the solvents to be used for each chemical (see **Section 5.8** for solubility results and SMT selections). If the solubility of an individual reference substance was different in 3T3 medium and NHK medium, the same solvent would be used for both test methods, rather than having different solvents for each method. For example, if solubility in one culture medium was ≥2 mg/mL and solubility in the other was <2 mg/mL, and the substance was soluble in DMSO at 200 mg/mL, the SMT would select DMSO as the solvent for both test methods (each test method using its respective culture medium).

Solubilizing sufficient reference substance to produce cytotoxicity was challenging for relatively insoluble, low toxicity, substances such as lithium carbonate (in the 3T3 NRU test method) but generally was not a problem for toxic substances that did not require as high a concentration to kill cells. Some insoluble and highly toxic reference substances were problematic, however, because the amount of powdered reference substance added to solvent was very small, and laboratory personnel found it difficult to determine the presence of solute particles in solution. Arsenic trioxide is an example of such a solute (see **Section 2.9.2**).

2.11 Summary

The *Guidance Document* NRU protocols were used as the basis of the validation study protocols. The SMT and participating laboratories made initial modifications to the protocols prior to implementation of the study. Other protocol modifications were made after commencement of testing and were the result of recommendations from the laboratories and the SMT, based on their experience with the initial protocols. The resulting optimized protocols were used in the main testing phase (Phase III) of the study.

The protocol components used in the validation study were similar for the 3T3 and NHK cells. The following procedures were common to the NRU protocols for both cell types:

- Testing was performed in four phases (Phases Ia, Ib, II, and III)
- Preparation of reference substances and positive control
- Cell culture environment conditions
- Determination of test substance solubility
- Configuration of 96-well plates for testing samples
- 48-hour exposure to test substance
- Range finder and definitive testing
- Microscopic evaluation of cell cultures for toxicity
- Measurement of NRU
- Data analysis

The main differences in the test methods for the two cell types were:

- The conditions of propagation of the cells in culture
- The cell growth medium components
- The volumes of reference substance added to the 96-well plate

A solubility protocol was developed which allowed the laboratories to identify the most appropriate solvent and appropriate limit concentrations for each test substance.

Three laboratories participated in testing the 72 reference substances in both cell types and one additional laboratory procured and distributed the coded reference substances and performed solubility tests prior to their distribution to the testing laboratories.