Pre-Publication Copy

Draft *In Vitro* Acute Toxicity Test Methods

Background Review Document (BRD)

30 October 2006

Do Not Cite, Quote, or Distribute

[This Page Intentionally Left Blank]

TABLE OF CONTENTS

Page Number

APPENDICES ..... **xiii**

LIST OF FIGURES ..... **xvii**

LIST OF TABLES ..... **xix**

LIST OF ACRONYMS AND ABBREVIATIONS ..... **xxiii**

ACKNOWLEDGEMENTS ..... **xxx**

PREFACE ..... **xxxiii**

EXECUTIVE SUMMARY ..... **xxxviii**

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 **1-3**

1.1 Historical Background and Rationale for the Use of *In Vitro* Cytotoxicity Assays to Predict Starting Doses for *In Vivo* Acute Oral Toxicity Tests **1-4**

1.1.1 The Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) Program 1-5

1.1.2 An International Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic Toxicity 1-7

1.1.3 The Registry of Cytotoxicity (RC) 1-9

1.1.4 The ZEBET Initiative to Reduce Animal Use 1-12

1.1.5 The International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity 1-13

1.1.6 The NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study 1-15

1.1.6.1 *Study Design* 1-16

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 **1-16**

1.2.1 Current Regulatory Testing Requirements for Acute Oral Toxicity 1-16

1.2.1.1 *Test Methods for Assessing Acute Oral Toxicity* 1-20

1.2.2 Intended Regulatory Uses for *In Vitro* Cytotoxicity Test Methods 1-20

1.2.3 Similarities and Differences in the Endpoints of *In Vitro* Cytotoxicity Test Methods and Rodent Acute Oral Toxicity Test Methods 1-21

1.2.4 Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard Assessment 1-23

1.3 Scientific Basis for the *In Vitro* NRU Test Methods 1-24
1.3.1 Purpose and Mechanistic Basis of *In Vitro* NRU Test Methods 1-24
1.3.2 Similarities and Differences in the Modes/Mechanisms of Action for *In Vitro* NRU Test Methods Compared with the Species of Interest 1-25
1.3.3 Range of Substances Amenable to the *In Vitro* NRU Test Methods 1-26

2.0 TEST METHOD PROTOCOL COMPONENTS OF THE 3T3 AND NHK *IN VITRO* NRU TEST METHODS 2-3

2.1 Basis for Selection of *In Vitro* NRU Cytotoxicity Test Methods 2-4
2.1.1 *Guidance Document* Rationale for Selection of *In Vitro* NRU Cytotoxicity Test Methods 2-5
2.1.2 *Guidance Document* Rationale for Selection of Cell Types 2-5

2.2 Overview of the 3T3 and NHK NRU Test Methods 2-6
2.2.1 The 3T3 NRU Test Method 2-9
2.2.1.1 *Initiating and Subculturing 3T3 Cells* 2-9
2.2.1.2 *Preparation of Cells for 96-well Plate Assays* 2-10
2.2.1.3 *Reference Substance Application* 2-10
2.2.2 The NHK NRU Test Method 2-10
2.2.2.1 *Initiating and Subculturing NHK Cells* 2-10
2.2.2.2 *Preparation of Cells for 96-well Plate Assays* 2-11
2.2.2.3 *Reference Substance Application* 2-11
2.2.3 Measurement of NRU in the 3T3 and NHK Test Methods 2-11

2.3 Descriptions and Rationales of the 3T3 and NHK NRU Test Methods 2-12
2.3.1 Materials, Equipment, and Supplies 2-12
2.3.1.1 *3T3 Cells* 2-12
2.3.1.2 *NHK Cells* 2-12
2.3.1.3 *Tissue Culture Materials and Supplies* 2-13
2.3.1.4 *Cell Culture Materials* 2-13
2.3.1.5 *Equipment* 2-14
2.3.1.6 *Culture Medium* 2-14
2.3.2 Reference Substance Concentrations/Dose Selection 2-16
2.3.2.1 *Range Finder Test* 2-16
2.3.2.2 *Definitive Test* 2-17
2.3.3 NRU Endpoints Measured 2-18
2.3.3.1 *NRU and Measurement* 2-18
2.3.3.2 *Determination of IC_{50}, IC_{20}, and IC_{80} Values* 2-18
2.3.4 Duration of Reference Substance Exposure 2-19
2.3.5 Known Limits of Use 2-19
2.3.5.1 *Solubility/Precipitation/Volatility* 2-19

2.3.5.2 *Biokinetic Determinations* 2-20
2.3.5.3 *Organ-Specific Toxicity* 2-20
2.3.5.4 *The Role of Cytotoxicity Tests in an In Vitro Battery Approach for Possible Replacement of In Vivo Acute Toxicity Testing* 2-21
2.3.6 Basis of the Response Assessed 2-22
2.3.7 Appropriate Vehicle, Positive, and Negative Controls 2-22
2.3.7.1 *Positive Control* 2-22
2.3.7.2 *Vehicle Control* 2-22
2.3.7.3 *Negative Control* 2-23
2.3.8 Acceptable Ranges of Control Responses 2-23
2.3.8.1 *Vehicle Controls as a Quality Control Tool* 2-24
2.3.9 Nature of Experimental Data Collected 2-24
2.3.9.1 *NRU OD Measurements* 2-25
2.3.9.2 *Information and Data Collected* 2-25
2.3.10 Data Storage Media 2-26
2.3.11 Measures of Variability 2-26
2.3.12 Methods for Analyzing NRU Data 2-27
2.3.13 Decision Criteria for Classification of Reference Substances 2-28
2.3.14 Information and Data Included in the Test Report 2-28

2.4 Proprietary Components of the *In Vitro* NRU Test Methods 2-30

2.5 Basis for the Number of Replicate and Repeat Experiments for the 3T3 and NHK NRU Test Methods 2-31

2.6 Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols 2-31
2.6.1 Phase Ia: Laboratory Evaluation Phase 2-31
2.6.1.1 *NR Dye Crystals* 2-31
2.6.1.2 *3T3 Cell Growth* 2-32
2.6.1.3 *NHK Cell Growth* 2-32
2.6.1.4 *Vehicle Control OD Limits* 2-33
2.6.1.5 *Precipitate Formation* 2-34
2.6.1.6 *Dilution Factor* 2-34
2.6.1.7 *Test Acceptance Criteria* 2-35
2.6.2 Phase Ib: Laboratory Evaluation Phase 2-35
2.6.2.1 *NR Crystal Formation* 2-35
2.6.2.2 *Heating of Reference Substance Solutions* 2-36
2.6.2.3 *Growth of Untreated Cells* 2-37
2.6.2.4 *Correction of Reference Substance OD Values* 2-38
2.6.2.5 *Laboratory Error Rates* 2-38
2.6.2.6 *Resultant Protocol Changes for Phase II* 2-39
2.6.2.7 *Test Acceptance Criteria* 2-40
2.6.3 Phase II: Laboratory Qualification Phase 2-40
2.6.3.1 *Testing of Volatile Reference Substances* 2-40
2.6.3.2 *Atypical Concentration-Responses* 2-43
2.6.3.3 *Hill Function* 2-45

2.6.3.4 *Insoluble Reference Substances* 2-46
2.6.3.5 *Inadequate Cell Growth in NHK Medium* 2-46
2.6.3.6 *Performance Standards for Media to Support NHK Growth* 2-47
2.6.3.7 *Test Acceptance Criteria for Phase II* 2-48
2.6.4 Phase III: Laboratory Testing Phase 2-48
2.6.4.1 *Required Cytotoxicity Values* 2-48
2.6.4.2 *Revisions to Data Analysis Procedures* 2-49

2.7 Differences Between the 3T3 and NHK NRU Protocols for the Validation Study and the *Guidance Document* Standard Protocols 2-49

2.8 Overview of the Solubility Protocol 2-51

2.9 Basis of the Solubility Protocol 2-52
2.9.1 Initial Solubility Protocol Development 2-53
2.9.2 Basis for Modification of the Phase II Protocol 2-54

2.10 Components of the Solubility Protocol 2-55
2.10.1 Medium, Supplies, and Equipment Required 2-55
2.10.1.1 *Medium and Chemical Supplies* 2-55
2.10.1.2 *Equipment* 2-55
2.10.1.3 *Procedures* 2-55
2.10.2 Data Collection 2-56
2.10.3 Variability in Solubility Measurement 2-57
2.10.4 Solubility Issues During the Testing of the Reference Substances ..2-57
2.10.5 Analysis of Solubility Data 2-57

2.11 Summary 2-58

3.0 REFERENCE SUBSTANCES USED FOR VALIDATION OF THE 3T3 AND NHK NRU TEST METHODS 3-3

3.1 Rationale for the 72 Reference Substances Selected for Testing 3-3
3.1.1 Reference Substance Selection Criteria 3-3
3.1.2 Candidate Reference Substances 3-5
3.1.3 Selection of Reference Substances for Testing 3-6

3.2 Characteristics of the Selected Reference Substances 3-7
3.2.1 Source Databases Represented by the Selected Reference Substances 3-8
3.2.2 Chemical Classes Represented by the Selected Reference Substances 3-31
3.2.3 Product/Use Classes Represented by the Selected Reference Substances 3-32
3.2.4 Toxicological Characteristics of the Selected Reference Substances 3-32
3.2.4.1 Corrosivity 3-32
3.2.4.2 Toxicity Targets 3-33

3.2.4.3 Metabolism 3-37
3.2.5 Selection of Reference Substances for Testing in Phases Ib and II 3-37
3.2.6 Unsuitable and Challenging Reference Substances 3-38

3.3 Reference Substance Procurement, Coding, and Distribution....3-40
3.3.1 Exceptions....3-41

3.4 Reference Substances Recommended by the *Guidance Document*....3-42

3.5 Summary 3-42

4.0 RODENT ACUTE ORAL LD_{50} REFERENCE VALUES USED TO ASSESS THE ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS 4-3

4.1 Methods Used to Obtain Rodent Acute Oral LD_{50} Reference Values....4-3
4.1.1 Identification of Candidate Rodent Acute Oral LD_{50} Reference Data 4-3
4.1.2 Criteria Used to Select Candidate Rodent Acute Oral Data for Determination of LD_{50} Reference Values....4-6
4.1.2.1 Final Exclusion Criteria....4-6
4.1.2.2 Assumptions Regarding Materials, Animals, and Methods 4-7
4.1.2.3 Calculation Reference LD_{50} Values....4-8
4.1.2.4 Use of Rat and Mouse Data....4-8

4.2 Final Rodent Acute Oral LD_{50} Reference Values 4-9

4.3 Relevant Toxicity Information for Humans....4-14

4.4 Accuracy and Reliability of the Rodent Acute Oral LD_{50} Reference Values 4-15
4.4.1 Variability Among the Acceptable LD_{50} Values....4-15
4.4.2 Comparison of Rodent Acute Oral LD_{50} Reference Values with the Corresponding RC LD_{50} Values....4-16
4.4.3 Comparison of the Variability Among Acceptable LD_{50} Values to Those Obtained in Other Studies....4-18

4.5 Summary 4-19

5.0 3T3 AND NHK NRU TEST METHOD DATA AND RESULTS 5-3

5.1 Study Timeline and Participating Laboratories 5-3
5.1.1 Statements of Work (SOW) and Protocols....5-3
5.1.2 Study Timeline....5-4
5.1.3 Participating Laboratories....5-4

5.2 Coded Reference Substances and GLP Guidelines....5-6
5.2.1 Coded Reference Substances 5-6
5.2.2 Lot-to-Lot Consistency of Reference Substances 5-6

5.2.3 Adherence to GLP Guidelines 5-6

5.3 3T3 and NHK NRU Test Method Protocols.....5-7
5.3.1 Phase Ia: Laboratory Evaluation Phase 5-7
5.3.1.1 *Protocol Changes and the Effect on the Data* 5-7
5.3.2 Phase Ib: Laboratory Evaluation Phase 5-8
5.3.2.1 *Protocol Changes and the Effect on the Data* 5-9
5.3.3 Phase II: Laboratory Qualification Phase 5-9
5.3.3.1 *Protocol Changes and the Effect on the Data* 5-10
5.3.4 Phase III: Main Validation Phase.....5-11
5.3.4.1 *Protocol Changes and the Effect on the Data* 5-12

5.4 Data Used to Evaluate Test Method Accuracy and Reliability 5-14
5.4.1 PC Data 5-14
5.4.1.1 *Phase Ib PC Data Acceptance Limits* 5-15
5.4.1.2 *Phase II PC Data Acceptance Limits*.....5-15
5.4.1.3 *Phase III PC Data Acceptance Limits* 5-16
5.4.2 Reference Substance Data.....5-17

5.5 Statistical Approaches to the Evaluation of 3T3 and NHK Data 5-31
5.5.1 Statistical Analyses for Phase Ia 5-31
5.5.1.1 *Outlier Determination for Replicate Well Concentration Data*.....5-31
5.5.1.2 *Curve Fit Criteria* 5-32
5.5.1.3 *Reproducibility Analyses for PC IC_{50} Values*.....5-32
5.5.2 Statistical Analyses for Phase Ib 5-33
5.5.2.1 *Outlier Determination for Replicate Well Concentration Data*.....5-33
5.5.2.2 *Reproducibility Analyses of the Reference Substance IC_{50} Values*.....5-33
5.5.3 Statistical Analyses for Phase II.....5-34
5.5.3.1 *Outlier Determination for Replicate Well Concentration Data*.....5-34
5.5.3.2 *Reproducibility Analyses of Reference Substance IC_{50} Values* 5-34
5.5.3.3 *Comparison of 3T3 and NHK Test Results to the RC Millimole Regression* 5-34
5.5.4 Statistical Analyses for Phase III.....5-35
5.5.4.1 *Outlier Determination for Replicate Well Concentration Data*.....5-35
5.5.4.2 *Reproducibility Analyses of the PC IC_{50} Data* 5-35
5.5.4.3 *Reproducibility Analyses of the Reference Substance IC_{50} Values*.....5-36
5.5.4.4 *Comparison of 3T3 and NHK Test Results to the RC Millimole Regression* 5-37
5.5.5 Summary of the Data Used for Statistical Analyses 5-39

5.6 Summary of NRU Test Results.....5-39

5.7 Availability of Data 5-46

5.8 Solubility Test Results.....5-47
5.8.1 Solubility Data.....5-47
5.8.2 Solubility and Volatility Effects in the Cytotoxicity Tests.....5-52

5.9 Summary 5-53

6.0 ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS 6-5

6.1 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Rodent Acute Oral Toxicity.....6-7
6.1.1 Linear Regression Analyses for the Prediction of Rat Acute Oral LD_{50} Values from *In Vitro* IC_{50} Values 6-7
6.1.2 Comparison of Combined-Laboratory 3T3 and NHK Regressions to the RC Millimole Regression.....6-11

6.2 Analysis of Outlier Substances for the RC Millimole Regression 6-11
6.2.1 Identification of Outlier Substances 6-13
6.2.2 Evaluation of Outlier Substances 6-15
6.2.2.1 *Physical Characteristics*.....6-16
6.2.2.2 *Chemical Class* 6-16
6.2.2.3 *Solubility*.....6-16
6.2.2.4 *Metabolism* 6-17
6.2.2.5 *Mechanism of Toxicity* 6-18

6.3 Improving the Prediction of *In Vivo* Rat Oral LD_{50} Values from *In Vitro* IC_{50} Data 6-21
6.3.1 The RC Rat-Only Millimole Regression 6-21
6.3.2 The RC Rat-Only Weight Regression 6-22

6.4 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting GHS Acute Oral Toxicity Category.....6-24
6.4.1 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Millimole Regression.....6-27
6.4.1.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Millimole Regression* 6-27
6.4.1.2 *Discordant Substances in the Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression*.....6-30
6.4.2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression.....6-30
6.4.2.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Weight Regression*.....6-32
6.4.2.2 *Discordant Substances from the Prediction of Acute Oral Toxicity Category with the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression*.....6-33

6.4.3 Summary of the Regressions Evaluated 6-33

6.5 Correlation of NRU Concentration-Response Slope with Rat Lethality Dose-Response Slope 6-35

6.6 Strengths and Limitations of the *In Vitro* Cytotoxicity Test Methods with the IC_{50}–LD_{50} Regressions for Prediction of Rodent Acute Oral Toxicity 6-36

6.6.1 *In Vitro* Cytotoxicity Methods 6-36

6.6.2 Use of Mole-Based vs. Weight-Based Regressions for the Prediction of Toxicity for Low and High Molecular Weight Substances 6-39

6.7 Salient Issues of Data Interpretation 6-42

6.8 Comparison of NRU Test Results to Established Performance Standards 6-42

6.9 Summary 6-49

7.0 RELIABILITY OF THE 3T3 AND NHK TEST METHODS 7-3

7.1 Reference Substances Used to Determine the Reliability of the 3T3 and NHK Test Methods 7-3

7.2 Reproducibility Analyses for the 3T3 and NHK Test Methods.....7-5

7.2.1 Comparison of Laboratory-Specific IC_{50}-LD_{50} Linear Regression Analyses to the Mean Laboratory Regression 7-6

7.2.2 ANOVA Results for the 3T3 and NHK NRU Test Methods.....7-6

7.2.2.1 Differences Among IC_{50} Values in Laboratories Using the 3T3 NRU Test Method 7-6

7.2.2.2 Differences Among IC_{50} Values in Laboratories using the NHK Test Method 7-22

7.2.3 CV Results for the 3T3 and NHK NRU Test Methods 7-22

7.2.3.1 Reproducibility of Intralaboratory CV Values 7-22

7.2.3.2 Reproducibility of Interlaboratory CV Values 7-35

7.2.3.3 Variation of CV with Chemical Property.....7-36

7.2.3.4 Results of the Intralaboratory CV Analysis.....7-36

7.2.3.5 Results of the Interlaboratory CV Analysis.....7-38

7.2.4 Comparison of Maximum to Minimum IC_{50} Values Using Laboratory Means.....7-41

7.2.5 Comparison of Maximum:Minimum IC_{50} Ratios with the Maximum:Minimum LD_{50} Ratios 7-43

7.2.6 Normalization of Reference Substance IC_{50} Values Using SLS IC_{50} Values.....7-44

7.3 Historical Positive Control Data.....7-48

7.3.1 ANOVA and Linear Regression Results for the 3T3 NRU Test Method 7-50

7.3.1.1 Variation of the SLS IC_{50} Values with Time.....7-50

7.3.1.2 Comparison of SLS IC_{50} Values Among the

Laboratories 7-50
7.3.2 ANOVA and Linear Regression Results for the NHK NRU Test Method 7-52
7.3.2.1 Variation SLS IC_{50} Values with Time 7-52
7.3.2.2 Comparison of SLS IC_{50} Values Among the Laboratories 7-52

7.4 Laboratory Concordance for Solvent Selection 7-54

7.5 Summary 7-57

8.0 3T3 AND NHK NRU TEST METHOD DATA QUALITY 8-3

8.1 Compliance with Good Laboratory Practice Regulations 8-3
8.1.1 Guidelines Followed for Cytotoxicity Testing 8-3
8.1.1.1 *Good Laboratory Practices* 8-3
8.1.1.2 *Spirit of GLP* 8-3
8.1.1.3 *Good Cell Culture Practices (GCCP)* 8-4
8.1.2 Quality Assurance (QA) for NRU Cytotoxicity Test Data 8-5
8.1.2.1 *Coded Reference Substances* 8-5
8.1.2.2 *Solubility Testing and Data Review* 8-5
8.1.2.3 *NRU Cytotoxicity Test Tallies* 8-6
8.1.3 Guidelines Followed for Rodent Acute Oral LD_{50} Data Collection 8-7
8.1.3.1 *Rodent Acute Oral LD_{50} Values Used in the Registry of Cytotoxicity (RC)* 8-7
8.1.3.2 *Rodent Acute Oral LD_{50} Values Collected by NICEATM from Other Sources* 8-7

8.2 Results of Data Quality Audits 8-8
8.2.1 QA Statements 8-8
8.2.2 QA Statements from the Laboratories 8-8
8.2.2.1 *BioReliance QA Statements* 8-8
8.2.2.2 *FAL QA Statements* 8-9
8.2.2.3 *ECBC QA Statements* 8-9
8.2.2.4 *IIVS QA Statements* 8-10
8.2.2.5 *Other QA Information* 8-11

8.3 Effect of Deviations or Non-compliance with GLPs 8-11
8.3.1 Laboratory Error Rates 8-12
8.3.2 Failure Rates for Definitive and PC Tests 8-13
8.3.3 Intralaboratory Reproducibility 8-15
8.3.4 Prediction of Globally Harmonized System (GHS) Acute Oral Toxicity Categories 8-15

8.4 Availability of Laboratory Notebooks 8-16

8.5 Summary 8-17

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS OF *IN VITRO* CYTOTOXICITY TEST METHODS AND THEIR ABILITY TO PREDICT IN VIVO ACUTE TOXICITY AND OTHER TOXIC EFFECTS 9-3

9.1 Relevant Studies 9-3
9.1.1 Correlation of NRU Cytotoxicity Values with Rodent Lethality 9-3
9.1.1.1 Peloux et al. (1992) 9-5
9.1.1.2 Fautrel et al. (1993) 9-6
9.1.1.3 Roguet et al. (1993) 9-6
9.1.1.4 Rasmussen (1999) 9-7
9.1.1.5 Creppy et al. (2004) 9-8
9.1.2 Use of *In Vitro* Cytotoxicity Data to Reduce the Use of Animals in Acute Oral Toxicity Testing 9-9
9.1.2.1 Halle et al. (1997): Animal Savings with the ATC Method Using Cytotoxicity Data 9-9
9.1.2.2 Spielmann et al. (1999): Animal Savings Using Cytotoxicity Data with the UDP Method 9-11
9.1.2.3 EPA (2004): U.S. EPA HPV Challenge Program Submission 9-13
9.1.3 Other Evaluations of 3T3 or NHK NRU Methods 9-14
9.1.3.1 Draize Eye Irritation 9-14
9.1.3.2 Predicting Human Lethal Blood Concentrations 9-17

9.2 Independent Scientific Reviews 9-17
9.2.1 *In Vitro* Acute Toxicity Testing for the Classification and Labeling of Chemicals 9-17
9.2.1.1 Seibert et al. (1996): ECVAM Workshop 16 9-17
9.2.2 Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing 9-19
9.2.2.1 ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the ATC Method 9-19
9.2.2.2 ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the UDP Method 9-21
9.2.3 Validation of 3T3 NRU Assay for Phototoxicity 9-23
9.2.3.1 NHK NRU Phototoxicity Assay 9-24

9.3 Studies Using *In Vitro* Cytotoxicity Test Methods with Established Performance Standards 9-25
9.3.1 *Guidance Document* (ICCVAM 2001b) 9-25
9.3.2 King and Jones (2003) 9-25
9.3.3 A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005) 9-26

9.4 Summary 9-27

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT) 10-3

10.1 Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Rodent Acute Oral Toxicity Assays 10-4

10.2 Reduction and Refinement of Animal Use for the UDP 10-5
10.2.1 *In Vivo* Testing Using the UDP 10-5
10.2.1.1 *Main Test* 10-5
10.2.1.2 *Limit Test* 10-6
10.2.2 Computer Simulation Modeling of the UDP 10-7
10.2.3 Animal Savings in the UDP When Using 3T3- and NHK-Based Starting Doses 10-9
10.2.3.1 *The Effect of the Dose-Mortality Slope on Animal Use* 10-9
10.2.3.2 *Mean Animal Use for UDP Simulations – Comparison of Regressions and Predictions for the 3T3 and NHK NRU Test Methods* 10-10
10.2.3.3 *Animal Savings in the UDP by Toxicity Category Using 3T3- and NHK-Based Starting Doses* 10-12
10.2.4 Refinement of Animal Use for the UDP When Using 3T3- and NHK-Based Starting Doses 10-19
10.2.5 Accuracy of the UDP Outcomes Using the IC_{50}-Based Starting Doses 10-19

10.3 Reduction and Refinement of Animal Use in the ATC Method 10-21
10.3.1 *In Vivo* Testing Using the ATC Method 10-21
10.3.1.1 *Main Test* 10-21
10.3.1.2 *Limit Test* 10-22
10.3.2 Computer Simulation Modeling of the ATC Method 10-22
10.3.3 Animal Savings for the ATC Method When Using 3T3- and NHK-Based Starting Doses 10-24
10.3.3.1 *The Effect of Dose-Mortality Slope on Animal Use* 10-24
10.3.3.2 *Mean Animal Use for ATC Simulations – Comparison of Regressions and Predictions from the 3T3 and NHK NRU Test Methods* 10-25
10.3.3.3 *Animal Savings in the ATC Method by GHS Acute Oral Toxicity Category Using* the *3T3- and NHK-Based Starting Doses* 10-28
10.3.4 Refinement of Animal Use in the ATC Method When Using 3T3- and NHK-Based Starting Doses 10-34
10.3.5 Accuracy of the ATC Method Outcomes Using IC_{50}-Based Starting Doses 10-34

10.4 The Impact of Accuracy on Animals Savings 10-36

10.5 The Impact of Prevalence on Animals Savings 10-39

10.6 Summary 10-39

11.0 PRACTICAL CONSIDERATIONS 11-3

11.1 Transferability of the 3T3 and NHK NRU Test Methods 11-3
11.1.1 Facilities and Major Fixed Equipment 11-4
11.1.1.1 *Facility Requirements* 11-4
11.1.1.2 *Cell Culture Laboratory* 11-4
11.1.1.3 *Major Equipment* 11-4
11.1.2 Availability of Other Necessary Equipment and Supplies 11-5
11.1.2.1 *General Equipment* 11-5
11.1.2.2 *Cell Culture Materials and Supplies* 11-6
11.1.2.3 *Cell Cultures* 11-6
11.1.3 Problems Specific to the NHK NRU Test Method 11-7

11.2 3T3 and NHK Test Method Training Considerations 11-7
11.2.1 Required Training and Expertise 11-8
11.2.1.1 *Specific Training and Expertise Needed* 11-8
11.2.1.2 *General Laboratory Expertise Needed* 11-9
11.2.2 Training Requirements to Demonstrate Proficiency 11-9
11.2.2.1 *Proficiency with GLP-Compliance* 11-10
11.2.3 Personnel Needed to Perform the *In Vitro* NRU Cytotoxicity Test Methods 11-10

11.3 Cost Considerations 11-11
11.3.1 3T3 and NHK NRU Test Methods 11-11
11.3.1.1 *Equipment Costs* 11-11
11.3.1.2 *Costs for Cell Cultures and Supplies* 11-12
11.3.1.3 *Commercial Testing* 11-12
11.3.2 Rodent Acute Oral Toxicity Testing 11-12

11.4 Time Considerations for Performing the 3T3 and NHK NRU Tests .. 11-15
11.4.1 The 3T3 NRU Test Method 11-15
11.4.2 The NHK NRU Test Method 11-15
11.4.3 Prequalification of NHK Medium 11-15
11.4.4 *In Vivo* Testing 11-16
11.4.5 Using NRU Cytotoxicity Methods to Determine Need to Perform the Limit Test 11-16

11.5 Summary 11-17

12.0 REFERENCES 12-1

13.0 GLOSSARY 13-1

APPENDICES

Appendix A NICEATM/ECVAM Validation Study Management A-1

Appendix B Validation Study Test Method Protocols (Phase III) B-1

B-1 Test Method Protocol for the BALB/c 3T3 Neutral Red Uptake (NRU) Cytotoxicity Test B-3

B-2 Test Method Protocol for the Normal Human Epidermal Keratinocyte (NHK) Neutral Red Uptake Cytotoxicity Test B-25

B-3 Test Method Protocol for Solubility Determination (Phase III) B-47

B-4 Test Method Procedure for Prequalification of Normal Human Epidermal Keratinocyte Growth Medium (Phase III) B-59

Appendix C Validation Study Test Method Protocols (Phases Ia, Ib, and II) C-1

C-1 Test Method Protocol for the BALB/c 3T3 Neutral Red Uptake Cytotoxicity Test (Phase Ia) C-3

C-2 Test Method Protocol for the Normal Human Epidermal Keratinocyte (NHK) Neutral Red Uptake Cytotoxicity Test (Phase Ia) C-23

C-3 Test Method Protocol for the BALB/c 3T3 Neutral Red Uptake Cytotoxicity Test (Phase Ib) C-41

C-4 Test Method Protocol for the Normal Human Epidermal Keratinocyte (NHK) Neutral Red Uptake Cytotoxicity Test (Phase Ib) C-63

C-5 Test Method Protocol for the BALB/c 3T3 Neutral Red Uptake Cytotoxicity Test (Phase II) C-85

C-6 Test Method Protocol for the Normal Human Epidermal Keratinocyte (NHK) Neutral Red Uptake Cytotoxicity Test (Phase II) C-109

Appendix D Relevant Publications D-1

D-1 Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. Available at the following URL: http://iccvam.niehs.nih.gov/methods/invidocs/guidance/iv_guide.htm D-3

D-2 Spielmann H, Genschow E, Liebsch M, Halle W. 1999. Determination of the starting dose for acute oral toxicity (LD_{50}) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Laboratory Anim 27:957-966. ..... D-5

Appendix E Neutral Red Dye Experiments E-1

E-1 Assessment of Protocol Variables in the NICEATM/ECVAM Evaluation of Cytotoxicity Assays E-5

E-2 Neutral Red (NR) Dye Experiments – 3T3 Cells – IIVS E-13

E-3 Neutral Red (NR) Dye Experiments – NHK Cells – IIVS E-19

E-4 Neutral Red (NR) Dye Experiments – 3T3 Cells – ECBC E-25

Appendix F Reference Substance Information F-1

F-1 NRU Test Information for the 72 Reference Substances F-3

F-2 Chemical, Physical, and Biological Information from the Literature for the 72 Reference Substances F-11

F-3 Candidate Reference Substances F-27

Appendix G Statement of Work (SOW) G-1

G-1 A Validation Study For *In Vitro* Basal Cytotoxicity Testing G-3

G-2 Procedures for Acquisition, Preparation, Solubility Testing, and Distribution of Test Chemicals for a Validation Study for *In Vitro* Basal Cytotoxicity Testing G-57

Appendix H Rat and Mouse Oral LD_{50} Database H-1

H-1 Rat and Mouse Oral LD_{50} Database H-3

H-2 Evaluation of the Candidate Reference Data H-39

Appendix I *In Vitro* NRU Data I-1

I-1 3T3 NRU Test Chemical Data I-3

I-2 NHK NRU Test Chemical Data I-79

I-3 3T3 NRU Positive Control (SLS) Data I-155

I-4 NHK NRU Positive Control (SLS) Data I-173

Appendix J LD_{50} and Toxicity Category Predictions J-1

J-1 3T3 NRU Predictions: RC Millimole Regression J-5

J-2 NHK NRU Predictions: RC Millimole Regression J-11

J-3 3T3 NRU Predictions: RC Rat-Only Millimole Regression J-17

J-4 NHK NRU Predictions: RC Rat-Only Millimole Regression J-23

J-5 3T3 NRU Predictions: RC Rat-Only Weight Regression J-29

J-6 NHK NRU Predictions: RC Rat-Only Weight Regression J-35

J-7 Comparison of Millimole Regression with Weight Regression Regarding Prediction of Toxicity (LD_{50}) for Low or High Molecular Weight Chemicals J-41

Appendix K NRU Data for Regressions K-1

K-1 NRU Data Used for Laboratory-Specific Regressions K-3

K-2 Geometric Mean of Laboratory NRU Data Used for Combined-Laboratory Regressions K-17

K-3 Data for RC Chemicals with Rat Oral LD_{50} Data K-25

Appendix L Outlier Information L-1

L-1 Outlier Characterization for the 3T3 and NHK NRU Test Methods with the RC Millimole Regression L-3

L-2 Discordant Substances for GHS Toxicity Category Predictions Using the 3T3 and NHK NRU Test Methods and RC Rat-Only Regressions L-9

L-3 Analysis of Outliers by Halle (1998) for the RC Millimole Regression..... L-17

Appendix M Acute Systemic Toxicity Test Guidelines M-1

M-1 OECD UDP Test Guideline..... M-3

M-2 EPA UDP Test Guideline..... M-31

M-3 OECD ATC Test Guideline..... M-71

M-4 OECD FDP Test Guideline M-87

Appendix N UDP/ATC Simulation Modeling Results **..... N-1**

N-1 UDP Results for the RC Millimole Regression –Starting One Dose Lower Than Estimated LD_{50} - 5000 mg/kg Upper Limit..... N-3

N-2 UDP Results for the RC Rat-Only Weight Regression – Starting One Dose Lower Than Estimated LD_{50} – 5000 mg/kg Upper Limit N-11

N-3 ATC Results for the RC Millimole Regression - 2000 mg/kg Upper Limit..... N-19

N-4 ATC Results for the RC Rat-Only Weight Regression – 2000 mg/kg Upper Limit N-27

Appendix O - *Federal Register* Notices O-1

O-1 **70FR14473** - Request for Nominations for an Independent Peer Review Panel To Evaluate *In Vitro* Testing Methods for Estimating Acute Oral Systemic Toxicity and Request for *In Vivo* and In Vitro Data..... O-3

O-2 **69FR61504** - Availability of Updated Standardized *In Vitro* Cytotoxicity Test Method Protocols for Estimating Acute Oral Systemic Toxicity; Request for Existing *In Vivo* and *In Vitro* Acute Toxicity Data..... O-5

O-3 **69FR11448 -** Notice of the Availability of Agency Responses to ICCVAM Test Recommendations for the Revised Up-and-Down Procedure for Determining Acute Oral Toxicity and *In Vitro* Methods for Assessing Acute Systemic Toxicity O-7

O-4 **66FR49686 -** Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity; Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity: Notice of Availability and Request for Public Comment..... O-9

O-5 **65FR57203 -** Notice of an International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, co-sponsored by NIEHS, NTP and the

U.S. Environmental Protection Agency (EPA): Workshop Agenda and Registration Information O-11

O-6 **65FR37400 -** Notice of an International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, co-sponsored by NIEHS, NTP and the U.S. Environmental Protection Agency (EPA): Request for Data and Suggested Expert Scientists O-14

Appendix P *In Vitro* Cytotoxicity Test Methods and the High Production Volume (HPV) Challenge Program P-1

P-1 Supplemental Acute Toxicity Protocol P-3

P-2 Office of Pollution Prevention and Toxics (OPPT) Letters to Manufacturers/Importers P-9

Appendix Q Additional UDP Simulation Modeling Results Q-1

Q-1 UDP Results for the RC Millimole Regression – Starting at Estimated LD_{50} - 5000 mg/kg Upper Limit Q-3

Q-2 UDP Results for the RC Rat-Only Weight Regression– Starting at Estimated LD_{50} - 5000 mg/kg Upper Limit Q-11

Appendix R SAS Codes R-1

R-1 SAS Code for ANOVA and Contrasts R-3

R-2 SAS Code for Regression Comparisons R-7

LIST OF FIGURES

Figure 1-1 RC Millimole Regression Between *In Vitro* Cytotoxicity (IC_{50x}) and Rat and Mouse Acute Oral LD_{50} Values for 347 Chemicals 1-11

Figure 1-2 NICEATM/ECVAM Validation Study Phases 1-17

Figure 2-1 Major Steps in the Performance of the NRU Test Methods 2-8

Figure 2-2 3T3 NRU OD for SLS as a Function of NR Concentration and Duration ... 2-37

Figure 2-3 SLS IC_{50} Values for Each NR Concentration and Incubation Duration (3T3 NRU) 2-38

Figure 2-4 Representative Concentration-Response for 2-Propanol in a 3T3 NRU Range Finder Test 2-42

Figure 2-5 Representative Concentration-Response for Aminopterin in a NHK NRU Range Finder Test 2-43

Figure 2-6 Representative Concentration-Response for Aminopterin in a NHK NRU Definitive Test 2-44

Figure 2-7 Flow Chart for Determination of Reference Substance Solubility in Medium, DMSO, or ETOH 2-53

Figure 3-1 The Fifty-Eight (58) Selected RC Reference Substances on the RC Regression 3-31

Figure 4-1 Correlation of LD_{50} Values for the 58 RC Chemicals With the Reference LD_{50} Values 4-17

Figure 5-1 Reference Substance IC_{50} Results and Data for the 3T3 NRU Test Method by Laboratory 5-28

Figure 5-2 Reference Substance IC_{50} Results and Data for the NHK NRU Test Method by Laboratory 5-29

Figure 5-3 RC IC_{50} Values vs 3T3 NRU IC_{50} Values for the 56 Substances in Common 5-45

Figure 5-4 RC IC_{50} Values vs NHK NRU IC_{50} Values for the 57 Substances in Common 5-46

Figure 6-1 Combined-Laboratory 3T3 and NHK NRU Regressions 6-7

Figure 6-2 Regression for 47 Substances Using RC Data 6-8

Figure 6-3 Regression for 47 Substances with the 3T3 and NHK Regressions 6-9

Figure 6-4 RC Rat-Only Millimole Regression (a) and RC Rat-Only Weight Regression (b) 6-11

Figure 6-5 Correlation of Dose-Mortality Slope to Hill Function Slope 6-13

Figure 6-6 *In Vitro – In Vivo* Regressions for Phases Ib and II for ECBC 6-38

Figure 6-7 *In Vitro – In Vivo* Regressions for Phases Ib and II for FAL 6-39

Figure 6-8 *In Vitro – In Vivo* Regressions for Phases Ib and II for IIVS.....6-40

Figure 7-1 Mean Laboratory and Laboratory-Specific 3T3 and NHK NRU Regressions 7-8

Figure 7-2 Frequency of Maximum:Minimum 3T3 NRU IC_{50} Ratios 7-42

Figure 7-3 Frequency of Maximum:Minimum NHK NRU IC_{50} Ratios 7-43

Figure 7-4 Comparison of Maximum:Minimum NRU IC_{50} Ratios to Maximum: Minimum LD_{50} Ratios.....7-45

Figure 7-5 SLS IC_{50} for Each Laboratory and Study Phase 7-49

LIST OF TABLES

Table ES-1 Datasets Used for Validation Study Analyses xx
Table 1-1 Summary of Current U.S. Legislation for Using Acute Toxicity Data for Product Labeling 1-18
Table 1-2 Regulatory Classification Systems for Acute Oral Toxicity 1-19
Table 2-1 Measured VC OD_{540} Values and Targets 2-23
Table 2-2 Refeeding/No Refeeding Data for the NHK NRU Test Method 2-33
Table 2-3 Error Rates in Phase Ib by Laboratory and Test 2-39
Table 2-4 Cell Seeding Densities 2-51
Table 2-5 Comparison of Concentrations Tested in Various Solubility Protocols 2-54
Table 3-1 GHS Classification Scheme for Acute Oral Toxicity 3-4
Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity 3-9
Table 3-3 Distribution of Candidate Substances and Reference Substances by Source and Toxicity Category 3-30
Table 3-4 Selected Substances: Distribution of RC Chemicals and RC Outliers by Toxicity Category 3-30
Table 3-5 Distribution of Chemical Class for the 72 Reference Substances by Toxicity Category 3-34
Table 3-6 Distribution of Product/Use Class for the 72 Reference Substances by Toxicity Category 3-36
Table 3-7 Reference Substances Metabolized to Active Metabolites 3-37
Table 3-8 Reference Substances Tested in Phases Ib and II 3-39
Table 4-1 Internet-Accessible Databases Searched for LD_{50} Information 4-4
Table 4-2 Rodent Acute Oral Reference LD_{50} Values Listed by GHS Category 4-10
Table 4-3 GHS Category Matches for the Rodent Acute Oral LD_{50} Initial and Reference Values 4-13
Table 4-4 Maximum:Minimum LD_{50} Ratios by GHS Toxicity Category 4-16
Table 5-1 Validation Study Timetable 5-5
Table 5-2 Reference Substances Affected by the Stopping Rule 5-13
Table 5-3 Positive Control (PC) Data and IC_{50} Results by Study Phase 5-15
Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory 5-18
Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory 5-23

Table 5-6 Key to Validation Study Reference Substances 5-30

Table 5-7 Datasets Used for Validation Study Analyses.....5-39

Table 5-8 Comparison of 3T3 and NHK NRU IC_{50} Geometric Means.....5-41

Table 5-9 Frequency of 3T3:NHK IC_{50} Ratios for Reference Substances 5-44

Table 5-10 Solubility Test Results (mg/mL).....5-48

Table 5-11 Reference Substances with Precipitate (PPT) and Volatility Issues).....5-52

Table 6-1 Linear Regression Analyses of the 3T3 and NHK and Rat Acute Oral LD_{50} Test Results.....6-6

Table 6-2 Outlier Substances for the RC and the 3T3 and NHK Methods When the RC Millimole Regression is Used.....6-10

Table 6-3 Substances With Mechanisms of Toxicity Not Expected to Be Active in the 3T3 or NHK Cells in Culture.....6-16

Table 6-4 Linear Regression Analyses to Improve the Prediction of Rodent Acute Oral LD_{50} Values from *In Vitro* NRU IC_{50} Using the RC Regression.....6-16

Table 6-5 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression 6-23

Table 6-6 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression 6-27

Table 6-7 Comparison of Regressions and *In Vitro* NRU Test Methods for Their Performance in Predicting GHS Acute Oral Toxicity Categories 6-31

Table 6-8 Reference Substances with Dose-Mortality and NRU Hill Slopes.....6-37

Table 6-9 Over- and Under- Prediction of Toxicity for Low and High Molecular Weight Substances Using RC Rat-Only Weight and Millimole Regressions.....6-37

Table 6-10 Linear Regressions for 11 Substances Tested in Phases Ib and II.....6-37

Table 7-1 Reference Substances Excluded from Reproducibility Analyses Because of Insufficient Cytotoxicity 7-4

Table 7-2 Number of Reference Substances Tested vs Number of Reference Substances Yielding IC_{50} Values from Each Laboratory, by GHS Acute Oral Toxicity Category.....7-5

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method.....7-9

Table 7-4 Reference Substances with Significant (ANOVA) Differences Among Laboratories Using the 3T3 NRU Test Method 7-21

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method.....7-23

Table 7-6 Reference Substances with Significant ANOVA Differences Among Laboratories for NHK NRU Test Method.....7-35

Table 7-7 Summary of CV Results for the 3T3 and NHK NRU Test Methods.....7-35

Table 7-8 Intralaboratory CV by Chemical Characteristics for the 3T3 and NHK NRU Test Methods 7-38

Table 7-9 Interlaboratory 3T3 and NHK NRU Test Method CV Values Sorted by Chemical Characteristics 7-40

Table 7-10 CV Values for 3T3 and NHK NRU Test Method IC_{50} Values and Normalized IC_{50} Values.....7-46

Table 7-11 ANOVA Results for the SLS IC_{50} Values in the 3T3 NRU Test Method ...7-51

Table 7-12 Linear Regression Analysis of SLS IC_{50} Values Over Time 7-52

Table 7-13 ANOVA Results for the SLS IC_{50} Values in the NHK NRU Test Method 7-53

Table 7-14 Solvent Determinations by Laboratory 7-55

Table 8-1 SMT-Recommended Documentation for FAL Laboratory.....8-4

Table 8-2 Phase III Error Rates in the Transfer of Data to the EXCEL® Template.....8-13

Table 8-3 Definitive Test and Positive Control (PC) Test Failure Rates in Phase III..8-14

Table 8-4 Combined Definitive and Positive Control (PC) Test Success Rates for the 3T3 and NHK Methods in Phase III.....8-15

Table 8-5 Coefficients of Variation for Definitive Tests.....8-15

Table 8-6 GHS Acute Oral Toxicity Category Predictions by Laboratory.....8-17

Table 9-1 Rat Acute Oral LD_{50} Ranges for Test Substances Used in Previous *In vitro* NRU Cytotoxicity Studies and the NICEATM/ECVAM Study 9-4

Table 9-2 Chemical Classes Represented by the Substances Used in Published Studies for Correlation of *In vitro* NRU Cytotoxicity with Rodent Acute Lethality 9-4

Table 9-3 EU Classes of Acute Oral Toxicity.....9-10

Table 9-4 Correct Identification of *In vivo* Phototoxicants by the NHK NRU Phototoxicity Assay 9-24

Table 10-1 Change in Animal Use with Dose-Mortality Slope for the UDP 10-11

Table 10-2 Mean Animal Use in the UDP Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the Different Regressions 10-13

Table 10-3 Animal Use for the UDP by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression.....10-17

Table 10-4 Animal Use for the UDP by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression.....10-18

Table 10-5 Animal Deaths in the UDP Using Starting Doses Based on the 3T3 and NHK NRU Test Methods 10-20

Table 10-6 Change in Animal Use with Dose-Mortality Slope in the ATC Method ... 10-25

Table 10-7 Animal Use for the ATC Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with Both Regressions 10-27

Table 10-8 Animal Savings for the ATC Method by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression 10-30

Table 10-9 Animal Savings for the ATC Method by GHS Acute Oral Toxicity Category Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression 10-33

Table 10-10 Animal Deaths for the ATC Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods 10-35

Table 11-1 Costs for Major Laboratory Equipment 11-11

Table 11-2 Costs for Cell Culture Materials and Commercial Laboratory *In Vitro* Cytotoxicity Testing 11-13

Table 11-3 Commercial Prices for Conducting *In Vivo* Acute Rat Toxicity Testing ... 11-14

Table 11-4 Comparison of Time Needed for *In Vitro* and *In Vivo* Testing 11-17

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}	Molar 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 Center 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	Hours
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 concentration
LD_{50}	Dose that produces lethality in 50% of test animals
LDH	Lactate dehydrogenase
ln	Natural logarithm
LPS	Lipopolysaccharide
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
NR_{90}	Concentration producing 10% growth inhibition in the NR assay
NR_{50}	Concentration producing 50% growth inhibition in the NR assay
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)
PM	Prediction Model
PPIS	EPA Pesticide Product Information System
PPT	Precipitate
QA	Quality assurance
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
SHS	Stabilized human serum
SIDS	OECD Screening Information Data Sets
SIS	Scientific Information Service
SLS	Sodium laurel 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)

ACKNOWLEDGEMENTS

The following individuals are acknowledged for their contributions to thein 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, Ph.D.
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D. (ICCVAM Co-Chair)

Department of Energy (DOE)

Po-Yung Lu, Ph.D.

Environmental Protection Agency (EPA)

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

Food and Drug Administration (FDA)

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

National Institute for Occupational Safety & Health (NIOSH)

Stephen Reynolds, Ph.D.

National Institute of Environmental Health Sciences (NIEHS)

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

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., Diplomate, ACLAM (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)
Dr. Martin Wenk – Principal Investigator

Institute for In Vitro Sciences (IIVS)
(Lead Laboratory – Protocols)
Mr. Hans Raabe – Study Director
Mr. Greg Mun – Laboratory Manager
Ms Angela Sizemore – Research Technician
Mr. Gregory O. Moyer – Research Technician
Dr. John Harbell – Scientific Director

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

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

Constella Group
(Statistical Analyses)
Dr. Patrick Crockett
Dr. Eric Harvey
Mr. Robert Lee
Ms Janine Wilcox

Private Statistical Consultant
Dr. Joseph Haseman

NIEHS
Dr. Grace Kissling – Contract Project Officer
Ms Molly Vallant - Contract Project Officer

Study Management Team (SMT)
NICEATM
Dr. William Stokes (NIEHS) – Director
Dr. Raymond Tice (NIEHS) – Asst. Director – Advisor
Dr. Judy Strickland (ILS, Inc.) – Project Coordinator
Mr. Michael Paris (ILS, Inc.) – Asst. Project Coordinator
Dr. Jeffrey Charles (ILS, Inc.) – Advisor

ECVAM
Dr. Thomas Hartung – Head of Unit
Dr. Silvia Casati – Task Leader
Dr. Michael Balls – (Head of Unit until June 2002)

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; 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

[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).

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

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 Center 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 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 two *in vitro* NRU cytotoxicity protocols using 3T3 cells or NHK cells in order to maximize intra- and inter-laboratory reproducibility
- Refine the prediction model drawn from the ZEBET approach
- Assess the accuracy of the two standardized *in vitro* basal cytotoxicity test methods for estimating rodent oral LD_{50} values across the five Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity as well as unclassified toxicities and estimating human lethal serum concentrations
- Estimate the reduction and refinement in animal use achievable from using *in vitro* basal cytotoxicity assays as one of the factors of the weight-of-evidence to identify starting doses for specific *in vivo* acute toxicity tests
- Generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of acute systemic toxicity

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. The public was invited to provide comments on the BRD and other documents and to attend the Panel meeting. Prior to this meeting, any 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. Diplomate A.C.L.A.M.
Director, NICEATM
Executive Director, ICCVAM

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

October 11, 2006

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 mouse fibroblast (BALB/c) 3T3 cells 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 both the numbers of animals per test and 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

[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).

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 and which involved 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).

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.

203 In its recommendations for further evaluations, ICCVAM concurred with the Workshop 2000
204 recommendation that near-term validation studies should focus on two standard basal
205 cytotoxicity assays: one using a human cell (normal human keratinocytes [NHK]) system and
206 one using a rodent cell (3T3) system. Historical data for *in vitro* cytotoxicity testing using
207 mouse 3T3 cells are available (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al.
208 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996), as are historical data for *in vitro*
209 basal cytotoxicity testing using NHK cells (e.g., Gettings et al. 1996; Harbell et al. 1997;
210 Sina et al. 1995; Willshaw et al. 1994).
211
212 NICEATM, in partnership ECVAM, designed an international, multi-laboratory validation
213 study to evaluate the reduction or refinement in animal use that could result from using
214 cytotoxicity data from the 3T3 and NHK NRU test methods to estimate starting doses for two
215 rodent acute oral toxicity test methods, the Up-and-Down Procedure (UDP; OECD 2001a;
216 EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001d). The NRU protocols,
217 as presented in the *Guidance Document,* were the initial basis of the NICEATM/ECVAM
218 validation study protocols. These protocols were originally derived from the BALB/c 3T3
219 Cytotoxicity Test, INVITTOX Protocol No. 46 (available at the FRAME-sponsored
220 INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]), the 3T3 cell studies by
221 Borenfreund and Puerner (1984, 1985) and the rat epidermal keratinocyte study of Heimann
222 and Rice (1983). A detailed description of the 3T3 and NHK NRU test method protocols
223 used in the NICEATM/ECVAM validation study is provided in **Section 2**.
224
225 ***Protocol Components***
226 The protocol components used in the validation study are similar for the 3T3 and NHK cells.
227 The following procedures are common to both cell types:
228 - Testing was performed in four phases (Ia, Ib, II, and III)
229 - Preparation of reference substances and positive control
230 - Cell culture environment conditions
231 - Determination of test substance solubility
232 - Configuration of 96-well plates for testing samples
233 - 48-hour exposure to test substance

234 • Range finder and definitive testing
235 • Microscopic evaluation of cell cultures for toxicity
236 • Measurement of NRU
237 • Data analysis
238
239 The main differences in the test methods for the two cell types are:
240 • The conditions of propagation of the cells in culture
241 • The cell growth medium components
242 • The volumes of reference substance added to the 96-well plate
243
244 Three laboratories participated in testing the 72 reference substances in both cell types:
245 • ECBC: The U.S. Army Edgewood Chemical Biological Center (Edgewood,
246 MD)
247 • FAL: Fund for the Replacement of Animals in Medical Experiments
248 Alternatives Laboratory (Nottingham, UK)
249 • IIVS: The Institute for *In Vitro* Sciences (Gaithersburg, MD)
250
251 BioReliance Corporation (Rockville, MD) procured and distributed the coded reference
252 substances and performed solubility tests prior to their distribution to the testing laboratories,
253 but did not perform any of the *in vitro* tests.
254
255 ***Validation Reference Substances***
256 The 72 reference substances were selected to represent: (1) the complete range of *in vivo*
257 acute oral toxicity (encompassing all five GHS acute toxicity categories as well as lower
258 toxicities [GHS; UN 2005]); (2) the types of substances regulated by various regulatory
259 authorities; and (3) substances with human toxicity data and/or human exposure potential. To
260 ensure that the complete range of toxicity was covered, 12 substances were selected for each
261 of the five acute oral toxicity categories, with an additional 12 substances with lower
262 toxicities (i.e., LD_{50} >5000 mg/kg). A discussion of the characteristics and sources of the
263 reference substances can be found in **Section 3**. The selected reference substances had the
264 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.

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 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

296 • Rats <4 weeks of age
297 • Anesthetized rats
298 • Test substance administered in food or capsule
299 • LD_{50} reported as a range or an inequality
300
301 **Table ES-1 Datasets Used for Validation Study Analyses[1]**
302

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	62	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

303 Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=normal human
304 keratinocytes; NRU=neutral red uptake.
305 [1]Number of substances.
306

307 For substances with multiple LD_{50} values (i.e., from different sources), the rodent reference
308 LD_{50} values for use in the validation study were determined by calculating a geometric mean
309 of the available values for each reference substance. The reference LD_{50} values for 19 (26%)
310 of the 72 substances varied sufficiently from the initial LD_{50} values that came from the RC
311 database and other summary sources, that the substances were reclassified into different GHS
312 categories.
313
314 The reliability of the calculated rat acute oral LD_{50} reference values was assessed by
315 comparison to other evaluations of the performance of rodent acute oral toxicity tests. For the

316 62 reference substances that had more than one LD_{50} value, the maximum:minimum ratios
317 ranged from 1.1 to 25.9, with most below an order of magnitude.
318
319 ***Test Method Accuracy***
320 Although the 3T3 and NHK NRU test methods are not intended to be used as replacements
321 for rodent acute oral toxicity tests, they were evaluated for their ability to correctly predict
322 the reference LD_{50} values (i.e., accuracy[3]). The rationale for evaluating the accuracy of LD_{50}
323 predictions is that the current acute oral toxicity test methods (i.e., UDP, ATC, and Fixed
324 Dose Procedure [FDP; OECD 2001c]) call for starting doses to be placed as close as possible
325 and just below the true LD_{50}. When the starting dose is close to the true LD_{50} for a test
326 substance, fewer animals are needed. When the starting dose is below the true LD_{50}, there is
327 reduced pain and suffering because doses tend to be lower, and the test outcome bias is more
328 conservative (i.e., higher toxicity). Regression models developed using IC_{50} and LD_{50} values
329 were used to derive estimated LD_{50} values from 3T3 or NHK NRU IC_{50} values.
330
331 A number of different analyses were performed in an attempt to improve the estimation of
332 the rat acute oral LD_{50}. IC_{50}-LD_{50} regressions (in millimole units) were calculated for each *in*
333 *vitro* cytotoxicity test method and participating laboratory using the 3T3 and NHK IC_{50}
334 values. Because the regressions for each NRU test method among laboratories were not
335 significantly different from one another (for each NRU test method, $p > 0.5$), the regression
336 for each NRU test method was based on data pooled across the laboratories. This combined-
337 laboratory regression was then compared to the RC data using a regression based on RC IC_{50}
338 and LD_{50} data for the 47 substances common to the validation study and the RC, with rat
339 acute oral LD_{50} reference values, and with both 3T3 and NHK IC_{50} values produced by all
340 three participating laboratories. The statistical comparison of slope and intercept
341 (simultaneously) using an F test showed that neither the 3T3 regression nor the NHK
342 regression was significantly different from the RC regression for the 47 substances ($p = 0.61$
343 and 0.76 respectively). These outcomes support use of the RC millimole regression.
344

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

345 Reference substances that fit the RC millimole regression poorly (i.e., outliers) were
346 evaluated to determine whether there were relationships between their outlier status and their
347 physical or chemical characteristics. Because the IC_{50}-LD_{50} regressions for the 3T3 and NHK
348 NRU test methods yielded results that were not different from the RC regression for 47
349 substances, the RC millimole regression was preferred for analysis of outliers because it was
350 based on a much larger data set and because it had established acceptance limits (Halle 1998,
351 2003). Certain chemical structural classes, boiling points, molecular weights, and log K_{OW}
352 values were related with outliers, but solubility in the 3T3 or NHK medium and the cells'
353 lack of xenobiotic metabolic capability did not correlate with outlier status. Because these *in*
354 *vitro* NRU test methods are based upon basal cytotoxicity, the mechanism of toxicity was
355 also considered as a characteristic to explain the presence of outliers. Twenty-two reference
356 substances were neurotoxic or cardiotoxic and were not expected to be active in the 3T3 and
357 NHK cell cultures. Of these 22 substances, 13 (59%) were outliers (i.e., they fit the RC
358 millimole regression poorly) using the 3T3 NRU and 12 (55%) were outliers using the NHK
359 NRU. These substances represented 13/28 (46%) and 12/31 (39%) of the outliers for the 3T3
360 and the NHK NRU test methods, respectively. More information on the outlier analysis is
361 presented in **Section 6.2**.
362
363 The potential variation produced by combining the LD_{50} values of two rodent species in the
364 RC millimole regression was eliminated by developing a regression based solely on RC
365 substances with rat LD_{50} data (i.e., the RC rat-only millimole regression). The RC rat-only
366 data were also converted to a weight basis for an additional regression, the RC rat-only
367 weight regression, for applicability to mixtures or to substances for which molecular weight
368 is unknown.
369
370 The accuracy of the *in vitro* NRU test methods when used with each of the IC_{50}-LD_{50}
371 regressions was characterized by determining the proportion of reference substances for
372 which their GHS categories (based on rat acute oral LD_{50} data) were correctly predicted. The
373 accuracy of the RC rat-only millimole regression was 31% (21/67 reference substances) and
374 29% (20/68 reference substances) with the 3T3 and the NHK NRU test methods,
375 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 skewedness of the reference substance set with respect to the fit of the reference substances to the regressions. 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 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. ANOVA analyses 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 33 of 64 (52%) reference substances having ratios between 1.5 and 2.5. The maximum:minimum mean laboratory IC_{50} ratios for the NHK NRU test method

437 ranged from 1.0 to 107.6, with 50 of 68 (74%) reference substances having ratios between
438 1.5 and 2.5.
439
440 ***Data Quality***
441 The laboratories reported no significant deviations from the protocols, and deviations that did
442 occur were acknowledged and addressed by the Study Directors. Tests that had deviations
443 affecting the data were rejected by the Study Directors and repeated. The computation of test
444 method and data collection errors showed that the non-GLP laboratory consistently had the
445 highest error rate and the lowest intralaboratory reproducibility for IC_{50} results; however, the
446 laboratory's GHS acute oral toxicity category predictions were comparable to that for the
447 other laboratories.
448
449 An electronic copy of all data for the validation study can be obtained from NICEATM upon
450 request by mail, fax, or e-mail to Dr. William S. Stokes, NICEATM, NIEHS, P. O. Box
451 12233, MD EC-17, Research Triangle Park, NC, 27709, (phone) 919-541-2384, (fax) 919-
452 541-0947, (e-mail) niceatm@niehs.nih.gov.
453
454 ***Other Scientific Reports and Reviews***
455 3T3 and NHK NRU methods have been evaluated for purposes other than the prediction of
456 starting doses for acute toxicity studies (e.g., ocular irritancy; human lethal blood
457 concentrations, *in vivo* phototoxicity). *In vitro* NRU cytotoxicity test methods using various
458 cell types have been evaluated for their correlation with rodent lethality endpoints (e.g.,
459 rat/mouse intravenous[i.v.], intraperitoneal [i.p.], and oral toxicity). Peloux et al. (1992) and
460 Fautrel et al. (1993) showed good correlations (r= 0.88) of *in vitro* cytotoxicity with rodent
461 i.p./i.v. and i.v. toxicity data, respectively. A 3T3 NRU test method has been validated by
462 ECVAM for the identification of *in vivo* phototoxic potential.
463
464 No *in vitro* test methods have been validated for the prediction of acute oral toxicity.
465 Estimations of animal savings using *in vitro* cytotoxicity data to estimate starting doses for
466 the UDP did not use actual *in vitro* cytotoxicity data. Instead, animal savings were estimated
467 by assuming that the *in vivo* starting dose equals the true LD_{50}, which is an approach that

468 assumes that cytotoxicity data can perfectly predict *in vivo* lethality. These theoretical
469 predictions of animal savings in the UDP ranged from 25-40% (ICCVAM 2001a), as
470 compared with the average animal savings of 5.3-7.8% predicted using computer simulation
471 modeling of the UDP for the reference substances tested in the NICEATM/ECVAM study.
472 Halle et al. (1997) used the *in vitro* cytotoxicity data in the RC to determine that an animal
473 savings of 32% can be attained for the ATC method by using the LD_{50} predicted by the RC
474 millimole regression as the starting dose. For the reference substances tested in the
475 NICEATM/ECVAM validation study, most of which were a poor fit to the RC millimole
476 regression, the average animal savings for the ATC, as determined by computer simulation
477 modeling, was 4.8-10.2%.
478
479 ***Animal Welfare Considerations: Reduction, Refinement, and Replacement***
480 Computer models were used to simulate testing of the reference substances according to the
481 UDP and ATC test methods. In principle, animal savings with the FDP could be estimated
482 even though death is not the primary endpoint, but the validation study did not include this
483 analysis. The number of animals that would be used, and the number of animals that would
484 survive or die during the UDP or ATC procedure, were determined for the default starting
485 doses and compared with those when starting dose using simulations for each substance and
486 starting dose was based on LD_{50} values determined from IC_{50} values for each reference
487 substance using the RC rat-only regressions.
488
489 Computer simulation of UDP testing showed that, for the reference substances used in this
490 validation study, using the 3T3 or NHK NRU test methods and the RC rat-only millimole
491 regression to identify the starting dose resulted in the use of fewer animals per test by an
492 average of 5.3% (0.50 animals) to 6.6% (0.53 animals), depending upon the assumed
493 mortality-response slope and *in vitro* NRU test method used. The RC rat-only weight
494 regression predicted mean animal savings of 6.0% (0.56 animals) to 7.8% (0.62 animals).
495 When substances were grouped by GHS acute oral toxicity category, there were no animal
496 savings for substances in the $50 < LD_{50} \leq 300$ mg/kg category because the default starting
497 dose is in this range. The greatest animal savings were observed for substances with $2000 <$
498 $LD_{50} \leq 5000$ 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

530 categories determined using default starting doses. This concordance was 97 to 99% for the
531 3T3 and NHK NRU test methods.
532
533 ***Practical Considerations***
534 Practical issues with respect to the implementation of these *in vitro* NRU test methods
535 include the need for, and availability of, appropriate cell culture equipment, training and
536 expertise, cost, and time expenditure. The ECVAM Good Cell Culture Practice Task Force
537 Report 1 (Hartung et al. 2002) encourages the establishment of laboratory practices and
538 principles that will reduce uncertainty in the development and application of *in vitro* test
539 methods.
540
541 All equipment and supplies are readily available, and the *in vitro* NRU test methods are
542 easily transferable to laboratories experienced with mammalian cell culture techniques. Much
543 of the training and expertise needed to perform the 3T3 and NHK NRU test methods are
544 common to all people with mammalian cell culture experience. Additional technical training
545 would not be intensive because these methods are similar in general performance to other *in*
546 *vitro* mammalian cell culture assays. GLP training should be provided to laboratory
547 personnel (including study directors and principal investigators) to ensure proper adherence
548 to test protocols and data documentation and verification procedures.
549
550 Prices for commercial *in vitro* NRU cytotoxicity testing to determine the IC_{50} for one
551 substance ranged from $1120 to $1850. It is not clear if the price of an *in vivo* test would be
552 reduced if it were preceded by an *in vitro* cytotoxicity test to set the starting dose. Thus, use
553 of these test methods may not reduce the overall cost of rodent acute oral toxicity testing and
554 may increase the cost, but their use has the potential to reduce the number of animals and the
555 time needed for a study. The greatest savings in time and animals will occur if the IC_{50} data
556 determine that the rodent acute oral toxicity limit test should be performed, rather than the
557 main test. Based on the cost and technical procedures associated with cell culture
558 maintenance, the 3T3 NRU test method is less expensive and less complicated to conduct
559 than the NHK NRU test method.
560

[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....1-3

1.1 Historical Background and Rationale for the Use of *In Vitro* Cytotoxicity Assays to Predict Starting Doses for Rodent Acute Oral Toxicity Tests 1-4
1.1.1 The Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) Program 1-5
1.1.2 An International Evaluation of Selected *In Vitro* Toxicity Test Systems for Predicting Acute Systemic Toxicity 1-7
1.1.3 The Registry of Cytotoxicity (RC) 1-9
1.1.4 The ZEBET Initiative to Reduce Animal Use.....1-12
1.1.5 The International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity 1-13
1.1.6 The NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study 1-15

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 1-18
1.2.1 Current Regulatory Testing Requirements for Acute Oral Toxicity.....1-18
1.2.2 Intended Regulatory Uses for *In Vitro* Cytotoxicity Test Methods.....1-21
1.2.3 Similarities and Differences in the Endpoints of *In Vitro* Cytotoxicity Test Methods and Rodent Acute Oral Toxicity Test Methods.....1-21
1.2.4 Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard Assessment 1-23

1.3 Scientific Basis for the *In Vitro* NRU Test Methods 1-24
1.3.1 Purpose and Mechanistic Basis of *In Vitro* NRU Test Methods 1-25
1.3.2 Similarities and Differences in the Modes/Mechanisms of Action for *In Vitro* NRU Test Methods Compared with the Species of Interest.....1-26
1.3.3 Range of Substances Amenable to the *In Vitro* NRU Test Methods.....1-27

[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 of these test methods were used to predict the starting dose.

[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).

The structure of the BRD follows the requested structure of the *ICCVAM 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
- 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 milimole regression, which is 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[2].

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.

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

132 and forensic toxicology handbooks and case reports of human poisonings (Ekwall et al.
133 1998a). The resulting data were presented and analyzed in a series of 50 MEIC Monographs.
134 Rat and mouse oral LD_{50} data were collected from the Registry of Toxic Effects for
135 Chemical Substances (RTECS®)[3].
136
137 The 50 reference substances were tested in as many as 61 different *in vitro* assays (Ekwall et
138 al. 1998b). The metric of interest was the IC_{50} (i.e., the concentration that inhibited the
139 response measured by 50%) for the endpoint measured. Of the 20 test methods that used
140 human-derived cells, 18 used cell lines and two used primary cell cultures. Of the 21 test
141 methods that used mammalian (but other than human) cells, 12 used cell lines and nine used
142 primary cell cultures. Eighteen test methods were ecotoxicological in nature and two used
143 cell-free systems. Cell viability and/or cell growth were the endpoints of choice in the
144 majority of the cell-based systems. The chemical exposure duration ranged from 5 minutes to
145 6 weeks, but most frequently was 24 hours (Clemedson et al. 1996).
146
147 The ability of the *in vitro* IC_{50} data to predict human acute oral lethality was assessed using
148 human LC values compiled from three different data sets (Ekwall et al. 2000):
149 • Clinically measured acute lethal serum concentrations
150 • Acute lethal blood concentrations measured post-mortem
151 • Peak LC values derived from approximate LC_{50} curves over time after
152 exposure
153
154 A partial least squares (PLS) analysis indicated that the IC_{50} data generated from as many as
155 61 test methods predicted the three sets of lethal blood concentration data well (R^2 = 0.77,
156 0.76, and 0.83, Q^2 = 0.74, 0.72, and 0.81, respectively, where R^2 is the determination
157 coefficient and Q^2 is the predicted variance according to cross-validation in the PLS model
158 used). A two component PLS model using rat and mouse oral LD_{50} values less accurately

---

[2] 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.
[3] 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.

159 predicted human lethal blood concentrations (R^2 = 0.65, Q^2 = 0.64). These results suggested
160 that *in vitro* basal cytotoxicity assays might be more effective in estimating human acute oral
161 lethality than rodent acute oral toxicity test methods.
162
163 Because the MEIC study showed that the *in vitro* test methods with the best predictivity
164 generally used human cell lines (Ekwall et al. 1998b), the MEIC management team identified
165 a battery of *in vitro* assays using three human cell lines that had maximal performance for
166 predicting peak acute LC values in humans (R^2=0.79 and Q^2=0.76) (Ekwall et al. 2000).
167 However, it was concluded that improvements in the prediction of human acute oral lethality
168 were necessary before *in vitro* cytotoxicity assays could replace animal tests. To adjust for
169 lethality produced by mechanisms other than basal cytotoxicity, the Evaluation-guided
170 Development of New *In Vitro* Tests (EDIT) program was proposed to address targeted
171 development of *in vitro* test methods for other endpoints, including biokinetics (e.g., gut
172 absorption, distribution, clearance), biotransformation, and target organ toxicity (Clemedson
173 et al. 2002).

174 1.1.2 An International Evaluation of Selected *In Vitro* Toxicity Test Systems for
175 Predicting Acute Systemic Toxicity

176 FRAME organized an international collaborative study conducted in 1992 - 1993 to evaluate
177 the prediction of rodent acute oral lethality by *in vitro* test methods (Fentem et al. 1993)[4].
178 The objective of the study was to identify *in vitro* systems and strategies that could be used
179 for the classification and labelling of new chemicals, thereby reducing, and possibly
180 replacing, the use of animals for acute oral toxicity testing.
181
182 The 42 substances tested in the study comprised a diverse group of organic and inorganic
183 chemical classes, including surfactants, pharmaceuticals, and pesticides (Fentem et al. 1993).
184 *In vitro* toxicity assays using different mammalian cell lines, exposure periods, and toxicity
185 endpoints were evaluated, including:

[4] 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])

186 • Two cell proliferation assays (total protein in mouse BALB/c 3T3 fibroblast
187 cells and MTT[5] reduction in Chinese hamster fibroblastoid V79 cells after a
188 72-hour exposure period)
189 • Two cytolethality assays (MTT reduction in V79 cells and lactate
190 dehydrogenase [LDH] release from primary rat hepatocytes after a 24-hour
191 exposure period)
192 • A cell function assay (myotube contractility inhibition in rat skeletal muscle
193 cells)
194
195 The resulting *in vitro* IC_{50} data were linearly regressed against the lowest available rat or
196 mouse oral LD_{50} values for each test substance. There were no significant differences among
197 the IC_{50}-LD_{50} regressions for the different *in vitro* test methods.
198
199 A subset of 26 to 40 of the 42 test substances, based on the availability of European Union
200 (EU) hazard classification data, was used to evaluate two approaches for using *in vitro* IC_{50}
201 data to classify chemicals into the four hazard categories used by the EU for acute oral
202 toxicity labelling (Fentem et al. 1993). One approach used the IC_{50} values obtained from the
203 five different *in vitro* test methods for each test substance to predict the LD_{50} value and
204 hazard category from the IC_{50}-LD_{50} regression. The accuracy of hazard classification for the
205 five *in vitro* tests was from 43 to 65%. The other approach used toxicokinetic parameters for
206 31 to 38 substances to convert the IC_{50} values to effective dose (i.e., ED_{50}) values. Hazard
207 classification accuracy was 43 to 55%.
208
209 In addition, to ensure assure that a variety of toxic mechanisms were evaluated during *in*
210 *vitro* testing, the lowest predicted LD_{50} or ED_{50} from the results of a battery of three tests: a
211 cell proliferation assay (total protein for 3T3 cells); a cytotoxicity/cytolethality assay using
212 primary rat hepatocytes (LDH release); and the rat skeletal muscle cell contractility assay,
213 was used also. The lowest predicted LD_{50} or ED_{50} of the three tests was then used to predict

[5] 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.

214 toxicity classification. The accuracy of classification using this approach was 48% for the
215 ED_{50} and 45% for the predicted LD_{50} values.
216
217 Based on the results obtained, a battery of *in vitro* tests was recommended for classifying
218 chemicals for their acute lethal potency in rodents (Fentem et al. 1993). The first order test in
219 the battery measures basal cytotoxicity. This study observed no major differences in the
220 performances of the *in vitro* test methods that measure inhibition of cell growth regardless of
221 the cell line (V79, 3T3-L1, or BALB/c 3T3), exposure duration (24-72 hours), or endpoint
222 measurement technique (MTT reduction, neutral red uptake [NRU], or protein
223 concentration). The second order test in the battery assesses hepatocyte-specific toxicity and
224 the role of biotransformation in cytotoxic activity. Co-cultures of rodent hepatocytes with
225 proliferating cells such as 3T3 cells were recommended because the use of hepatocytes alone
226 would not indicate that a chemical requires bioactivation to produce its toxic effects. The
227 third order test in the battery detects chemicals that interfere with electrically excitable
228 membranes at non-cytotoxic concentrations (e.g., a contractility assay using primary cultures
229 of rat muscle cells) (Fentem et al. 1993).

230 1.1.3 The Registry of Cytotoxicity (RC)

231 The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS®, and
232 published IC_{50} values from *in vitro* cytotoxicity assays that used a variety of cell lines and
233 cytotoxicity endpoints for substances with known molecular weights (Halle 1998, 2003). The
234 main purpose for compiling the RC was to evaluate, using data from substances with a wide
235 range of rodent acute oral toxicities, whether basal cytotoxicity (averaged over various cell
236 types, cell lines, and/or toxicity endpoints) accurately predicted acute oral lethality in rats and
237 mice. The RC currently contains data for 347 different substances (Halle 1998, 2003) and
238 efforts are underway to increase the number to 500 (ICCVAM 2001a). The RC does not
239 contain data on chemical mixtures.
240
241 The RC contains cytotoxicity data for substances that met the following criteria (Halle 1998,
242 2003):
243 • At least two different IC_{50} values needed to be available, from studies using
244 either different cell types, different cell lines, or different cytotoxicity endpoints

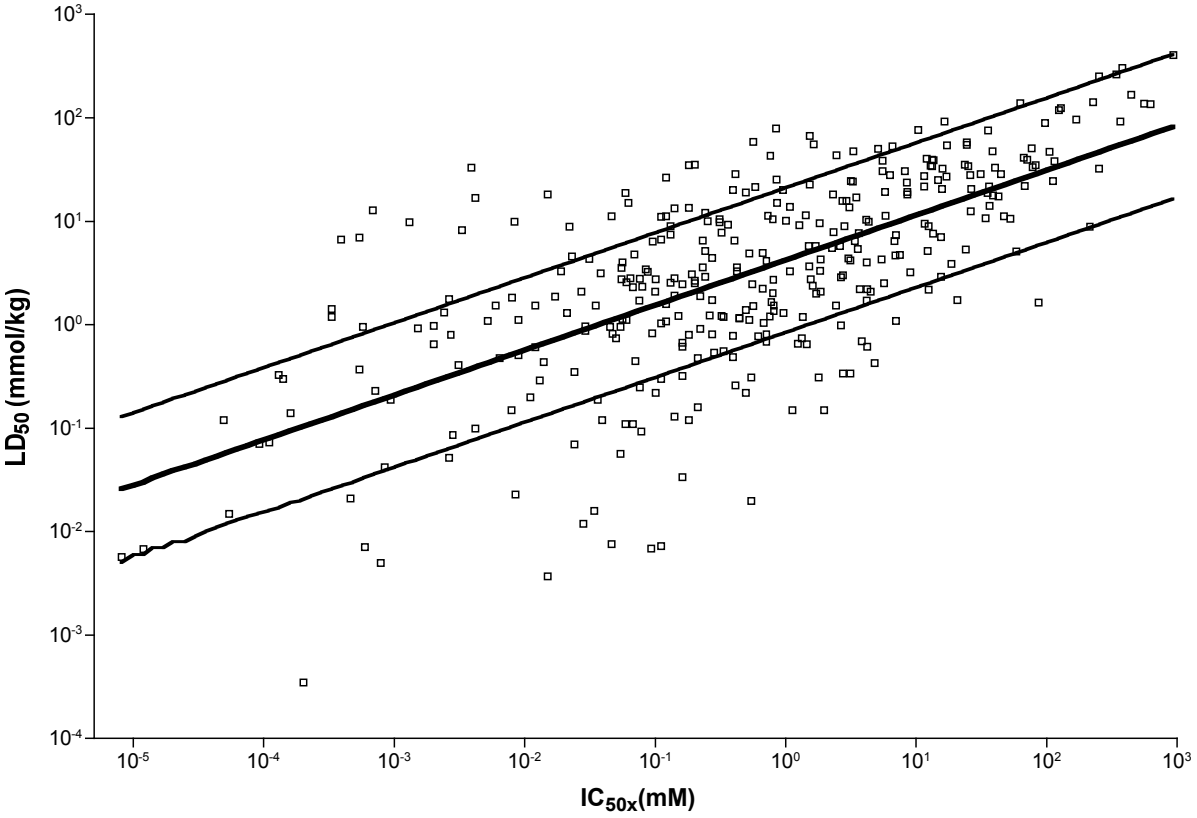
245 • Data had to be generated using mammalian cells only (although data from
246 studies using hepatocytes or related cells were excluded)
247 • The chemical exposure duration had to be at least 16 hours, with no upper limit
248
249 The following cytotoxicity endpoints were accepted:
250 • Cell proliferation: cell number; cell protein; DNA content; DNA synthesis; ^{3}H-
251 thymidine intake; colony formation
252 • Cell viability/metabolic indicators: metabolic inhibition test (MIT-24);
253 mitochondrial reduction of tetrazolium salts into an insoluble (MTT) or soluble
254 (2,3-bis(2-methoxy-4-nitro-5- sulfophenyl)-2H-tetrazolium-5-carboxanilide
255 [XTT]) dye
256 • Cell viability/membrane indicators: NRU; trypan blue exclusion; cell
257 attachment; cell detachment
258 • Differentiation indicators, such as functional and/or morphological changes
259 among and within cells
260
261 IC_{50} values (1,912) for 347 substances were obtained from 157 original publications (Halle
262 1998, 2003). The two to 32 IC_{50} values for each substance were averaged as geometric means
263 to produce one IC_{50x} value for each substance. The rodent LD_{50} values used in the RC were
264 obtained from RTECS®. For the first 117 substances, designated as the training data set (RC-
265 I), LD_{50} values were not revised when subsequent issues of RTECS® reported lower values[6].
266 For the most recent 230 substances, designated as the verification set (RC-II), the LD_{50}
267 values were taken from the 1983/84 RTECS® publication. Whenever obtainable, oral LD_{50}
268 data from rats were used (282 values). If rat data were unavailable, LD_{50} data from mice were
269 used (65 values). Combining rat and mouse data in the regression was deemed to be justified
270 when separate regressions for the mouse and rat LD_{50} values against the IC_{50x} values did not
271 result in significant differences between the slopes and intercepts of the two regressions
272 (Halle 1998, 2003).
273

[6] RTECS® published the lowest LD50 reported for a substance and updates the information periodically.

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:

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

Figure 1-1 RC Millimole Regression Between *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).

293 To identify an acceptability range for practical use and research purposes, the acceptable
294 prediction interval for the LD_{50} was empirically defined as approximately one-half an order
295 of magnitude on either side of the best-fit linear regression (i.e., ± log 5, or ± 0.699) (Halle
296 1998, 2003). This interval was based on eight linear regressions calculated for *in vitro*
297 mammalian cell cytotoxicity data using various endpoints and oral LD_{50} values from rat,
298 mouse, or rat and mouse from five publications. The prediction interval approximates the
299 predicted LD_{50} range for the eight regressions across about eight orders of magnitude of IC_{50}
300 values. When this approach was used, 73% (252/347) of the RC substances fall within the
301 prediction interval.

302 1.1.4 The ZEBET Initiative to Reduce Animal Use

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

307 • The active concentration *in vitro* (i.e., the actual concentration available to the
308 cultured cells)
309 • The *in vitro* concentrations that produce basal cytotoxicity, hepatocyte
310 toxicity, and selective cytotoxicity (i.e., effects on cell-specific functions such
311 as transport processes or cell-to-cell communication)
312 • The effect of biokinetic processes on acute oral toxicity in rodents
313 • *In vitro* tests that provide the physicochemical parameters needed to estimate
314 equivalent body doses from *in vitro* data
315

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

321 • Acute Toxic Class method (ATC; OECD draft Test Guideline [TG] 423
322 [ICCVAM 2001a])
323 • 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 for the UDP, 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

355 their usefulness and limitations for generating information for acute toxicity testing had not
356 been assessed. However, there was agreement that: (1) in the near-term, *in vitro* basal
357 cytotoxicity test methods would be useful for estimating the starting dose for rodent acute
358 oral toxicity studies, and (2) further development, optimization, and validation of *in vitro* test
359 methods that considered target organ specificity and *in vivo* factors like adsorption,
360 distribution, metabolism, and excretion (ADME) that modulate the lethality of a xenobiotic
361 were needed (ICCVAM 2001a). Furthermore, the approach proposed by ZEBET (i.e., the use
362 of *in vitro* basal cytotoxicity test methods to predict the starting dose for the sequential rodent
363 acute oral toxicity test methods) (Halle 1998, 2003; Spielmann et al. 1999) was
364 recommended for rapid adoption so that data could be generated to establish its usefulness
365 with a larger number of substances (ICCVAM 2001a). To assist in the adoption and
366 implementation of the ZEBET approach, several workshop participants prepared the
367 *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute*
368 *Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b).
369
370 The *Guidance Document* recommended testing 10 to 20 RC substances (of high purity) from
371 the RC in a candidate *in vitro* basal cytotoxicity assay to be used for predicting starting doses
372 for acute oral lethality tests (ICCVAM 2001b). The substances were to cover a wide range of
373 toxicities and fit the RC prediction model (i.e., the linear regression line) as closely as
374 possible. The *in vitro* test methods recommended and provided as examples were NRU
375 assays using 3T3 and normal human keratinocytes (NHK) cells. The IC_{50} results from testing
376 the selected substances would be used to calculate a regression against the LD_{50} values used
377 by the RC. If the resulting regression were parallel to the RC millimole regression and within
378 the ± log 5 (i.e., ± 0.699) prediction interval for the RC, the *Guidance Document*
379 recommended using the *in vitro* cytotoxicity assay to predict starting doses for LD_{50} assays.
380 If the regression from the *in vitro* assay did not meet these criteria, then the *Guidance*
381 *Document* advised either (a) adjusting the slope or (b) using the NRU protocols offered in the
382 *Guidance Document* (considered the most efficient approach).
383
384 Based on the conclusions and recommendation of the Workshop 2000 participants, ICCVAM
385 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[7] 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 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

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

414 Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity, as
415 well as unclassified toxicities
416 • Estimate the reduction and refinement in animal use achievable from using the
417 *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods to identify starting
418 doses for *in vivo* acute oral toxicity tests, assuming that no other information
419 were available
420 • Develop high quality *in vivo* acute oral lethality and *in vitro* NRU cytotoxicity
421 databases that can be used to support the investigation of other *in vitro* test
422 methods necessary to improve the prediction of *in vivo* acute oral lethality

423 1.1.6.1 *Study Design*

424 The planning phase of the validation study included the selection of reference substances for
425 testing, which is described in **Section 3,** and the identification of rodent oral LD_{50} values for
426 the reference substances, which is described in **Section 4**. The validation study proceeded in
427 several phases (see **Figure 1-2**) so that the Study Management Team (SMT) could evaluate
428 the reproducibility of results after each phase and refine the protocols, if necessary, before
429 proceeding to the next phase. The resulting NRU data collected were used to evaluate linear
430 regression formulas for the prediction of LD_{50} values from IC_{50} values (see **Section 6**).
431 Computer simulation modeling of acute oral toxicity test outcomes was then performed to
432 determine potential animal savings using the NRU-predicted starting doses compared with
433 the default starting dose for the UDP and the ATC (see **Section 10**). Study management and
434 study participant information is provided in **Appendix A.**
435

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).

499 **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

500 Abbreviations: EPA=U.S. Environmental Protection Agency; OECD=Organisation for Economic Co-operation and Development;
501 LD_{50}=dose producing death in 50% of the animals tested; CPSC=U.S. Consumer Product Safety Commission; FIFRA = Federal
502 Insecticide, Fungicide, and Rodenticide Act;; OSHA=U.S. Occupational Safety and Health Administration; DOT=U.S. Department of
503 Transportation; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
504 [1]*Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used*
505 *in Safety Evaluation* calls for humane killing of moribund animals (OECD 2000). Moribund animals that are humanely euthanized are
506 accepted as deaths.
507

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

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

520 1.2.1.1 *Test Methods for Assessing Acute Oral Toxicity*

521 The current internationally recognized test methods for acute oral toxicity testing are the FDP
522 (OECD 2001c), the ATC (OECD 2001d), and UDP (OECD 2001a; EPA 2002a) (see
523 **Appendix M** for test method guidelines). Information on toxic doses and signs of acute
524 toxicity and target organs can be obtained using any of these three methods. All three
525 methods are sequential tests in which the outcome of testing one or more animals at the first
526 dose is used to determine the second dose that should be tested. The FDP differs from the
527 UDP and ATC in that it involves using more animals per dose, and the primary endpoint of
528 interest is evident toxicity[8] rather than lethality. Both the FDP and the ATC methods provide
529 a range estimate of the LD_{50} for classification purposes. The UDP generally provides a point
530 estimate of the LD_{50} with a confidence interval (EPA 2002a).
531
532 Each of the test method guidelines includes a limit test in which up to five or six animals are
533 tested at the limit, or upper bound, dose depending on the dose chosen (OECD 2001a, c, d, e;
534 EPA 2002a). The limit test can be performed using 2000 or 5000 mg/kg, depending on the
535 regulatory need.

---

[8] *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).

536 1.2.2 Intended Regulatory Uses for the *In Vitro* Cytotoxicity Test Methods

537 *In vitro* cytotoxicity test methods currently cannot serve as replacements for acute oral
538 toxicity tests in animals. However, such test methods can be used as adjuncts for rodent acute
539 oral toxicity tests. The current test guidelines for acute oral toxicity tests recommend using
540 information from structurally-related substances and the results of any other toxicity tests
541 (EPA 2002b), including *in vitro* cytotoxicity test method (OECD 2001a, c, d, e; EPA 2002a)
542 to select the starting *in vivo* dose. The 3T3 and NHK NRU test methods may be used as part
543 of this weight-of-evidence approach to select starting doses in order to reduce and refine the
544 use of animals for acute oral toxicity testing.
545
546 **Section 10** presents computer simulation analyses that characterize the extent of animal
547 reduction and refinement that may occur by using the *in vitro* NRU test methods to estimate
548 the starting doses for the UDP and the ATC method, by estimating the numbers of animals
549 used and the numbers of animal that die. These simulations determined (1) the numbers of
550 animals used when using the default starting dose and, (2) the number of animals used when
551 using a starting dose determined from the *in vitro* NRU test methods. These calculations
552 determined the reduction in animal use that can be achieved when using the *in vitro* NRU test
553 methods. To characterize the extent of refinement produced using the NRU-determined
554 starting dose, the number of animals that would have died with the NRU-determined starting
555 dose was compared with the number of animals that would have died when using the default
556 starting dose. Because there is a lack of information for specific substances about the dose at
557 which evident toxicity occurs and the relationship of this dose to the LD_{50}, the FDP will not
558 be considered further in this document. However, the use of *in vitro* cytotoxicity data to
559 determine starting doses may also reduce the use of animals in the FDP.

560 1.2.3 Similarities and Differences in the Endpoints of the *In Vitro* Cytotoxicity Test
561 Methods and Rodent Acute Oral Toxicity Test Methods

562 The endpoint measured in the *in vitro* NRU test methods is cell death. Neutral red dye is
563 taken up and accumulated only by live cells; the primary measure of interest is the IC_{50} (i.e.,
564 the test substance concentration that causes a 50% inhibition of NRU). In contrast, the
565 endpoint measured in acute oral toxicity assays is usually animal morbidity or death. Cell
566 death and animal death may have similar mechanistic bases because all cells, regardless of

567 whether they are in animals or *in vitro* cell cultures, have similar cellular mechanisms; for
568 example, energy production and maintenance of cell membrane integrity.
569
570 Death of an animal death and death of a cultured cell due to toxicity both involve interference
571 with vital cell processes or physical injury. Cell death in a culture system involves the death
572 of a single cell type, but through mechanisms that also operate in the animal. In contrast,
573 cellular injury in an animal, if sufficiently widespread or in a critical process, can lead to
574 injury or loss of function of other cell types in a tissue not directly affected by the treatment,
575 resulting in organ failure. Major organ system failures (e.g., liver and kidney failure),
576 gastrointestinal corrosion, and bone marrow depression, can be fatal. Examples of
577 mechanisms leading to such organ failures are disruption of membrane structure or function,
578 inhibition of mitochondrial function, disturbance of protein turnover, and disruption of
579 energy production (Gennari et al. 2004). Alternatively, the tissue injury could affect non-
580 exposed vital organs or tissues through interference with homeostatic signaling mechanisms
581 (Gennari et al. 2004). For example, respiratory depression leading to death may be due to
582 depression of the central nervous system (CNS) rather than a direct assault on the respiratory
583 system itself.
584
585 Animal and cell culture systems are also different with respect to how a substance or toxicant
586 is delivered to the cell and how it is distributed within the cell, metabolized, and excreted.
587 After oral administration, animals must absorb the toxicant from the gastrointestinal tract,
588 which involves the passage through membranes, many of which are selective with respect to
589 what molecules they will allow to pass. The toxicant may or may not be bound to serum
590 proteins, thereby reducing its availability to the target organ. The toxicant may be
591 metabolized before, during, and/or after its distribution to the target organs, or the toxicant or
592 its metabolites may be excreted before reaching the target organ or reacting with its
593 components. As a consequence, the most critical target organs may not be exposed to the
594 active metabolite, or be exposed for only a limited time or to a relatively small fraction of the
595 administered dose.
596
597 In contrast, in a cell culture system, the test substance is applied directly to the target cells

598 and the only membranes that must be passed are those of the target cell and its subcellular
599 organelles. No absorption and distribution by other cellular systems is required. Cell culture
600 systems may or may not include serum proteins, which could reduce the availability of
601 toxicant to the target site. For example, the 3T3 cell culture medium includes serum while the
602 NHK cell culture medium does not. 3T3 and NHK cells have little to no capacity to
603 metabolize xenobiotic compounds, and added cell-free metabolic activation systems, such as
604 rat liver homogenates, may not accurately mimic all phases of *in vivo* metabolism. Excretion
605 from the cell culture milieu is not a consideration because anything excreted from the cell
606 remains in the culture medium and is available to the other cells in the culture. As a result,
607 the cells in culture (as opposed to cells in an animal) may be exposed to a test substance for
608 the entire duration of the test protocol.
609
610 Animals and cell culture systems may also differ with respect to the target on which a
611 toxicant acts. If a toxicant acts in a specialized organ system *in vivo*, it may not produce a
612 toxic effect by the same mechanism in cultured cells that are derived from a tissue different
613 from the target organ. For example, a substance that affects a neuroreceptor-mediated
614 pathway in animals would not be expected to produce a similar toxicity in 3T3 or NHK cells,
615 which are derived from fibroblasts and skin cells, respectively, and do not contain similar
616 neuroreceptors; if toxicity is seen in these cell cultures, it may be from a different mechanism
617 or in a different concentration relationship than *in vivo*. Even if a neurotoxin were applied to
618 neuronal cells in culture, the cultured cells may not respond in the same way as neuronal
619 cells in an animal because cells in culture, especially cell lines, may not retain the same
620 functionalities as cells *in vivo*.

621 1.2.4 Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard
622 Assessment

623 In the overall strategy of hazard or safety assessment, the intended regulatory use of the *in*
624 *vitro* NRU test methods is to reduce and refine the use of animals in current acute toxicity
625 assays. The *in vitro* systems would serve as adjuncts to the *in vivo* test methods but are not
626 intended as replacements for the rodent acute oral toxicity test methods. For the OECD
627 alternative acute oral toxicity assays (the ATC and UDP), the number of animals used
628 depends on the starting dose. The number of dosing steps (and animals) is reduced if the

629 starting dose is close to the true toxicity class (ATC) or the true LD_{50} (UDP) (Spielmann et
630 al. 1999; ICCVAM 2001b).
631
632 As noted earlier, Spielmann et al. (1999) and the *Guidance Document* (ICCVAM 2001b)
633 suggest that the RC millimole regression analysis be used with *in vitro* cytotoxicity data to
634 predict starting doses for the ATC and UDP. The RC millimole regression cannot be applied
635 to unknown substances or to mixtures (e.g., product formulations) because such materials
636 cannot be assigned molecular weights. Therefore, the NICEATM/ECVAM validation study
637 also evaluated the classification accuracy and the reduction in animal use associated with a
638 regression based on weight units (with IC_{50} in μg/mL and LD_{50} in mg/kg) (see **Section 10**).
639 This regression would potentially be appropriate for predicting the starting dose for mixtures
640 and undefined substances.

641 **1.3 Scientific Basis for the *In Vitro* NRU Test Methods**

642 Cytotoxicity has been defined as the adverse effects resulting from interference with
643 structures and/or processes essential for cell survival, proliferation, and/or function (Ekwall
644 1983). Ekwall (1983) described the concept of "basal cell functions" (mitochondrial activity,
645 plasma membrane integrity, etc.) that virtually all cells possess and suggested that, for most
646 substances, toxicity is a consequence of non-specific alterations in those cellular functions,
647 which may then lead to adverse effects on organ-specific functions and/or death of the
648 organism. These effects may involve the integrity of membranes and the cytoskeleton,
649 cellular metabolism, the synthesis and degradation or release of cellular constituents or
650 products, ion regulation, and cell division.
651
652 Ekwall (1983) and others (e.g., Grisham and Smith 1984) concluded that, because the actions
653 of substances that produce injury and death are ultimately exerted at the cellular level, *in*
654 *vitro* cytotoxicity assays might be useful for the prediction of acute lethality potency, as well.
655 Considerable research has been undertaken to develop and evaluate *in vitro* tests for use as
656 screens and as potential replacements for rodent LD_{50} tests, and numerous groups have
657 reported good agreement between *in vitro* cytotoxicity and animal lethality (see reviews by
658 Phillips et al. 1990; Garle et al. 1994; Guzzie 1994). However, none of the proposed *in vitro*
659 models have been evaluated in any formal studies for reliability and relevance, and their

660 usefulness and limitations for generating information to meet regulatory requirements for
661 acute toxicity testing data have not been assessed.

662 1.3.1 Purpose and Mechanistic Basis of the *In Vitro* NRU Test Methods

663 A number of basal cytotoxicity endpoints can be used to measure cell death or interference
664 with cell proliferation. The NRU test methods were chosen for the NICEATM/ECVAM
665 validation study because they were recommended in the *Guidance Document* for the purpose
666 of obtaining cytotoxicity information to determine starting doses for rodent acute oral
667 toxicity assays (ICCVAM 2001b). Both the 3T3 and NHK NRU test methods were
668 reproducible in previous validation studies (ICCVAM 2001b). In addition, both cell types are
669 easily obtainable from commercial sources and the *Guidance Document* provided
670 preliminary evidence that these assays could reproduce the RC millimole regression.
671 Additionally, the assays can be automated and they require no radioactivity or highly
672 dangerous reagents (see **Section 2** for discussion of the protocols and **Appendix B** for the
673 protocols).
674
675 Neutral red is a weakly cationic water-soluble supravital dye that stains living cells
676 (Borenfreund and Puerner 1985). It readily diffuses through the plasma membrane and
677 concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix.
678 Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility
679 and other adverse changes that gradually become irreversible. Thus, cell death and/or
680 inhibition of cell growth decreases the amount of neutral red retained by the culture.
681 Borenfreund and Puerner (1985) were the first to publish a protocol for the NRU assay using
682 3T3 cells as a method to objectively quantify toxicity previously assessed by subjective,
683 visual observation. The NRU assay, which was standardized for a 96-well plate format,
684 correlated two measurements of toxicity from the exposure of 3T3 cells to six surfactants: (1)
685 a visual morphological evaluation of the cells under an inverted phase microscope, and (2) a
686 quantitative measurement of NRU. The visual evaluation was designed to identify the highest
687 concentration of toxicant that causes only minimal morphological changes (i.e., the highest
688 tolerated dose [HTD]). Because Borenfreund and Puerner (1985) found that the HTD in the
689 NRU test was comparable to the concentration that produced 10% inhibition (i.e., the IC_{10})
690 compared with the controls, the IC_{10} value was deemed to be a good index for comparing the

691 relative toxicities of experimental agents. The assay was described as a rapid, reliable,
692 inexpensive, and reproducible *in vitro* test method for screening potentially toxic agents
693 (Borenfreund and Puerner 1985). Furthermore, the authors suggested that the test method
694 was a good candidate for inclusion in a battery of assays for toxicity screening with the
695 purpose of reducing the use of animals for toxicity tests.

696 1.3.2 Similarities and Differences in the Modes/Mechanisms of Action for the *In Vitro*
697 NRU Test Methods Compared with the Species of Interest

698 Although the ultimate species of interest for acute oral toxicity concerns is humans, labeling
699 and hazard identification requirements are based on rodents. There are differences between
700 humans and rodents in terms of absorption, distribution, metabolism, excretion, and the
701 intrinsic sensitivity of target organs to xenobiotic compounds. The differences are largely
702 substance-specific and quantitative, although there are a number of substances where the
703 human may produce metabolites not seen in the rodent and vice versa. *In vitro* cytotoxicity
704 studies have also noted differences in sensitivity between human cells and other mammalian
705 cells (Clemedson et al. 1996b). It is important to note that, for certain chemicals, there can
706 also be large differences in sensitivity among different human cell types and cell lines
707 (Clemedson et al. 1996b, 1998a, b).
708
709 Because of the differences in sensitivity between humans and rodents, it might be likely that
710 cultured human cells would predict human lethality better than cultured rodent cells and that
711 cultured rodent cells would predict rodent lethality better than human cells. Ekwall et al.
712 (1998b) showed that *in vitro* cytotoxicity test methods using human cell lines generally
713 predicted human toxicity more accurately than did test methods using nonhuman mammalian
714 cells.
715
716 In addition to being derived from different species, there are several other differences
717 between 3T3 and NHK cells, all of which may contribute to differences in sensitivity.

718 - 3T3 cells are an immortal line, while the NHK cells are primary cells.
719 - The cells originate from different tissues; 3T3 cells are derived from
720 embryonic fibroblasts, while the NHK cells are isolated from neonatal
721 foreskin tissue.

722 • NHK cells grow more slowly in culture than the 3T3 cells (i.e., after seeding
723 into 96-well plates, NHK cells require 48-72 hours for growth to the
724 appropriate confluence while 3T3 cells require approximately 24 hours; see
725 **Appendix B**).
726 • NHK cells have greater ability to metabolize xenobiotic compounds, in that
727 they exhibit minimal cytochrome P450 activity (Babich et al. 1991), whereas
728 3T3 cells have practically no ability to metabolize xenobiotic compounds
729 (INVITTOX 1991).

730 1.3.3 Range of Substances Amenable to the *In Vitro* NRU Test Methods

731 The *in vitro* NRU test methods can be applied to a wide range of substances as long as they
732 can be dissolved in the cell culture medium or in a nontoxic solvent (at the concentration
733 used), and do not react with the culture medium. Although these test methods may to be
734 applicable to mixtures, none were evaluated in this validation study. The toxicity of
735 substances that act by mechanisms not expected to be active in 3T3 or NHK cells (e.g., those
736 that are specifically neurotoxic or cardiotoxic) will likely be underpredicted by these test
737 methods. Therefore, until more appropriate cell lines are developed, the results from basal
738 cytotoxicity testing with such substances may not be relevant for predicting *in vivo* effects.
739
740 Insoluble substances or those unstable in aqueous environments are not compatible with the
741 test systems. Volatile substances may yield acceptable results if CO_2 permeable plastic film
742 is used to seal the test plates. Testing for corrosive substances is unnecessary since there is no
743 regulatory requirement for acute oral toxicity testing for known corrosives. The 3T3 NRU
744 test method may underestimate the toxicity of substances that are highly bound to serum
745 proteins because the culture medium contains 5% serum during substance exposure. The
746 toxicity of substances that specifically affect lysosomes may be overestimated because they
747 may affect NRU binding, and therefore, retention, in the cell. Red substances (and other
748 colored substances) that absorb light in the optical density range of NR may interfere with
749 the test if they remain inside the cell in sufficient amounts after washing and are soluble in
750 the NR solvent.
751

751
752
753
754
755
756
757
758
759
760
761
762
763
764
765 ***[This Page Intentionally Left Blank]***
766

2.0 TEST METHOD PROTOCOL COMPONENTS OF THE 3T3 AND NHK *IN VITRO* NRU TEST METHODS 2-3

2.1 Basis for Selection of *In Vitro* NRU Cytotoxicity Test Methods 2-4
2.1.1 *Guidance Document* Rationale for Selection of *In Vitro* NRU Cytotoxicity Test Methods 2-5
2.1.2 *Guidance Document* Rationale for Selection of Cell Types 2-5

2.2 Overview of the 3T3 and NHK NRU Test Methods 2-6
2.2.1 The 3T3 NRU Test Method 2-9
2.2.2 The NHK NRU Test Method 2-10
2.2.3 Measurement of NRU in the 3T3 and NHK Test Methods 2-11

2.3 Descriptions and Rationales of the 3T3 and NHK NRU Test Methods 2-11
2.3.1 Materials, Equipment, and Supplies 2-12
2.3.2 Reference Substance Concentrations/Dose Selection 2-15
2.3.3 NRU Endpoints Measured 2-17
2.3.4 Duration of Reference Substance Exposure 2-18
2.3.5 Known Limits of Use 2-19
2.3.6 Basis of the Response Assessed 2-21
2.3.7 Appropriate Vehicle, Positive, and Negative Controls 2-21
2.3.8 Acceptable Ranges of Control Responses 2-22
2.3.9 Nature of Experimental Data Collected 2-24
2.3.10 Data Storage Media 2-26
2.3.11 Measures of Variability 2-26
2.3.12 Methods for Analyzing NRU Data 2-27
2.3.13 Decision Criteria for Classification of Reference Substances 2-27
2.3.14 Information and Data Included in the Test Report 2-27

2.4 Proprietary Components of the *In Vitro* NRU Test Methods 2-29

2.5 Basis for the Number of Replicate and Repeat Experiments for the 3T3 and NHK NRU Test Methods 2-30

2.6 Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols 2-31
2.6.1 Phase Ia: Laboratory Evaluation Phase 2-31
2.6.2 Phase Ib: Laboratory Evaluation Phase 2-34
2.6.3 Phase II: Laboratory Qualification Phase 2-39
2.6.4 Phase III: Laboratory Testing Phase 2-47

2.7 Differences Between the 3T3 and NHK NRU Test Method Protocols for the Validation Study and the *Guidance Document* Standard Protocols 2-48

2.8 Overview of the Solubility Protocol 2-50

2.9 Basis of the Solubility Protocol.....2-51
2.9.1 Initial Solubility Protocol Development 2-52
2.9.2 Basis for Modification of the Phase II Protocol 2-53

2.10 Components of the Solubility Protocol 2-54
2.10.1 Medium, Supplies, and Equipment Required.....2-54
2.10.2 Data Collection 2-55
2.10.3 Variability in Solubility Measurement.....2-56
2.10.4 Solubility Issues During the Testing of the Reference Substances 2-56
2.10.5 Analysis of Solubility Data 2-56

2.11 Summary.....2-57

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 which was used in Phase III. **Section 2.7** describes the differences between the protocols used in this study and the

93 protocols outlined in the *Guidance Document*. **Sections 2.8**, **2.9**, and **2.10** provide details on
94 the solubility protocol evaluated during the validation study and used to identify the
95 appropriate solvent for dissolving the reference substances.
96
97 The 3T3 and NHK NRU test method protocols were provided to the three laboratories that
98 participated in the validation study (see **Section 5.6.3** for additional laboratory information).
99 These were:
100 - The U.S. Army Edgewood Chemical Biological Center (ECBC)
101 - The FRAME Alternatives Laboratory (FAL)
102 - The Institute for *In Vitro* Sciences (IIVS)
103
104 A fourth laboratory (BioReliance Corporation, Rockville, MD) was used to procure and
105 distribute the coded reference substances, and to perform solubility tests on the validation
106 study reference substances prior to their distribution to the participating laboratories

107 **2.1 Basis for Selection of the *In Vitro* NRU Cytotoxicity Test Method**

108 As stated in **Section 1**, in agreement with the recommendations of the Workshop 2000
109 participants (ICCVAM 2001a), ICCVAM made the following recommendations and
110 forwarded them to U.S. Federal agencies along with the Workshop 2000 Report (ICCVAM
111 2001a) and *Guidance Document* (ICCVAM 2001b).
112 "ICCVAM concurs with the Workshop recommendation that near-term
113 validation studies should focus on two standard cytotoxicity assays: one
114 using a human cell system and one using a rodent cell system. Since the
115 murine BALB/c 3T3 cytotoxicity assay has been evaluated for only a
116 limited number of chemical classes, there is merit in determining its
117 usefulness with a broader array of chemical classes. Cell lines established
118 from the rat rather than the mouse might also be considered, as most acute
119 oral toxicity testing is conducted in this species. Human cell lines should
120 also be considered since one of the aims of toxicity testing is to make
121 predictions of potential toxicity in humans. Future validation studies
122 should therefore compare rodent and human *in vitro* data with one another,
123 with rodent *in vivo* data, and with human *in vivo* data. Correlations

124 between *in vitro* and *in vivo* data might help in selecting cytotoxicity
125 assays for further evaluation". (ICCVAM 2001a)
126
127 Based on this recommendation and the *Guidance Document* recommendation, NICEATM
128 and ECVAM selected the 3T3 and NHK NRU basal cytotoxicity test methods for validation.

129 2.1.1 *Guidance Document* Rationale for Selection of *In Vitro* NRU Cytotoxicity Test
130 Methods

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

136 2.1.2 *Guidance Document* Rationale for Selection of Cell Types

137 The Workshop 2000 participants (ICCVAM 2001a) concluded that there were no significant
138 differences between the basal cytotoxicity results obtained using permanent mammalian cell
139 lines, primary human cells (e.g., NHK cells), or the IC_{50x} approach of Halle and Spielmann
140 (Halle 1998, 2003; Spielmann et al. 1999; Halle and Spielmann 1992). Further, the *Guidance*
141 *Document* recommended that *in vitro* basal cytotoxicity test methods not use hepatocytes (or
142 related metabolically competent cells) or other types of highly differentiated cells because
143 they may not give the best prediction of acute lethality for the large variety of chemicals
144 likely to be tested (ICCVAM 2001b). However, it was recognized that, ultimately, simple
145 predictive systems (*in vitro* or *in silico*) would be needed for early identification of those
146 substances likely to be metabolized to more toxic or less toxic species than the parent
147 chemical as well as those that were likely to exhibit cell-specific toxicity (e.g., Fentem et al.
148 1993; Seibert et al. 1996; Curren et al. 1998; Ekwall et al. 1999).
149
150 Established rodent cell lines were recommended for validation because (ICCVAM 2001b):
151 • It was assumed that such cells would give the best prediction of rat and mouse
152 acute oral lethality (i.e., like correlates with like).

153 • The use of a readily available, easy to culture, immortalized cell line for *in*
154 *vitro* cytotoxicity testing would accelerate the development of a database that
155 can be used to analyze the usefulness of this approach.
156
157 Human cells also offer potential advantages. As determined in the MEIC project, the *in vitro*
158 test methods with the best predictivity for peak acute LC values in humans generally used
159 human cell lines (Ekwall et al. 1998b). Thus, a long-term advantage of using human cells is
160 that *in vitro* human cell cytotoxicity data can be added to human toxicity databases to
161 facilitate the development of test methods that may better predict acute oral human lethality.
162
163 3T3, an immortalized mouse fibroblast cell line, and NHK, primary human cells, were
164 selected as representative rodent and human cells, respectively, for the NICEATM/ECVAM
165 validation study. Historical data for the 3T3 NRU test were available from a variety of
166 studies, including controlled and blinded validation studies, indicating the reliability of this
167 test method (Gettings et al. 1991, 1992, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996;
168 Balls et al. 1995; Brantom et al. 1997). NHK cells have also been used in validation studies
169 for basal cytotoxicity test methods with good results (Willshaw et al. 1994; Sina et al. 1995;
170 Gettings et al. 1996; Harbell et al. 1997).

171 **2.2 Overview of the 3T3 and NHK NRU Test Methods**

172 The *Guidance Document* (ICCVAM 2001b) includes a proposed 3T3 NRU test method
173 protocol based on the 3T3 Cytotoxicity Test (INVITTOX Protocol No. 46; available from the
174 FRAME-sponsored INVITTOX database [http://embryo.ib.amwaw.edu.pl/invittox/]), which
175 in turn was based on the Borenfreund and Puerner (1985) protocol, as elaborated on in
176 Spielmann et al. (1991, 1996). This protocol was updated based on experience obtained
177 during the validation of the 3T3 NRU Phototoxicity Test (INVITTOX Protocol No. 78; also
178 available at the FRAME INVITTOX database). The RC millimole regression for prediction
179 of acute oral rat and mouse toxicity (Halle 1998, 2003; Spielmann et al. 1999) was included
180 as the prediction model (see **Section 1.1.2**).
181
182 The NHK NRU protocol provided in the *Guidance Document* was based on the protocol used
183 by IIVS, which was based on a NRU protocol of Borenfreund and Puerner (1984) and a rat

184 epidermal keratinocytes protocol (Heimann and Rice 1983). Formulations for the media and
185 solutions, and general NHK cell culture techniques, correspond to Clonetics® products from
186 the CAMBREX Corporation.
187
188 The protocol components for the 3T3 and NHK NRU test methods used in this validation
189 study are similar (see **Figure 2-1**). The nature of the NRU response is described in **Section**
190 **1.3.1**. **Figure 2-1** provides an overview to the major steps for performance of the *in vitro*
191 NRU test methods. The following procedures are common to both cell types:
192 • Preparation of substances and the PC
193 • Cell culture environmental conditions
194 • Determination of test substance solubility
195 • 96-well plate configuration for testing samples
196 • Range finder and definitive tests
197 • Microscopic evaluation of cell cultures for toxicity based on morphological
198 alterations
199 • Procedures for measurement of NRU
200 • Data analysis procedures
201
202 The main protocol differences between the two cell lines are:
203 • The conditions of propagation of the cells in culture (e.g., time needed to
204 reach appropriate confluence)
205 • The growth media components
206 • The volumes of substances applied to the 96-well plates
207 • The number of cell divisions undergone by each cell line during exposure to a
208 test substance
209
210

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 37ºC (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 37ºC 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 37ºC

⇓

(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)[1]

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

---

[1] 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.

257 2.2.1 The 3T3 NRU Test Method

258 2.2.1.1 *Initiating and Subculturing 3T3 Cells*

259 Each laboratory initially prepared a large pool of 3T3 cells (described further in **Section**
260 **2.3.1.1**), cryogenically preserved multiple ampules of these cells in liquid nitrogen, and
261 periodically removed an ampule when needed. Although the NRU protocols used for each
262 study phase provided cell culture density guidelines for subculturing the cells, each
263 laboratory refined the final seeding density to achieve optimal growth.
264
265 Cryopreserved 3T3 cells were thawed, resuspended in a culture medium containing
266 Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with non-heat-
267 inactivated 10% newborn calf serum (NCS), transferred into tissue culture flasks (25 or 75 -
268 80 cm^2), and incubated at 37°C ± 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air. When
269 cells reached 50 to 80% confluence (as estimated from a visual inspection of cell density),
270 they were removed from the flask by trypsinization. A single-cell suspension was added to
271 new flasks for propagation and the cells were passaged/subcultured at least two times[2] before
272 seeding into 96-well plates for testing. This study did not evaluate the potential effects that
273 cell passage number may have on the performance of the 3T3 NRU test method.

274 2.2.1.2 *Preparation of Cells for 96-well Plate Assays*

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

[2] 3T3 cells could be 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 proposing 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.

283 that approximately a 50% confluent monolayer was present at the time of substance
284 application.

285 2.2.1.3 *Reference Substance Application*

286 After the appropriate incubation period to achieve a half-confluent monolayer, the medium
287 was removed and 50 µL of culture medium with 10% NCS were added to each well. Then,
288 50 µL treatment medium containing the appropriate substance concentrations were added for
289 a final concentration of 5% NCS. The cells were then incubated at 37ºC for 48 ± 0.5 hours.
290 At the end of the incubation period, the cells were microscopically evaluated for changes in
291 morphology and their appearance was documented (as per Visual Observation Codes in the
292 protocol) prior to measurement of NRU.

293 2.2.2 <u>The NHK NRU Test Method</u>

294 2.2.2.1 *Initiating and Subculturing NHK Cells*

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

310 2.2.2.2 *Preparation of Cells for 96-well Plate Assays*

311 After subculturing the cells, 125 µL of the cell suspension ($2.0 – 2.5x10^3$ cells/well) were
312 placed in the appropriate wells and 125 µL of cell-free culture medium were dispensed into

313 the peripheral wells (blanks). One plate per reference substance was prepared. The cells were
314 incubated at 37ºC for 48 - 72 hours and checked to be sure that cultures were at 20 to 50%
315 confluence at the start of exposure to the reference substance.

316 2.2.2.3 *Reference Substance Application*

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

323 2.2.3 Measurement of NRU in the 3T3 and NHK Test Methods

324 The treatment medium was removed from the 96-well plates, the cells were rinsed with
325 phosphate buffered saline (PBS), and 250 µL NR dye medium was added to the wells (25 µg
326 NR/mL for 3T3 cells; 33 µg NR/mL for NHK cells). The plates were then incubated (37°C ±
327 1°C, 90% ± 5% humidity, and 5.0% ± 1% CO_2/air) for three hours. After incubation, the NR
328 medium was removed, the cells were rinsed with PBS, and 100 µL of the desorb solution
329 were applied. The plates were shaken on a microtiter plate shaker for 20 to 45 minutes to
330 extract NR from the cells and to form a homogeneous solution. The optical density (OD) of
331 the resulting colored solution was measured (within 60 minutes of adding the desorb
332 solution) at 540 nm ± 10 nm (OD_{540}) in a spectrophotometric microtiter plate reader, using
333 the blank wells as reference. Data from the plate reader were transferred to a Microsoft®
334 EXCEL® (Microsoft Corporation, Redmond, WA, USA) spreadsheet template (hereafter
335 know as EXCEL® template) designed by the SMT and the testing laboratories for statistical
336 analyses.

337 **2.3 Descriptions and Rationales of the 3T3 and NHK NRU Test Methods**

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

343 2.3.1 Materials, Equipment, and Supplies

344 2.3.1.1 *3T3 Cells*

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

357 2.3.1.2 *NHK Cells*

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

367 2.3.1.3 *Tissue Culture Materials and Supplies*

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

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

381 2.3.1.4 *Cell Culture Materials*

382 Laboratory items needed include the following:

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

393 2.3.1.5 *Equipment*

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

396 • Incubator (37ºC ± 1ºC, 90% ± 5% humidity, 5.0% ± 1% CO_2/air)
397 • Laminar flow clean bench/cabinet (standard: "biological hazard")
398 • Water bath (37ºC ± 1ºC)
399 • Inverted phase contrast microscope (with 10X to 40X objectives)
400 • Centrifuge (capable of 220 x g)
401 • Laboratory balance (capable of measuring to 10 mg)
402 • Spectrophotometer for reading 96-well plates (i.e., microtiter plate reader)
403 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 -70ºC)
- Cryostorage container (and liquid nitrogen supply)
- Magnetic stirrer
- Antistatic bar ionizer
- Personal computer
- Osmometer
- pH meter

2.3.1.6 *Culture Medium*

<u>For 3T3 Cells</u>

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.

435 For NHK Cells

436 Although the contents of the NHK basal culture medium are proprietary, the formulation is
437 based on a commercially available, non-proprietary basal medium (MCDB 153 medium
438 formulation [Tsao et al. 1982]; e.g., MCDB 153 medium - SIGMA-ALDRICH product
439 number #M 7403 http://www.sigmaaldrich.com/sigma/datasheet/m7403dat.pdf). The
440 laboratories recommended this medium for use with the CAMBREX Clonetics® NHK cells
441 because they all had access to this supplier. Other media are acceptable for NHK NRU
442 testing if the performance standards prescribed in the media prequalification protocol are met
443 (see **Appendix B4** and **Section 2.6.3.5**).
444
445 The serum-free culture medium used for NHK cells was Clonetics® keratinocyte basal
446 medium (KBM®) supplemented with KBM® SingleQuots® (epidermal growth factor, insulin,
447 hydrocortisone, antimicrobial agents, bovine pituitary extract [BPE]) and Calcium SingleQuots®
448 (calcium) [all from CAMBREX Corporation] to make keratinocyte complete growth medium.
449 Although the keratinocyte complete growth medium is a defined serum-free medium, it
450 contains BPE collected from bovine pituitary glands. BPE contains growth factors and
451 hormones, and is added to serum-free medium as a mitogenic supplement. Variability in the
452 composition of the BPE could be a factor in cell growth kinetics. However, it is suggested
453 that the undefined BPE components could be replaced with defined growth supplements,
454 such as insulin, epidermal growth factor, and fibroblast growth factor, without adversely
455 affecting the cellular proliferation rates and general physiology of human keratinocytes (Life
456 Technologies, Inc. 1997).

457 2.3.2 Reference Substance Concentrations/Dose Selection

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

463 2.3.2.1 *Range Finder Test*

464 A range finder 3T3 or NHK NRU test was performed to determine the concentrations of a
465 reference substance to be used for the definitive (concentration-response) test (see **Section**

466 **2.3.2.2**). The range finder test used eight concentrations of the reference substance prepared
467 by diluting the stock solution using log intervals to cover a large concentration range (e.g.,
468 1:10, 1:100, 1:1000, etc.; up to eight orders of magnitude). The highest concentrations
469 applied to the cells were 10 mg/mL for substances dissolved in culture medium and 1 mg/mL
470 in medium for substances dissolved in DMSO, unless precluded by solubility. ETOH was not
471 used as a solvent for any of the substances in the validation study (see **Sections 2.8, 2.9,** and
472 **2.10**).
473
474 If the range finder test did not produce cytotoxicity, then a second range finder test was
475 conducted at higher concentrations (e.g., the highest concentration would be >10 mg/mL if in
476 medium, >1 mg/mL if in DMSO) unless precluded by solubility. If the substance being tested
477 was insoluble or poorly soluble, then more stringent solubility procedures were employed to
478 increase the stock concentration (to the maximum concentration specified in **Appendices B1**
479 and **B2**). If the range finder test produced a biphasic dose-response curve[3] for NR uptake, the
480 concentrations selected for the definitive tests covered the response range that included the
481 lowest concentration that reduced viability by 50% (see **Section 2.6.3.2**).

482 2.3.2.2 *Definitive Test*

483 The concentration-response determination is referred to as the definitive test because it is
484 used to determine the IC_{50} value of the substance being tested. The concentration closest to
485 the calculated IC_{50} value in the range finder test served as the midpoint of the eight
486 concentrations tested in the definitive test. In the absence of other information (e.g.,
487 knowledge of the slope of the toxic response), the recommended dilution factor was 1.47
488 ($\sqrt[6]{10}$), which divides a log interval into six equidistant steps (e.g., 10, 14.7, 21.5, 31.6, 46.4,
489 68.1, 100). The *Guidance Document* considered a progression factor of 1.21 ($\sqrt[12]{10}$) to be the
490 smallest factor practically achievable, and this was the lowest required concentration interval.
491 The PC was tested similarly to the reference substances in the definitive test and the same
492 recommended dilution factors were used (dilution factor at the discretion of the Study
493 Director).
494

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

495 A definitive test was considered successful if it met all of the test acceptance criteria outlined
496 in the NRU protocols. Definitive tests were repeated as per the protocols if the test failed to
497 meet all of the test acceptance criteria. **Section 2.5** addresses the basis for replicate testing.
498
499 If minimal or no cytotoxicity was observed in the range finder test, the maximum
500 concentration for the definitive test was determined as follows:

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

514 2.3.3 NRU Endpoints Measured

515 2.3.3.1 *NRU and Measurement*

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

526 = 100% viability). Cell viability in other test wells was computed in reference to the mean
527 VC OD value (i.e., [well OD/mean VC OD] x 100 = % viability).

528 2.3.3.2 *Determination of IC_{50}, IC_{20}, and IC_{80} Values*

529 IC_{50} values were determined from the concentration-response curve using a Hill function,
530 which is a four parameter logistic mathematical model relating the concentration of a
531 substance to the response (typically following a sigmoidal shape). Modifications to the Hill
532 function used in later phases of the study are described in **Section 2.6.3**.
533
534 Data from the EXCEL® template were transferred to a template designed by the SMT for
535 GraphPad PRISM® 3.0, a commercially available statistical software (GraphPad Software,
536 Inc., San Diego, CA, USA – hereafter known as PRISM® template). The PRISM® template
537 calculated the IC_{50}, IC_{20}, and IC_{80} concentrations, reported as μg/mL of reference substance
538 in solution. IC_{20} and IC_{80} data were collected for potential use in designing a prediction
539 model for estimating human lethal blood concentrations.

540 2.3.4 Duration of Reference Substance Exposure

541 The SMT and laboratory representatives reevaluated the reference substance exposure
542 duration recommended in the *Guidance Document* (ICCVAM 2001b) before initiating the
543 study. The *Guidance Document* recommended an exposure of 24 hours for the 3T3 cells and
544 48 hours for the NHK cells. However, Riddell et al. (1986) showed large differences in
545 cytotoxicity for 3T3 cells in response to some chemicals, depending on whether the exposure
546 duration was 24 or 72 hours. Although the toxicity induced by substances that damage, for
547 example, cell membranes is likely to be observed in a relatively shorter time, the toxic effects
548 of substances that interfere with cell functions/processes specifically relating to DNA
549 replication (e.g., protein and nucleic acid synthesis) and cell division (e.g., mitotic spindle
550 formation) are more pronounced after longer exposure periods. This occurs because cells are
551 affected only at certain phases of the cell cycle.
552
553 IIVS conducted studies to evaluate the effect of exposure durations of 24, 48, and 72 hours
554 and of 48 and 72 hours on the sensitivity of 3T3 cells and NHK, respectively, to six
555 chemicals selected from the list in Riddell (1986). Because the closest fit to the RC millimole
556 regression occurred when a 48-hour exposure duration was used, this exposure duration was

557 selected for use with both cell types in the validation study (Curren et al. 2003) (see
558 **Appendix E**).

559 2.3.5 Known Limits of Use

560 2.3.5.1 *Solubility/Precipitation/Volatility*

561 *In vitro* test methods cannot be used for substances that cannot be dissolved in media,
562 DMSO, or ETOH at a sufficiently high concentration to induce cytotoxicity in excess of
563 50%. Also, chemicals that are unstable or exothermic in water cannot be adequately tested
564 with these *in vitro* test methods (as well as *in vivo* methods).
565
566 Precipitation of a test substance in the dosing solution or in the culture medium after the
567 substance to be tested has been added can affect the concentration-response and thus reduce
568 the accuracy of the calculated IC_{50}. Some reference substances used in the validation study
569 had precipitates in their medium/DMSO 2X concentrations prior to dilution for application to
570 the test wells. Precipitates were also observed for some substances in a number of test wells
571 after addition of the media/DMSO 1X solutions (see **Section 5.8** and **Table 5-10**) to the
572 cultures and/or at the end of the exposure period.
573
574 Volatility was detected for a number of reference substances during the range finder tests by
575 observance of cross contamination (i.e., high cytotoxicity) in VC wells. Plate sealers were
576 used during the definitive tests to control volatility (see **Section 2.6.3** – *Testing Volatile*
577 *Reference Substances*), and could be used during the range finder tests if the Study Director
578 suspected that the reference substance might be volatile. The use of plate sealers required
579 additional laboratory training, and some volatile substances were difficult to test even with
580 the use of plate sealers. Furthermore, some test substances (e.g., organic solvents) may react
581 chemically with the plastic in the sealers.

582 2.3.5.2 *Biokinetic Determinations*

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

585 "Results obtained from *in vitro* studies in general are often not directly applicable to the
586 *in vivo* situation. One of the most obvious differences between the situation *in vitro* and

587 *in vivo* is the absence of processes regarding absorption, distribution, metabolism and
588 excretion (i.e., biokinetics) that govern the exposure of the target tissue in the intact
589 organism. The concentrations to which *in vitro* systems are exposed may not correspond
590 to the actual situation at the target tissue after *in vivo* exposure. In addition, the
591 occurrence of metabolic activation and/or saturation of specific metabolic pathways or
592 absorption and elimination mechanisms may also become relevant for the toxicity of a
593 compound *in vivo*. This may lead to misinterpretation of *in vitro* data if such information
594 is not taken into account. Therefore, predictive studies on biological activity of
595 compounds require the integration of data on the mechanisms of action with data on
596 biokinetic behavior."
597
598 The 3T3 and NHK NRU test methods do not account for biokinetics.

599 2.3.5.3 *Organ-Specific Toxicity*

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

607 2.3.5.4 *The Role of Cytotoxicity Tests in an In Vitro Battery Approach for Possible*
608 *Replacement of In Vivo Acute Toxicity Testing*

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

612 Step 1: Perform a physico-chemical characterization and biokinetic modeling.
613 Step 2: Evaluate basal cytotoxicity using, for example, the 3T3 or NHK NRU test
614 methods.
615 Step 3: Evaluate the potential that metabolism will mediate the basal cytotoxicity
616 effect.
617 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 Reference 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.

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

659 2.3.7.2 *Vehicle Control*

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

665 2.3.7.3 *Negative Control*

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

669 2.3.8 Acceptable Ranges of Control Responses

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

677

678 **Table 2-1 Measured VC OD_{540} Values[1] and Target**
679

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

680 Abbreviations: VC=Vehicle control; OD_{540}=Optical density at 540 nM; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral Red
681 Uptake; NHK=Normal human keratinocytes; ECBC=Edgewood Chemical Biological Center; FAL=Fund for Replacement
682 of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
683 [1]Lowest to highest OD values for tests that meet test acceptance criteria.
684 [2]Ranges used for all laboratories. Ranges for Phases Ia and Ib were test acceptance criteria. Ranges for Phases II and III
685 were used as target ranges, rather than as test acceptance criteria.
686

687 In Phase III, 99.5% (914/919) of all 3T3 mean VC OD values and 97% (913/944) of all NHK
688 mean VC OD values were within the target ranges. Most OD values outside the ranges were
689 from range finding tests and were usually the result of volatile reference substances affecting
690 the VC cells adjacent to the highest reference substance concentration wells.
691

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

696 • Some tests in Phases Ia and Ib exhibited NR crystals that caused higher OD
697 readings.
698 • Cell seeding densities were revised from 2.5 x 10^3 cells/well to a range of 2.0
699 – 3.0 x 10^3 cells/well.
700

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

702 • The minimum percent confluence of cells necessary before the reference
703 substance could be applied was reduced from 30% to 20% confluence.

704 • Cell growth was reduced in some tests in the later study phases as a result of
705 medium and supplement issues (e.g., certain lots of basal medium and
706 medium supplements for NHK cells did not provide optimum growth
707 conditions for the keratinocytes).

708 2.3.8.1 *Vehicle Controls as a Quality Control Tool*

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

717 2.3.9 Nature of Experimental Data Collected

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

722 2.3.9.1 *NRU OD Measurements*

723 At the conclusion of the NRU desorb step, the OD of the resulting colored solution in each
724 well of the 96-well plates was measured at 540 ± 10 nm in a spectrophotometric microtiter
725 plate reader. Each laboratory followed its in-house Standard Operating Procedure (SOP) for
726 use of the microplate readers. These SOPs included instructions for operation and calibration
727 of the instruments. Critical specifications such as alignment, accuracy, reproducibility, and
728 linearity were included as standard parameters for review and routine calibration. Raw OD
729 data from the plate reader was electronically transferred to the EXCEL® template. The
730 template converted the raw data from each treatment well (six wells/reference substance
731 concentration) to derived data by subtracting the mean blank OD value (two blank
732 wells/reference substance concentration) from each reference substance well OD. There were
733 12 VC wells and 20 associated blank wells. The corrected VC OD values were used to

734 calculate the mean VC OD, which was then used to calculate relative viability (% of mean
735 VC OD) in each test well for the reference substance or PC. The percent viability values
736 were then transferred to the PRISM® template for calculation of the IC_{20}, IC_{50}, and IC_{80}
737 values.

738 2.3.9.2 *Information and Data Collected*

739 Originals of the raw data (i.e., the Study Workbook and computer printouts of absorbance
740 readings from the plate reader) and copies of other raw data, such as instrument logs, were
741 collected and archived under the direction of the Study Director according to Good
742 Laboratory Practice (GLP)-compliant procedures.
743
744 The Study Director/technicians entered the following information into the EXCEL® template:

745 • Testing identification for: test facility, chemical code, study number, 96-well
746 plate number, experiment number
747 • Reference substance preparation: solvent used, solvent concentration in
748 dosing solutions, highest stock concentration, dilution factor, pH of 2X dosing
749 solutions, medium clarity/color, presence/absence of precipitate in 2X
750 solutions, PC concentration range
751 • Cell line/type: cell supplier, lot number, cryopreserved passage number,
752 passage number in assay
753 • Cell culture conditions: medium, supplements, suppliers and lot numbers,
754 serum concentrations
755 • Timeline: dates of cell seeding, dose application, OD_{540} determination
756 • Raw data: OD values from each well from the microtiter plate reader
757 • Test results: mean corrected OD_{540} value, Hill function R^2 value, logs of IC_{20},
758 IC_{50}, and IC_{80} (PRISM® template presents data as logs of the IC_x; EXCEL®
759 converts values to μg/mL)
760 • Test acceptance criteria: acceptable number of values on each side of the IC_{50}
761 (i.e., number of points >0 and ≤50% viability, and >50 and <100% viability),
762 acceptable percent difference for the VCs, acceptable Hill function R^2 value
763 (coefficient of determination) for the PC, and calculated IC_{50} concentration for
764 the PC

765 • Visual observations: protocol codes for cell culture conditions for all reference
766 substance concentrations (i.e., relative level of cell cytotoxicity, cell
767 morphology, presence of precipitate)

768 2.3.10 Data Storage Media

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

773 2.3.11 Measures of Variability

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

775 1) Each plate contained VCs on each end of the plate (columns 2 and 11) (see
776 **Figure 1** in **Appendix B1** for plate map). The difference between the mean
777 NRU OD for each VC column and mean of the pooled VC wells was used as a
778 test acceptance criterion. The Study Director rejected the test if the difference
779 was greater than 15%, which indicated cross-contamination from a volatile
780 substance or possible cell seeding errors.
781 2) A mean relative viability was determined for each concentration of the
782 substance tested along with the SD and coefficient of variation
783 (%CV=SD/mean x 100).
784 3) Macros were included in the EXCEL® template to perform an outlier test
785 (Dixon and Massey 1981) on the data for the six replicate wells for each
786 concentration. Outliers (i.e., individual well values that exceeded the 99%
787 confidence interval [CI] for the replicate wells) were highlighted and could be
788 excluded from the resulting analysis to improve curve fit. The Study Director
789 made the decision as to whether or not to remove outliers and provided a
790 justification for the decision.
791

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

793 • Each set of assays for reference substances included a PC plate. If the SLS PC
794 test did not meet test acceptance criteria, then the tests for the associated
795 reference substances were rejected. The SMT recommended testing a

796 manageable number of definitive test plates (e.g., 4 to 6) with each PC to limit
797 the number of definitive NRU tests rejected for PC failure. In this validation
798 study, 4.2% of all definitive tests performed were rejected because the PC
799 failed (i.e., the PC IC_{50} was outside the acceptable confidence limits).
800 • SDs and CVs were determined for mean IC_{50} values from replicate tests.
801 Replicate testing included three definitive tests for each reference substance,
802 each performed on a different day.

803 2.3.12 Methods for Analyzing NRU Data

804 Relative cell viability for each reference substance concentration was calculated using the
805 ODs of the six replicate values (minimum of four acceptable replicate wells) per test
806 concentration. Relative cell viability was expressed as a percentage of the mean VC OD.
807 Absolute OD data from the microtiter plate reader was transferred to the EXCEL® template
808 for performance of these calculations. Where possible, the concentration range (eight
809 concentrations) tested for each reference substance ranged from no effect to 100% toxicity.
810
811 The IC_{20}, IC_{50}, and IC_{80} values were determined from the concentration-response curve using
812 the PRISM® template and applying a Hill function to the % viability data. The IC_{20} and IC_{80}
813 values were calculated for potential use in the development of a human prediction model
814 (reported elsewhere).

815 2.3.13 Decision Criteria for Classification of Reference Substances

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

823 2.3.14 Information and Data Included in the Test Report

824 *Test and Control Substances*

825 With the exception of the PC, the laboratories tested coded substances and had

826 minimal information about the test substances' properties (see **Section 3.4** for the
827 reference substance information provided to the laboratories). The following
828 describes the test and test substance information that should be included in an NRU
829 test method report.

830 • Chemical name(s) and synonyms, if known
831 • The CASRN, if known
832 • Formula weight, if known
833 • Purity and composition of the substance or preparation (in percentage[s] by
834 weight)
835 • Physicochemical properties (e.g., physical state, volatility, pH, stability,
836 chemical class, water solubility)
837 • Solubilization of the test/control substances (e.g., vortexing, sonication,
838 warming, grinding) prior to testing, if applicable

839 *Information Concerning the Sponsor and the Test Facility*

840 • Name and address of the sponsor, test facilities, study director, and
841 participating laboratory technicians
842 • Justification of the test method and specific protocol used

843 *Test Method Integrity*

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

846 *Criteria for an Acceptable Test*

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

853 *Test Conditions*

854 • Experiment start and completion dates
855 • Details of test procedures used
856 • Test concentration(s) used and how they were derived

857 • Cell type used and source of cells
858 • Description of modifications made to the test procedure
859 • Reference to historical data of the test model (e.g., solvent and PCs)
860 • Description of the evaluation criteria used
861 *Results*
862 • Tabulation of data from individual test samples (e.g., IC_{50} values for the
863 reference substance and the PC, absolute and derived OD readings, reported in
864 tabular form, including data from replicate repeat experiments as appropriate,
865 and the means and standard deviations for each experiment)
866 *Description of Other Effects Observed*
867 • Cell morphology, precipitate, NR crystals, etc.
868 *Discussion of the Results*
869 *Conclusion*
870 *Quality Assurance (QA) Statement for GLP-Compliant Studies*
871 • A statement describing all inspections and other QA activities during the
872 study, and the dates results were reported to the Study Director. This
873 statement will also serve to confirm that the final report reflects the raw data.
874
875 During the validation study, the GLP-compliant laboratories, IIVS and ECBC, followed
876 additional reporting requirements provided in the relevant GLP guidelines (e.g., OECD 1998;
877 EPA 2003a, b; FDA 2003).
878
879 The SMT and laboratories developed standard forms for data collection (i.e., EXCEL® and
880 PRISM® templates). The solubility test form was derived from a standard form provided by
881 IIVS. The EXCEL® template was an adaptation of a template format presented in the
882 *Guidance Document* (ICCVAM 2001b).

883 **2.4 Proprietary Components of the *In Vitro* NRU Test Methods**

884 The only proprietary components used in these test methods are the NHK cells and the NHK
885 basal culture medium obtained from CAMBREX Clonetics®. All other components are
886 readily available through various scientific product suppliers.
887

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**):

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

922 **2.6 Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols**

923 2.6.1 Phase Ia: Laboratory Evaluation Phase

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

925 2.6.1.1 *NR Dye Crystals*

926 NR dye crystals formed in the 96-well test plates when used at 50 μg/mL (OD values
927 measured in the blanks increased from ~ 0.05 to 0.10) in both NRU test procedures.
928 Troubleshooting efforts included incubating the NR medium overnight; centrifuging and
929 filtering the NR medium prior to application to the 96-well plates; and reducing the
930 concentration of NR dye. The laboratories performed tests using a reduced NR concentration
931 of 33 μg/mL. Since there were no quantitative differences in results between tests with 50
932 μg/mL and tests with 33 μg/mL NR, the SMT accepted tests with both concentrations.
933
934 *Protocol Revision:* The NR dye concentration was reduced to 33 μg/mL for both cell types in
935 subsequent test Phases.

936 2.6.1.2 *3T3 Cell Growth*

937 The growth rate of 3T3 cells (as determined by monolayer confluence) was slower than
938 expected. As a result, the cells required more time in culture to obtain the proper density after
939 seeding.
940
941 *Protocol Revision:* The 3T3 cells must be passaged 2-3 times after thawing before being used
942 for the test. The protocol also emphasized attainment of the appropriate percentage of cell
943 confluence (not more than 50% for 3T3 cells) required at the time the cells were exposed to
944 the reference substance, rather than using the time in culture as the guide.

945 2.6.1.3 *NHK Cell Growth*

946 The NHK cells had an additional growth problem that manifested as a ring of dead/dying
947 cells around the center of the wells. Troubleshooting efforts included evaluating various

948 brands of 96-well plates (laboratories were not required to use the same brand of plates) and
949 eliminating the change of medium prior to reference substance treatment. All laboratories
950 participated in evaluating the effect of changing (i.e., refeeding) or not changing (i.e., no
951 refeeding) the medium by performing a small study with the PC (SLS). Tests were
952 performed: 1) after refeeding the cells with fresh medium, and 2) by adding SLS to the
953 medium already on the cells. Control ODs were generally higher in the tests in which the
954 medium was not replenished, but sensitivity to SLS was generally unchanged (see **Table 2-**
955 **2**). FAL was experiencing difficulties in NHK cell growth at this stage of the study which
956 may account for the difference in the refeeding and no refeeding SLS IC_{50} values. The SMT
957 accepted tests with refeeding and those without refeeding (for Phase Ia) as long as they met
958 the test acceptance criteria.
959

960 **Table 2-2 Refeeding/No Refeeding Data for the NHK NRU Test Method**
961

	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

962 Abbreviations: NHK=Normal human keratinocyte; NRU=Neutral Red Uptake; ECBC=Edgewood Chemical
963 Biological Center; IIVS=Institute for *In Vitro* Sciences; FAL=Fund for the Replacement of Animals in Medical
964 Experiments Alternatives Laboratory; VC=Vehicle control; OD=Optical density; SLS=Sodium lauryl sulfate.
965 Note: OD values for SLS IC_{50} were extrapolated from the concentration-response curve data
966 [1]Mean ± standard deviation (uncorrected for blank absorbance
967

968 IIVS presented detailed information on the ring of dead cells issue (Raabe 2004). The
969 laboratory showed that the ring of cell death coincided with the formation of a meniscus
970 resulting from the residual medium left in the well after removal of the spent medium. The
971 problem was resolved by eliminating the removal of medium before applying test chemical
972 rather than requiring a standard brand of 96-well plates.
973

974 *Protocol Revision:* Step 2 of the NHK NRU test method was eliminated (change of medium
975 prior to addition of reference substance). The volume of medium (with cells) was changed
976 from 250 µL/well to 125 µL/well.
977
978 FAL, in contrast to the other two laboratories, used 80-cm^2 culture flasks for culturing the
979 thawed cells from the ampules of cryogenically-preserved pool of cells and encountered
980 difficulties in obtaining a satisfactory number of adhering NHKs.
981
982 *Protocol Revision (FAL only):* Culture flasks were coated with fibronectin-collagen to
983 promote cell adherence.

984 2.6.1.4 *Vehicle Control OD Limits*

985 In Phase I, the acceptable range of VC OD values designated in the protocols (0.3 ≤ OD
986 ≤1.1) were frequently unattainable in both test methods. Despite this, the Study Directors
987 reported that the cells were adequately responsive. The SMT withdrew the VC OD limits as a
988 test acceptance criterion.
989
990 *Protocol Revision for Phase Ib:* OD ranges were provided as guidelines for each cell type
991 based on OD data from all laboratories, a review of the concentration-response data, and the
992 ability of each test to pass the other test acceptance criteria. Each laboratory developed its
993 own VC OD acceptability range based on its historical data.

994 2.6.1.5 *Precipitate Formation*

995 During solubility testing, it was observed that some substances, when tested at the same
996 concentrations, precipitated in the 3T3 medium but not in the NHK medium. When a liquid
997 reference substance (i.e., 2-propanol) produced this effect, the precipitate was attributed to
998 the protein in the serum in the 3T3 medium rather than insolubility.
999
1000 *Protocol Revision:* The reference substances were dissolved in 3T3 medium without NCS to
1001 make the 2X solutions. The dissolved 2X reference substance was added to medium
1002 containing 10% NCS to reach the final 5% NCS and 1X reference substance concentrations.
1003

1004 2.6.1.6 *Dilution Factor*

1005 After a range finder test was performed, the definitive tests were to be performed using a
1006 $\sqrt[6]{10} = 1.47$ dilution scheme centered on the IC_{50} that was calculated from the range finder.
1007 In Phase Ia, the Study Directors, for various reasons related to the specific substance being
1008 tested, sometimes deviated from this requirement and used other dilution factors. The SMT
1009 agreed that the dilution factor requirements should be modified to allow more flexibility in
1010 setting up tests. The SMT accepted data generated using dilution factors other than the
1011 recommended 1.47 for definitive tests if all other test acceptance criteria were met. The use
1012 of smaller dilution factors generally increased the number of concentrations in the 10% to
1013 90% viability range, which improved the precision of the IC_{50} calculation.
1014
1015 *Protocol Revision:* The $\sqrt[6]{10} = 1.47$ dilution scheme was a suggested starting range, rather
1016 than a specific test acceptance criterion in subsequent test Phases.

1017 2.6.1.7 *Test Acceptance Criteria*

1018 The test acceptance criteria at the beginning of Phase Ia were:

1019 • The IC_{50} for SLS had to be within the 95% CI of the historical PC mean
1020 established by the Test Facility (*rescinded after commencement of Phase Ia*)
1021 • The OD_{540} of the VCs (with blank subtracted) had to be ≥ 0.3 and ≤ 1.1
1022 (*rescinded after commencement of Phase Ia*)
1023 • Mean OD values of the left and right VCs (columns 2 and 11 in the 96-well
1024 test plate) must not differ by more than 15% from the mean of all VC OD
1025 values
1026 • At least two cytotoxicity values, one on either side of the IC_{50} but between
1027 10% and 90% viability, needed to be present (*added after commencement of*
1028 *Phase Ia*)
1029 • The Hill function curve fits ($R^2 > 0.9$ <u>or</u> $0.8 < R^2 < 0.9$) were evaluated on a
1030 case by case basis for acceptability by the SMT (*added after commencement*
1031 *of Phase Ia*).

1032 2.6.2 <u>Phase Ib: Laboratory Evaluation Phase</u>

1033 All protocol revisions developed during Phase Ia were implemented during Phase Ib unless
1034 otherwise stated.

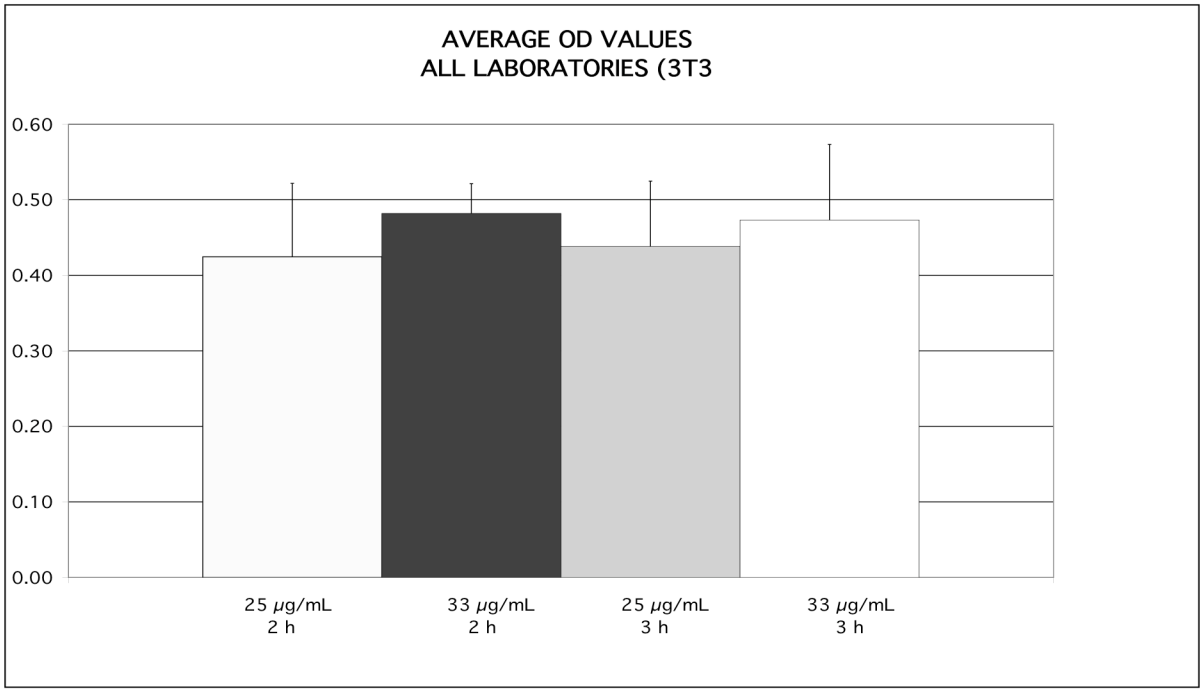
1035 2.6.2.1 *NR Crystal Formation*

1036 FAL and ECBC routinely observed NR crystals forming in the 96-well test plates in the 3T3
1037 NRU tests when 33 µg/mL NR was used. All laboratories tested 25 and 33 µg/mL NR
1038 concentrations and 2- and 3-hour NR incubation periods to determine which NR
1039 concentration and incubation period would provide optimal NRU measurements without
1040 crystal formation. In addition to determining whether NRU had reached a plateau at these
1041 concentrations and incubation time, the laboratories also determined whether the response to
1042 SLS differed under these conditions. Crystals were observed only at 33 µg/mL NR when
1043 present for three hours. **Figure 2-2** shows that the average OD results were similar for all NR
1044 concentrations and incubation periods tested. **Figure 2-3** shows that the SLS IC_{50} values
1045 were equivalent at the different NR concentrations and incubation periods. To minimize
1046 changes to the 3T3 protocol, the NRU concentration was lowered from 33 to 25 µg/mL,
1047 while the NR incubation period was maintained at three hours. The NR concentration and the
1048 incubation period for the NHK NRU test method remained at 33 µg/mL and three hours,
1049 respectively.
1050
1051 *Protocol Revision for Phase II*: The NR concentration for the 3T3 NRU test method was
1052 reduced to 25 µg/mL for the three-hour incubation period. Revised methods for preparation
1053 of the NR dye solution included filtration of the solution, maintenance of the solution at 37ºC
1054 prior to application to the cells, and application of the NR solution to the cells within 15
1055 minutes after removing it from 37ºC. Also, cells were observed during the NR incubation
1056 period to monitor possible crystal formation.

1057 2.6.2.2 *Heating of Reference Substance Solutions*

1058 The laboratories had difficulty solubilizing arsenic trioxide, one of the reference substances
1059 used in Phase Ib. Heating and mechanical applications for increasing the laboratory's ability
1060 to solubilize substances into culture medium were reviewed and revised.
1061
1062 *Protocol Revision for Phase II*: The duration range for heating a stock solution at 37ºC (if
1063 heating is needed) was increased from 5 to 10 minutes to 5 to 60 minutes.
1064
1065

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



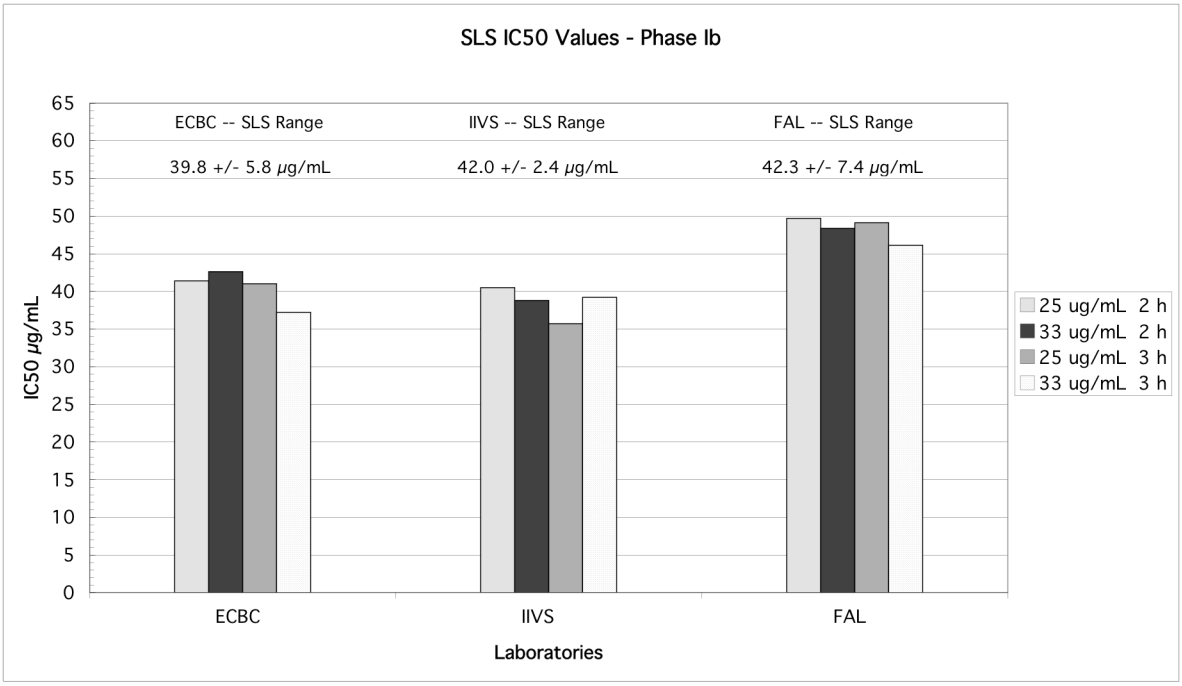
Abbreviations: OD=Optical density; NR=Neutral Red; SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral Red Uptake; h=Hours.
Note: Error bars are one standard deviation.

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).**

1083 **Figure 2-3 SLS IC_{50} Values for Each NR Concentration and Incubation Duration**
1084 **(3T3 NRU)**
1085



1086

1087 Abbreviations: SLS=Sodium lauryl sulfate; IC_{50}=Test substance concentration that reduces cell viability by
1088 50%; NR=Neutral red; 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; ECBC=Edgewood Chemical
1089 Biological Center; IIVS=Institute for *In Vitro* Sciences; FAL=Fund for the Replacement of Animals in
1090 Medical Experiments Alternatives Laboratory.
1091 Note: SLS range is mean IC_{50} value ± one standard deviation.
1092

1093 2.6.2.4 *Correction of Reference Substance OD Values*

1094 Each reference substance concentration was applied to six treatment wells and to two cell-
1095 free wells (i.e., blank wells) used to generate the background OD_{540} values to adjust for
1096 potential interference with the NR dye. The mean blank well OD (absolute OD) for each
1097 reference concentration was subtracted from the reference substance concentration ODs to
1098 provide the corrected OD for each replicate well.

1099 2.6.2.5 *Laboratory Error Rates*

1100 The SMT determined the Phase 1b error rates (number of tests with errors/total number of
1101 tests conducted) for each laboratory (**Table 2-3**) and compiled a list of the types of errors
1102 encountered. The vast majority of errors were transcriptional and typographical errors in the
1103 data sheets provided to the SMT.
1104

1105 **Table 2-3 Error Rates[1] in Phase Ib by Laboratory and Test**
1106

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%)

1107 Abbreviations: NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes;
1108 ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals In Medical Experiments
1109 Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences
1110 [1]Number of tests with errors/total number of tests (some data files had more than one error)
1111 Note: Most errors were transcriptional and typographical and not technical.

1112

1113 2.6.2.6 *Resultant Protocol Changes for Phase II*

1114 Following the completion of Phases Ia and Ib, IIVS sponsored a weeklong laboratory training
1115 exercise for all the laboratories at their facility to help standardize the level of training among
1116 the technical staff and to identify any further 3T3 and NHK NRU protocol revisions that
1117 might be needed. Protocol revisions made because of this exercise included:

1118 • Multi-channel repeater pipettes can be used for dispensing cells into the 96-
1119 well plates and dispensing plate rinse solutions, NR medium, and desorb
1120 solution but are not accurate enough to dispense the PC or the reference
1121 substances to the treatment wells.
1122 • Use of 8-channel reservoirs for applying dosing solutions to the wells so that
1123 multi-channel single delivery pipettes could be used
1124 • Use of a standardized length of time that the HBSS rinse remains on the cell
1125 monolayers in flasks during the cell subculture step
1126 • Protection of plates from light during the shaking step for NR extraction; all
1127 laboratories will cover plates with a light-impermeable barrier (e.g., aluminum
1128 foil) during this step
1129 • Allow plates to stand for at least five minutes after the shaking step is
1130 complete and eliminate any bubbles in media observed in the wells before
1131 measuring the OD
1132 • Change the allowable seeding density range for 3T3 NRU test method from
1133 $2.5x10^3$ cells/well to $2 – 3x10^3$ cells/well

1134 • Change the NHK culture flask size used at FAL for start-up of cryopreserved
1135 cells from 80-cm^2 to 25-cm^2 (the size the other laboratories had been using),
1136 and discontinue using a fibronectin-collagen coating.

1137 2.6.2.7 *Test Acceptance Criteria*

1138 The test acceptance criteria were revised as follows:

1139 • The IC_{50} for SLS (PC) should be within 2 SDs (approximately 95%) of the
1140 historical mean established by each laboratory in Phase Ia.
1141 • The mean OD values of the left and right VCs (columns 2 and 11 in the 96-
1142 well test plate) should not differ by more than 15% from the mean of all VC
1143 OD values on that plate.
1144 • At least one calculated cytotoxicity value should be between 10% and 50%
1145 viability, and one value between 50% and 90% viability.
1146 • The Hill function curve fit (R^2 >0.9 or 0.8 < R^2 <0.9) should be evaluated on a
1147 case-by-case basis for acceptability by the SMT.
1148 • VC OD criteria were based on Phase Ia data (mean ± two SDs): 0.3 to 0.8 for
1149 the 3T3 test method, and 0.6 to 1.7 for the NHK NRU test method
1150 (requirement for use of VC OD criteria as test acceptance criteria was
1151 rescinded after commencement of Phase Ib)

1152 2.6.3 <u>Phase II: Laboratory Qualification Phase</u>

1153 All protocol and acceptance criteria revisions were implemented during Phase II unless
1154 otherwise stated.

1155 2.6.3.1 *Testing of Volatile Reference Substances*

1156 When 2-propanol was tested in 3T3 and NHK cells, vapors from the highest concentration
1157 wells contaminated the adjacent VC wells and also appeared to affect some lower
1158 concentration wells (i.e., the wells exhibited unexpectedly reduced levels of NRU). An
1159 example range finder, concentration-response curve is shown in **Figure 2-4**. Such tests failed
1160 the VC criterion. When lower concentrations were used to avoid contaminating the VC wells
1161 adjacent to the highest concentration, the toxicity was inadequate to produce an IC_{50}. To
1162 address this problem, IIVS repeated their tests using film plate sealers, which isolated
1163 individual wells from one another; this was sufficient to prevent the cross-well

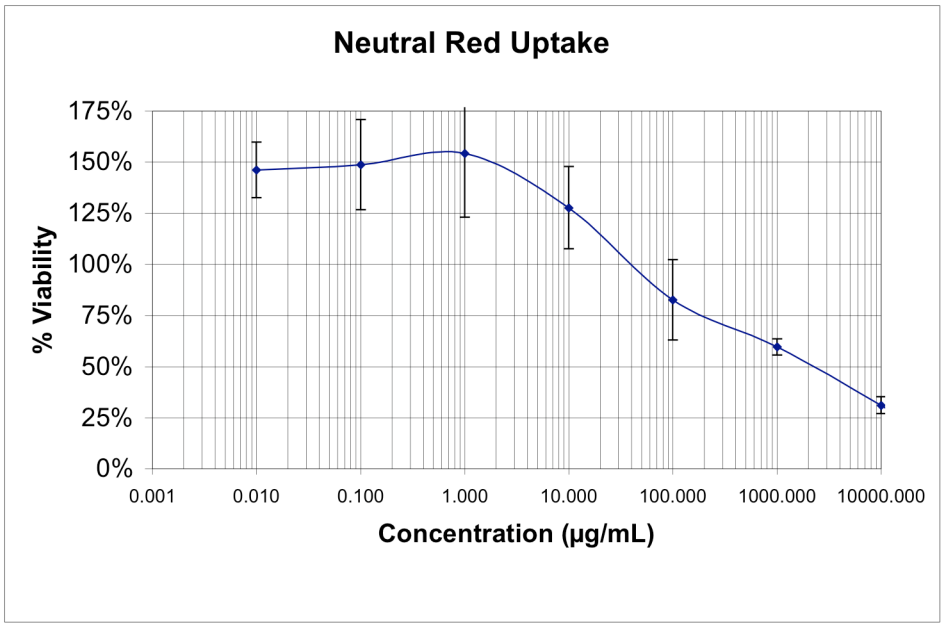
contamination, and acceptable results were obtained. Based on these data, the SMT recommended to the other two laboratories that film plate sealers be used when testing 2-propanol.

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

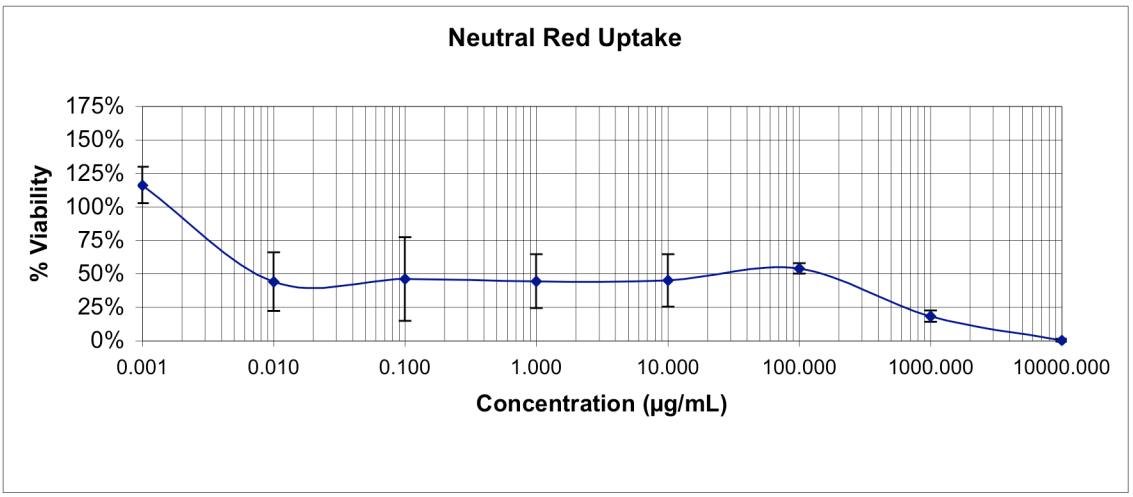
Abbreviations: 3T3=BALB/c 3T3 fibroblasts; VC=Vehicle control; C1 to C8=Test substance concentrations (C1-highest concentration, C8-lowest concentration); OD_{540}=Optical density at 540 nm; A to H=Row identification.
Note: %Difference of the two VC columns from the average VC was 63%. The mean corrected 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.

1195 2.6.3.2 *Atypical Concentration-Responses*

1196 Atypical concentration-responses are defined for this study as response curves that differ
1197 from a basic sigmoidal shaped curve. Curves that show a biphasic response as well as those
1198 that exhibited a plateau-like response at toxicity levels than 100% were considered atypical.
1199

1200 Two of the laboratories observed biphasic concentration-responses in the range finder tests
1201 for aminopterin and colchicine. When the range finder tests produced a biphasic response
1202 (see **Figure 2-5** for an example), the SMT advised the laboratories to focus the definitive
1203 tests on the lowest concentrations that produced at least a 50% loss in viability. Although
1204 doing so eliminated the biphasic response in the definitive tests, the highest tested
1205 concentrations did not reduce cell viability to 0% (see **Figure 2-6**). This effect with
1206 colchicine was very reproducible across laboratories in the NHK NRU test, but only FAL
1207 achieved this biphasic type of response with colchicine in the 3T3 NRU test. Aminopterin
1208 produced similar concentration-responses in the NHK NRU test at ECBC and FAL, but not
1209 at IIVS. In the 3T3 NRU test, only FAL obtained a biphasic response with aminopterin.
1210

1211 **Figure 2-5 Representative Concentration-Response for Aminopterin in a NHK NRU**
1212 **Range Finder Test**
1213

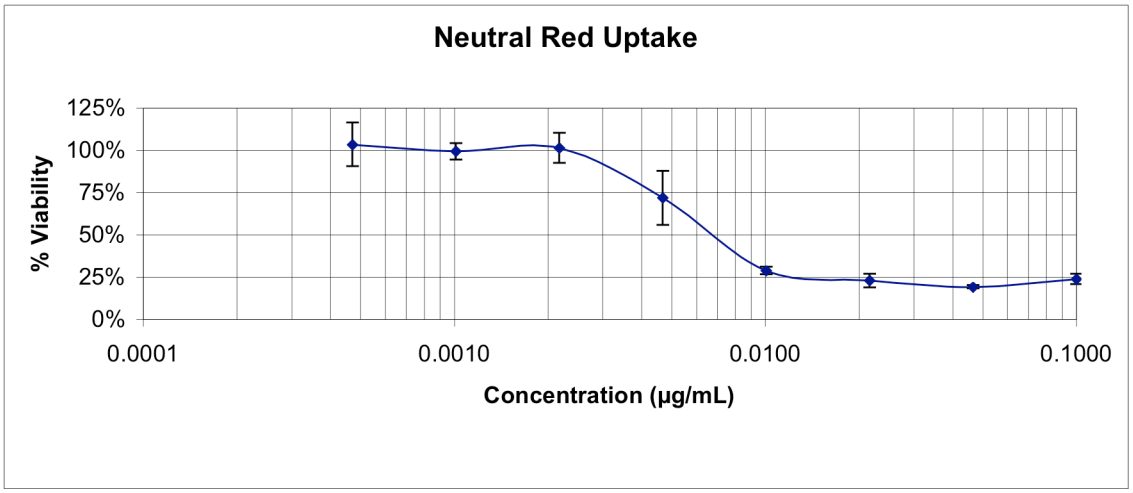


1214

1215 Abbreviations: NHK=Normal human keratinocytes; NRU=Neutral red uptake
1216 Representative dose-response for aminopterin in a NHK range finder test. Laboratories were instructed to
1217 focus the definitive tests on the lowest concentration that produced a 50% reduction in viability in the range
1218 finder test.
1219

1220

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



Abbreviations: NHK=Normal human 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

[4] Hormesis is a dose-response characterized by a compound's ability to cause a stimulatory effect in low doses or an inhibitory effect in high doses.

1244 binds to microtubular protein and interferes with function of mitotic spindles, which arrests
1245 cell division (NLM 2003). Aminopterin blocks the use of folic acid by the cells, inhibiting
1246 metabolism, RNA production, and protein synthesis, which is lethal during the S phase of the
1247 cell cycle by (NLM 2002). The variability of IC_{50} results for these substances among the
1248 laboratories may be due to different levels of cell confluence in the cultures at the time of
1249 treatment.

1250 2.6.3.3 *Hill Function*

1251 The Hill function used in the various phases of this study was defined as follows:
1252

1253 $$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC50 - \log X)\text{HillSlope}}}$$

1254 where Y= response (i.e., % viability), X is the substance concentration producing the
1255 response, Bottom is the minimum response (0% viability, maximum toxicity), Top is the
1256 maximum response (maximum viability), EC_{50} is the substance concentration at the response
1257 midway between Top and Bottom, and HillSlope describes the slope of the curve. When Top
1258 = 100% viability and Bottom = 0% viability, the EC_{50} is the equal to the IC_{50}.
1259
1260 Responses that do not achieve 100% cytotoxicity with increasing substance concentration do
1261 not fit the Hill function well. The R^2 values from such tests often failed the acceptance
1262 criterion. To obtain a better model fit, the Bottom parameter was estimated without
1263 constraints (the previous practice was to use Bottom = 0). However, when Bottom ≠ 0, the
1264 EC_{50} reported by the Hill function was not the same as the IC_{50} because the Hill function
1265 relies on $EC_{50,}$ which is defined as the point midway between the Top and Bottom responses.
1266 Thus, the Hill function calculation using the Prism® software was rearranged to calculate the
1267 IC_{50} as follows:

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

1268
1269 where IC_{50} is the concentration producing 50% toxicity, EC_{50} is the concentration producing
1270 a response midway between the Top and Bottom responses; Top being the maximum

1271 response (maximum survival), Bottom is the minimum response (0% viability, maximum
1272 toxicity), Y = 50 (i.e., 50% response), and HillSlope describes the slope of the response. The
1273 X from the standard Hill function equation is replaced, in the rearranged Hill function
1274 equation, by the IC_{50}.
1275
1276 IIVS performed the recalculations for their NHK NRU colchicine tests and the SMT
1277 performed the necessary recalculations for the other laboratories. Tests that were recalculated
1278 by the SMT are noted in the data summaries.
1279
1280 *Protocol Revision*: The protocol was revised to state that if a range finding test produces a
1281 biphasic response, then the concentrations selected for the subsequent tests should cover the
1282 most toxic dose-response range.

1283 2.6.3.4 *Insoluble Reference Substances*

1284 Lithium carbonate was insoluble in 3T3 medium. Only ECBC managed to expose 3T3 cells
1285 to sufficient lithium carbonate to produce three tests that met the acceptance criteria.
1286 Precipitate was reported for two of those tests at the three highest concentrations in the wells.
1287 Because the third highest concentration, 510.2 μg/mL, was approximately the IC_{50} (average
1288 was 564 μg/mL), the true IC_{50} for lithium carbonate may actually be lower than was
1289 calculated, and therefore the LD_{50} value would be underestimated. However, the data were
1290 reproducible and were not discarded.
1291
1292 *Protocol Revision for Phase III*: The protocol was revised to allow an increase in the
1293 stirring/rocking duration in an incubator from one to three hours if cytotoxicity in the range
1294 finder test was limited by solubility. Also, a **Stopping Rule for Insoluble Chemicals** was
1295 added (see **Section 2.5** and **Appendices B1** and **B2**) so that the laboratories would not
1296 continue repeated testing of insoluble substances in order to obtain three acceptable definitive
1297 tests.

1298 2.6.3.5 *Inadequate Cell Growth in NHK Medium*

1299 IIVS and FAL had several NHK NRU test failures that were attributed to poor cell growth.
1300 The SMT compiled KBM® and SingleQuot® lot numbers that the laboratories were using,

1301 along with the laboratory assessments of NHK cell growth. The information was used to
1302 identify the lots that produced adequate growth. The SMT also obtained quality assurance
1303 and quality control test results from CAMBREX Clonetics® on the lots of KBM®, but the
1304 information provided was inadequate for determining how the medium would perform in the
1305 NHK NRU test method.
1306
1307 *Resolution*: A protocol for prequalifying the medium was developed (see **Appendix B4**). For
1308 Phase III, the SMT asked IIVS to prequalify new lots of KBM® and SingleQuots® for use by
1309 all laboratories.

1310 2.6.3.6 *Performance Standards for Media to Support NHK Growth*

1311 A prequalification-of-medium protocol (**Appendix B4**) was developed and used by IIVS to
1312 test several different lots of medium and supplements to find combinations that maintained
1313 the typical growth characteristics of the NHK cells used in this study. The laboratories then
1314 reserved samples of the acceptable lots at CAMBREX so that testing would not be
1315 interrupted due to unavailability of adequate materials.
1316
1317 *Test Acceptance Criteria for Prequalifying Media Using SLS*

1318 • The fit of the SLS dose-response to the Hill model should be $R^2 \geq 0.85$ (i.e.,
1319 from PRISM® software).
1320 • The difference between the mean of all VCs and (a) the left mean VC, and (b)
1321 the right mean VC should be ≤15%.
1322 • At least one concentration should exhibit >0% and ≤50% viability and at least
1323 one should exhibit >50% and <100% viability.
1324 • After meeting all other acceptability criteria, the SLS IC_{50} must be within the
1325 historical range (± 2.5 SD) established by the laboratory.
1326
1327 *Other Criteria for Prequalifying Media (for consideration by a Study Director)*

1328 • General observations: rate of cell proliferation; percent confluence; number of
1329 mitotic figures per field; colony formation; distribution of cells in the flask;
1330 absence or presence of contamination

1331 • Cell morphology observations should include overall appearance (e.g., good,
1332 fair, poor), and presence of abnormal cells
1333 • Mean corrected OD_{540} of the VCs (e.g., are the values high/low when
1334 compared to historical data)
1335 • Cell morphology and confluence of the VC wells at the end of the 48-hour
1336 treatment
1337 • Cell doubling time, as compared to the doubling time with the previous
1338 batches of medium

1339 2.6.3.7 *Test Acceptance Criteria for Phase II*

1340 • The IC_{50} for SLS (PC) should be within 2.5 SDs of the historical mean
1341 established by the laboratory (*Phases Ia and Ib*)
1342 • Mean OD values of the left and right VCs (columns 2 and 11 in the 96-well
1343 test plate) do not differ by more than 15% from the mean of all VC well OD
1344 values. At least one calculated cytotoxicity value ≥10 % and ≤50 % viability
1345 and at least one value >50 % and ≤90 % viability
1346 • R^2 ≥0.90. The test fails if R^2 <0.80. If the 0.80 ≤ R^2 <0.90, the SMT evaluates
1347 the model fit (Note: The Study Director makes this determination for non-
1348 validation studies.)

1349 2.6.4 <u>Phase III: Laboratory Testing Phase</u>

1350 The changes below were made in the Phase III protocols based on the data and results in
1351 Phase II.

1352 2.6.4.1 *Required Cytotoxicity Values*

1353 Obtaining at least one calculated cytotoxicity value >0 % and ≤50 % viability and at least one
1354 that is >50 % and <100 % viability may be difficult or unattainable for substances with steep
1355 dose responses.
1356
1357 *Protocol Revision*: The test acceptance criterion was qualified so that tests with only one
1358 concentration between 0 and 100% viability were acceptable if the smallest practical dilution
1359 factor (i.e., 1.21) was used and all other test acceptance criteria were met.
1360

1361 Tests for three reference substances were accepted that met this new criterion in the 3T3
1362 NRU test method: diquat dibromide (1/9 tests); epinephrine bitartrate (2/9 tests); 1,1,1-
1363 trichloroethane (2/8 tests). No NHK tests required the use of these criteria (i.e., one point
1364 between 0 % and 100% viability at the lowest dilution factor).

1365 2.6.4.2 *Revisions to Data Analysis Procedures*

1366 The following revisions to data analysis procedures were made in Phase III NRU protocols:

1367 • If the Bottom parameter of the Hill function was fit to a value <0%, then the
1368 parameter was set to zero (0) for the IC calculations.
1369 • If toxicity plateaued above 20% viability (i.e., toxicity was <80%), the IC_{80}
1370 was not determined. The IC_{20} and IC_{50} values were calculated from the range
1371 of available toxic responses.
1372 • The requirement for substance dose-responses to fit the Hill equation with R^2
1373 ≥0.90 was rescinded. The Hill equation was used to characterize the shape of
1374 the response rather than to establish an acceptance criterion. The PC
1375 acceptance criterion was modified to R^2 ≥0.85.

1376 **2.7 Differences Between the 3T3 and NHK NRU Protocols for the Validation**
1377 **Study and the *Guidance Document* Standard Protocols**

1378 As the validation study progressed through Phases I and II, the protocols provided in the
1379 *Guidance Document* (ICCVAM 2001b) were optimized to address problems that were
1380 encountered during the validation study phases. Changes to the *Guidance Document*
1381 protocols are described below.

1382 • 3T3 cell seeding density for 96-well plates was decreased from $1x10^4$
1383 cells/well to $2.0 – 3.0x10^3$ cells/well.
1384 • The calcium concentration in NHK medium was changed from 0.15 mM to
1385 0.10 mM. The test laboratories had expressed concern that cell differentiation
1386 would occur at the higher concentration and requested a lower concentration.
1387 CAMBREX Clonetics®, the supplier of the NHK cells and NHK medium used
1388 in this study, normally grows NHK cells in 0.15 mM calcium and has seen no
1389 differentiation. The supplier agreed that the cells would grow well at 0.10 mM
1390 but should not be cultured at concentrations <0.10 mM in order to avoid

1391 morphological and growth rate changes (CAMBREX technical division,
1392 personal communication).

1393 • NHK cells were subcultured once prior to being distributed to the test wells,
1394 rather than for three passages. The laboratories expressed concern about the
1395 possibility of cell differentiation with subsequent passages in culture.

1396 • The highest recommended final concentrations of DMSO and ETOH in the
1397 culture media were reduced from 1% to 0.5%. IIVS performed experiments
1398 with both cell types to determine the concentration necessary to avoid solvent
1399 toxicity. 3T3 cells were tested with 0.5, 1, and 2% ETOH and DMSO at 0.1,
1400 0.2, 0.3, 0.4, 0.5, 1, and 2% concentrations. The 0.5% concentrations of both
1401 solvents were chosen as optimal because that concentration of ETOH
1402 produced no toxicity. Although 0.5% DMSO produced slight toxicity (i.e.,
1403 cells were 91% viable as compared to the control cells; See **Appendix E**), this
1404 concentration was chosen by the SMT and laboratories as an acceptable trade-
1405 off between slight toxicity and the ability to test substances at higher
1406 concentrations, and was used throughout the study for all reference substances
1407 that needed solvents other than culture medium (see Curren et al. 2003).
1408 DMSO was the preferred solvent if the test substance was not soluble in
1409 culture medium, and ETOH was not used in this study.

1410 • The pH of the reference substance solutions was not adjusted with NaOH or
1411 HCl regardless of whether solutions became acidic or basic (optimum
1412 mammalian cell culture pH is approximately 7.4 [Freshney, 2000]) upon
1413 addition of the test substance because some of the basal cytotoxicity produced
1414 by test substances may be due to pH effects. See **Appendix F1** for pH values
1415 of the reference substances in culture medium.

1416 • The CO_2 concentration in the incubator was reduced from 7.5% to 5.0%
1417 because the laboratories were already set up to use 5% CO_2, which is a typical
1418 optimum CO_2 concentration for mammalian cell culture.

1419 • Washing and fixing the cells with a formaldehyde solution prior to NR elution
1420 from the cells was eliminated. Formaldehyde disposal was problematic in
1421 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 E**).
- 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]	1.25×10^4	2.5×10^3	3.5×10^3	$2 – 2.5 \times 10^3$
Phase Ia	$0.42 – 1.68 \times 10^4$	2.5×10^3	$2.5 – 9 \times 10^3$	$2 – 2.5 \times 10^3$
Phase Ib	$0.42 – 1.68 \times 10^4$	2.5×10^3	$2.5 – 9 \times 10^3$	$2 – 2.5 \times 10^3$
Phase II	$0.42 – 1.68 \times 10^4$	$2 – 3 \times 10^3$	$2.5 – 9 \times 10^3$	$2 – 2.5 \times 10^3$
Phase III	$0.42 – 1.68 \times 10^4$	$2 – 3 \times 10^3$	$2.5 – 9 \times 10^3$	$2 – 2.5 \times 10^3$

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

2.8 Overview of the Solubility Protocol

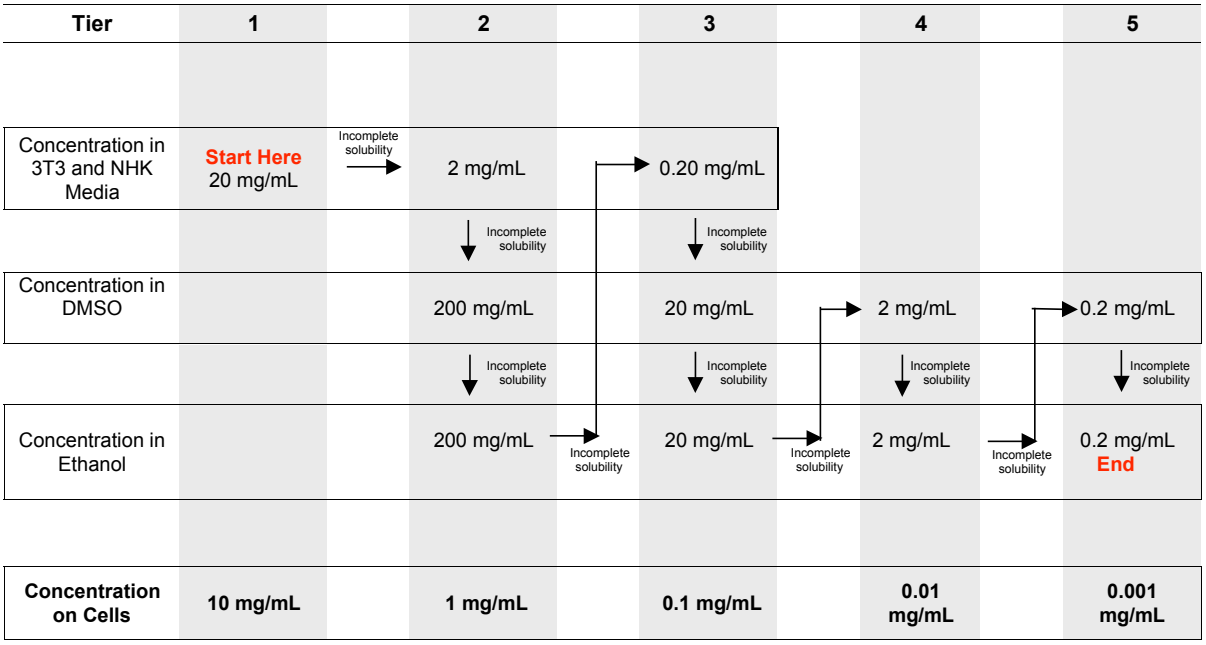
The SMT, with assistance from the laboratories, developed a solubility protocol to provide guidance for determining the most appropriate solvent for each test substance. The solubility protocol was based on an EPA guideline (EPA 1998) that involved testing for solubility in a particular solvent, beginning 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

1450 the solubility data provided by BioReliance and each testing laboratory, and designated the
1451 solvent to be used by all laboratories for each reference substance. This eliminated one
1452 potential variable in the NRU test results among laboratories.
1453
1454 The solubility protocol used by the *in vitro* laboratories during Phase III required the
1455 sequential testing of reference substances in the various solvents at concentrations that would
1456 be equivalent to the concentration that would be applied to the cell cultures. The solubility
1457 flow chart in **Figure 2-7** shows, for example, that 2 mg/mL medium and 200 mg/mL DMSO
1458 or ETOH were equivalent concentrations because they yielded 1 mg/mL in cell culture.
1459 Medium was diluted by one-half when applied to cultures. The 0.5% [v/v] final
1460 concentrations were achieved by diluting DMSO and ETOH by 200-fold. At each
1461 concentration, the following mixing procedures were employed, as necessary, to completely
1462 dissolve the reference substance in the sequence: vortex (1 to 2 minutes); sonication (up to 5
1463 minutes); warming to 37°C (5 to 60 minutes [NRU protocols allow warming to be extended
1464 to three hours if cytotoxicity in the range finder test was limited by solubility]). If the
1465 reference substance was still not dissolved, the next lower concentration, or a different
1466 solvent, was tested.

1467 **2.9 Basis of the Solubility Protocol**

1468 The solubility protocol used by BioReliance, which tested solubility of the reference
1469 substances prior to testing by the *in vitro* laboratories, is provided in **Appendix G**. The
1470 protocol is based largely on information from the literature and Internet searches for
1471 solubility procedures, the experience of the SMT and IIVS, and solubility and IC_{50}
1472 information from the RC chemicals database (Halle 1998, 2003). The only formal solubility
1473 protocol discovered was the EPA Product Properties Test Guideline, OPPTS 830.7840 Water
1474 Solubility Column Elution Method; Shake Flask Method (EPA 1998).
1475
1476
1477

1477 **Figure 2-7 Flow Chart for Determination of Reference Substance Solubility in**
1478 **Medium[1], DMSO, or ETOH**



1479
1480 Abbreviations: DMSO=Dimethyl sulfoxide; ETOH=Ethanol; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal
1481 human keratinocytes.
1482 [1]3T3 Medium - DMEM (Dulbecco's Modification of Eagle's Medium) with supplements; NHK medium -
1483 KBM® (Keratinocyte Basal Medium) with supplements (from CAMBREX Clonetics®).
1484 Note: DMSO is U.S.P. analytical grade. ETOH is U.S.P. analytical grade (100% non-denatured).
1485
1486

1487 2.9.1 Initial Solubility Protocol Development

1488 BioReliance evaluated the solubility of each reference substance in cell culture media at
1489 2000, 400, and 200 mg/mL, and if not soluble at those concentrations, in DMSO and then
1490 ETOH, at the same concentrations (initial protocol). It was apparent that these concentrations
1491 were not low enough when the laboratory was unable to achieve solubility for arsenic
1492 trioxide. The solubility protocol was revised twice to lower the range of concentrations tested
1493 (see **Table 2-5**). An extra tier of concentrations ≤1 mg/mL was added for poorly soluble and
1494 insoluble substances. The protocol used by the laboratories was further revised to reduce the
1495 number of steps required (by testing in log units) and to test in tiers using concentrations that
1496 reflected the concentrations anticipated in the cell cultures (see **Figure 2-7**).
1497
1498

1498 **Table 2-5 Comparison of Concentrations Tested in the Various Solubility Protocols**
1499

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

1500 Abbreviations: DMSO=Dimethyl sulfoxide; ETOH=Ethanol; Medium=Cell culture medium; NA=Not applicable
1501 Note: DMSO is U.S.P. analytical grade. ETOH is U.S.P. analytical grade (100% non-denatured).

1502

1503 In Phases Ib and II, the SMT used the data from BioReliance to select the solvents to be used
1504 for testing the various chemicals. When it became apparent that the laboratories sometimes
1505 obtained different solubility results than those reported by BioReliance, the SMT used the
1506 cytotoxicity results from the laboratories to determine the solvents to be used for Phase III
1507 reference substances.

1508

1509 The final protocol provided a tiered approach for determining the 2X stock concentration for
1510 each reference substance (see **Figure 2-7**). This protocol had the advantage of reducing the
1511 number of steps for testing (compared to that used by BioReliance) (see **Appendix B3**).

1512 2.9.2 Basis for Modification of the Phase II Protocol

1513 All three testing laboratories found arsenic trioxide (tested in Phase Ib) less soluble (see
1514 **Table 5-9**) than was reported by BioReliance (BioReliance values: 0.25 mg/mL in 3T3
1515 medium and 0.05 mg/mL in NHK medium). This chemical was not soluble using the
1516 procedures in the initial solubility protocol. IIVS warmed the stock solution (at least 200
1517 μg/mL for 2X) for longer than the protocol specified (i.e., 30 to 50 minutes) but still had
1518 persistent, small, undissolved particles. ECBC obtained a clear solution (highest 2X
1519 concentration was 30 to 50 μg/mL), but found precipitated particles after the solution stood at
1520 room temperature. Sonication time was increased to 15 to 30 minutes, and heating time to
1521 approximately 30 minutes to get a finer suspension. This procedure achieved a more

1522 homogeneous mixture, resulting in more uniform serial dilutions and a more even application
1523 of the reference substance to the cells. FAL stirred the suspension (approx. 20 to 90 µg/mL)
1524 in the CO_2 incubator for 1.5 to 2 hours to get clear medium.
1525
1526 *Protocol Revision for Phase II*: The duration of the heating step was altered from 5 to 20
1527 minutes to 5 to 60 minutes.

1528 **2.10 Components of the Solubility Protocol**

1529 2.10.1 Medium, Supplies, and Equipment Required

1530 2.10.1.1 *Medium and Chemical Supplies*

1531 • 3T3 culture medium: DMEM without L-glutamine and containing Hanks'
1532 salts and high glucose [4.5gm/l]; L-glutamine, 200 mM; NCS
1533 • NHK culture medium: Keratinocyte Basal Medium without Ca^{++} (KBM®,
1534 Clonetics® CC-3104); KBM® SingleQuots® medium supplements (Clonetics®
1535 CC-4131): epidermal growth factor, insulin, hydrocortisone, bovine pituitary
1536 extract; Calcium SingleQuots® (Clonetics® CC-4202); penicillin/streptomycin
1537 solution (antimicrobial agents)
1538 • United States Pharmacopoeia (U.S.P.) analytical grade DMSO
1539 • U.S.P. analytical grade (100%, non-denatured) ETOH

1540 2.10.1.2 *Equipment*

1541 • Waterbath (37°C)
1542 • Sonication apparatus
1543 • Vortex mixer
1544 • Micropipettors
1545 • Balance (capable of weighing 10 mg)
1546 • pH meter

1547 2.10.1.3 *Procedures*

1548 The initial Phase III solubility protocol required the dissolving of approximately 10 mg of
1549 reference substance in approximately 0.5 mL medium (both 3T3 and NHK media were used)
1550 for a final concentration of 20 mg/mL (see **Appendices B1** and **B2**). In order, the mixture
1551 was vortexed for 1 to 2 minutes, sonicated for up to 5 minutes, and warmed to 37°C for 5 to

1552 60 minutes, as necessary, to dissolve the substance. The endpoint for dissolution was a clear
1553 solution with no noticeable precipitate. If the reference substance was not soluble in medium
1554 at 20 mg/mL, then more medium was added to a concentration of 2 mg/mL (i.e., a total
1555 volume of approx. 5 mL) (Step 2). The mixing procedures were repeated as necessary to
1556 dissolve the reference substance. If the reference substance did not dissolve, approximately
1557 10 mg reference substance was added to approximately 0.5 mL DMSO in an attempt to
1558 dissolve it at a concentration of 200 mg/mL (Step 3). If the reference substance was not
1559 dissolved, the same concentration was attempted in 100% ETOH (Step 4). Step 5 began in
1560 the same way, with 0.2 mg/mL medium and then progressed to 20 mg/mL DMSO, and then
1561 20 mg/mL ETOH.
1562
1563 Determination of reference substance solubility was limited to visual observation of the
1564 resulting solution. If a solution appeared clear, then solubility testing ceased. If particles were
1565 visible or if the solution appeared cloudy, then more stringent mixing and/or heating
1566 procedures were employed. If necessary, the solubility procedure proceeded to the next
1567 solvent/concentration tier. The duration of the solubility test was dependent on the
1568 procedures used to achieve solubility. Some reference substances were immediately
1569 solubilized (e.g., liquids) and others required up to 60 minutes of heating and agitation or
1570 sonication.

1571 2.10.2 Data Collection

1572 All laboratories (including the reference substance distribution laboratory, BioReliance) used
1573 a worksheet designed to capture the solubility information for each reference substance. The
1574 endpoint for each step was a visual observation of the solution, a documented comment
1575 describing the observation, the concentration, and a conclusion of soluble or insoluble. Each
1576 worksheet contained:

1577 - Reference substance code number and physical description
1578 - Solvent used (3T3 medium, NHK medium, DMSO, ETOH)
1579 - Amount of reference substance (mg) used in the initial stage
1580 - Volume of solvent added and final volume (mL)
1581 - Test substance concentration (µg/mL) in the solvent
1582 - pH and color of the solution

1583 • Mechanical procedures used (vortexing, sonication, heating), duration, and
1584 temperature
1585 • Comments (soluble/insoluble at the particular concentration; visual
1586 observations; reactivity with solvent)
1587
1588 The solubility test information and data from the laboratories were transferred via email to
1589 the SMT and stored on the NICEATM server and as hard-copy printouts. Each laboratory
1590 also maintained electronic and hard-copy files of its data.

1591 2.10.3 Variability in Solubility Measurements

1592 Solubility determinations were not replicated because within-laboratory results were not
1593 expected to vary. Comparison of the results to determine inter-laboratory concordance for the
1594 72 reference substances (see **Section 5.8** for results) provided a measure of variability among
1595 the laboratories and information about the reproducibility of the solubility determinations
1596 (see **Section 7.4**).

1597 2.10.4 Solubility Issues During the Testing of the Reference Substances

1598 Substance solutions were monitored throughout all aspects of the test procedures, and
1599 observations were documented. The 2X and 1X solutions were permitted to contain
1600 precipitates for the range finder tests, but not for the definitive tests. The lowest
1601 concentration of the substance in a 2X solution that contained observable precipitates,
1602 particles, globules, or oily droplets, was documented in the EXCEL® template. After
1603 substance exposure, all wells of the 96-well test plates were observed microscopically and
1604 scored using a visual observation code. The code addressed growth characteristics and the
1605 presence or absence of precipitates (see **Appendix B** [test method protocols] for the
1606 observation codes used). The Study Directors made determinations of test acceptance based
1607 on the recommended concentration levels and the presence of precipitates.

1608 2.10.5 Analysis of Solubility Data

1609 During Phase III, the SMT used the solubility data from all laboratories to determine the
1610 solvents to be used for each chemical (see **Section 5.8** for solubility results and SMT
1611 selections). If the solubility of an individual reference substance was different in 3T3
1612 medium and NHK medium, the same solvent would be used for both test methods, rather

1613 than having different solvents for each method. For example, if solubility in one culture
1614 medium was ≥2 mg/mL and solubility in the other was <2 mg/mL, and the substance was
1615 soluble in DMSO at 200 mg/mL, the SMT would select DMSO as the solvent for both test
1616 methods (each test method using its respective culture medium).
1617
1618 Solubilizing sufficient reference substance to produce cytotoxicity was challenging for
1619 relatively insoluble, low toxicity, substances such as lithium carbonate (in the 3T3 NRU test
1620 method) but generally was not a problem for toxic substances that did not require as high a
1621 concentration to kill cells. Some insoluble and highly toxic reference substances were
1622 problematic, however, because the amount of powdered reference substance added to solvent
1623 was very small, and laboratory personnel found it difficult to determine the presence of
1624 solute particles in solution. Arsenic trioxide is an example of such a solute (see **Section**
1625 **2.9.2**).

1626 **2.11 Summary**

1627 The *Guidance Document* NRU protocols were used as the basis of the validation study
1628 protocols. The SMT and participating laboratories made initial modifications to the protocols
1629 prior to implementation of the study. Other protocol modifications were made after
1630 commencement of testing and were the result of recommendations from the laboratories and
1631 the SMT, based on their experience with the initial protocols. The resulting optimized
1632 protocols were used in the main testing phase (Phase III) of the study.
1633
1634 The protocol components used in the validation study were similar for the 3T3 and NHK
1635 cells. The following procedures were common to the NRU protocols for both cell types:
1636 • Testing was performed in four phases (Ia, Ib, II, and III)
1637 • Preparation of reference substances and positive control
1638 • Cell culture environment conditions
1639 • Determination of test substance solubility
1640 • Configuration of 96-well plates for testing samples
1641 • 48-hour exposure to test substance
1642 • Range finder and definitive testing
1643 • 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.

3.0 REFERENCE SUBSTANCES USED FOR VALIDATION OF THE 3T3 AND NHK NRU TEST METHODS 3-3

3.1 Rationale for the 72 Reference Substances Selected for Testing 3-3
3.1.1 Reference Substance Selection Criteria 3-3
3.1.2 Candidate Reference Substances 3-5
3.1.3 Selection of Reference Substances for Testing 3-6

3.2 Characteristics of the Selected Reference Substances.....3-7
3.2.1 Source Databases Represented by the Selected Reference Substances 3-8
3.2.2 Chemical Classes Represented by the Selected Reference Substances 3-31
3.2.3 Product/Use Classes Represented by the Selected Reference Substances...3-32
3.2.4 Toxicological Characteristics of the Selected Reference Substances.....3-32
3.2.5 Selection of Reference Substances for Testing in Phases Ib and II.....3-37
3.2.6 Unsuitable and Challenging Reference Substances.....3-38

3.3 Reference Substance Procurement, Coding, and Distribution 3-40
3.3.1 Exceptions 3-43

3.4 Reference Substances Recommended by the *Guidance Document* 3-42

3.5 Summary.....3-42

[This Page Intentionally Left Blank]

3.0 REFERENCE SUBSTANCES USED FOR VALIDATION OF THE 3T3 AND NHK NRU TEST METHODS

3.1 Rationale for the 72 Reference Substances Selected for Testing

This section describes the procedures used to select the 72 reference substances selected for testing in Phase Ia of the validation study.

3.1.1 Reference Substance Selection Criteria

The SMT (see **Appendix A**) selected reference substances for testing using a process based on general recommendations made by Workshop 2000 participants (ICCVAM 2001a). The following criteria were used:

- The toxicities of the reference substances should be evenly distributed across the expected range of rodent LD_{50} values, using the GHS classification for acute oral toxicity as a guide (UN 2005).
- The reference substances should cover a wide range of structural and use classes, and be relevant to the needs of the various user communities.
- Substances with human toxicity data and/or human exposure potential (i.e., substances of interest to society) should be included. Substances with human acute toxicity data were particularly important to ECVAM for determining the relationship of the NRU IC_{50} values to human blood/serum LC.

Table 3-1 shows the GHS scheme for classifying substances into six toxicity categories (five with measured LD_{50} ranges and an unclassified category with LD_{50} values greater than 5000 mg/kg) based on acute rodent oral LD_{50} values (UN 2005). The SMT used this scheme for the classification of candidate substances to assure that the reference substances selected for the validation study represented the full range of acute oral toxicity.

Table 3-1 GHS Classification Scheme for Acute Oral Toxicity

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 5$
2	$5 < LD_{50} \leq 50$
3	$50 < LD_{50} \leq 300$
4	$300 < LD_{50} \leq 2000$
5	$2000 < LD_{50} \leq 5000$
Unclassified	$LD_{50} > 5000$

Abbreviations: UN=United Nations; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
LD_{50}=dose that produces lethality in 50% of the test animals.

For the purposes of the initial toxicity classification, the rodent oral LD_{50} values for the individual substances were obtained from readily available toxicological databases. These rodent oral LD_{50} values were re-evaluated in **Section 4** for the purpose of identifying the most appropriate reference LD_{50} values to use for the accuracy analyses (i.e., determine to what extent there is agreement between a test method result and an accepted reference value [see **Section 6]**). Rat LD_{50} data were preferred because:

- The current acute oral toxicity test guidelines recommend using rats (OECD 2001a, c, d; EPA 2002a)
- The majority of LD_{50} data used in the RC millimole regression were from studies using rats (282 rat data points and 65 mouse data points) (Halle 1998, 2003)
- The great majority of acute oral systemic toxicity testing is performed with rats

Mouse oral LD_{50} values were used (10 substances) for the initial toxicity classification when rat data were unavailable, however, mouse data were not used in the regression analyses presented in **Section 6**. The toxicological databases, in order of preference, were:

- The RC, which contains LD_{50} values that came largely from the 1983/84 RTECS® (Halle 1998, 2003). The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for chemicals with known molecular weights.

- The current RTECS® (MDL Information Systems 2001, 2002)
- The current Hazardous Substances Data Bank (HSDB; U.S. National Library of Medicine [NLM] 2001, 2002).

To insure that a wide range of structural and use classes were selected, reference substances of interest to the various U.S. regulatory agencies, as determined from substance lists received from the various agencies, were included. Substances with human toxicity data and/or human exposure potential were chosen by mining publicly available databases (e.g., the NTP test database, the MEIC database) for potential candidates.

3.1.2 Candidate Reference Substances

The process of identifying the 72 reference substances started with the compilation of a database of 116 candidates. The intent of the SMT was to compile a database with at least 12 substances in each GHS toxicity category that also met the other criteria, and then to prioritize the substances within each category to select the 72 to be tested. As recommended by Workshop 2000 (ICCVAM 2001a), the following publicly available databases and other sources were used to identify candidate substances:

- The MEIC program, which collected human toxicity data and *in vitro* toxicity data from 61 test methods for 50 substances (Ekwall et al. 1998)
- The EDIT program, which targeted development of *in vitro* test methods for endpoints other than basal cytotoxicity; includes 17 chemicals that are a subset of the MEIC chemicals
- The RC (Halle 1998, 2003), which contains *in vitro* cytotoxicity and *in vivo* rodent LD_{50} data for 347 substances
- The Toxic Exposure Surveillance System (TESS) (Litovitz et al. 2000), which compiles reports of toxic human exposures from poison control centers throughout the United States
- Pesticides recommended for consideration by the EPA Office of Pesticide Programs (OPP)
- The *Guidance Document* (ICCVAM 2001b), which reported *in vitro* NRU results for 11 RC substances using protocols similar to those to be used in the validation study

- The U.S. NTP test database, which contains information on the toxicity of substances relevant to human exposure (NTP 2002)
- The EPA High Production Volume (HPV) Challenge Program list of chemicals. The HPV is a voluntary testing program to provide the public with a complete set of baseline health and environmental effects data for each chemical that is manufactured within or imported into the United States at amounts >1 million pounds/year (EPA 2000)

The candidate substances from the list of 116 that were not selected as reference substances to use in the validation study are listed in **Appendix F3**, grouped by GHS category, the rat or mouse oral LD_{50} value, the database(s) or other source(s) used to identify the substance as a potential candidate, and the type of product and/or use for the substance.

3.1.3 Selection of Reference Substances for Testing

Using the candidate substance database, 72 reference substances (12 GHS-unclassified substances and 12 substances from each of the five GHS acute oral toxicity hazard categories) were selected. This number of substances per GHS category was considered adequate by the ICCVAM Acute Toxicity Working Group (ATWG), ICCVAM, ECVAM, and the SMT to accurately evaluate the performance of these two *in vitro* NRU test methods for identifying the starting dose for rodent acute oral toxicity tests across the range of toxic levels that would be encountered during testing. The criteria used for prioritizing the candidate substances were:

- The extent of rodent acute oral toxicity data
- The extent of human acute oral toxicity data and/or relevance for human exposure
- The level of volatility (because the cells are exposed for 48 hours while incubated at 37°C in 96-well plates, volatilization from wells containing a volatile reference substance would affect the accuracy of the IC_{50} calculation and potentially contaminate other wells)
- Not a controlled substance according to the U.S. Drug Enforcement Agency (DEA). Excluding substances that are listed in DEA Schedules I and II from

consideration obviates the requirement for U.S. laboratories to obtain a DEA license and adhere to the DEA substance storage and control procedures

- Practical considerations such as cost and disposal

If more than 12 candidate substances in a GHS category met the above criteria, then selection was based on two further considerations. One consideration was the distribution of substance toxicities within each toxicity category so as to select substances that represented the entire range of toxicity within each category. Another consideration, which applied only to candidate substances selected from the RC database, was the fit of the toxicity to the RC millimole regression. Substances with the best fit to the RC millimole regression were preferentially selected to prevent the entire set of reference substances from having proportionally more "outlier" substances (i.e., greater than one-half log from the RC millimole regression) than the entire RC database.

The final list of selected reference substances is sorted by GHS acute oral toxicity category in **Table 3-2.**

3.2 Characteristics of the Selected Reference Substances

The physical/chemical and toxicological information in **Appendix F** may be useful for characterizing the performance of the *in vitro* NRU test methods for various chemical types (e.g., chemical class, toxic effect class). **Appendix F1** lists the reference substances in alphabetical order with information on the CASRN, purity, supplier, pH (of the use concentration), and concentrations tested. **Appendix F2** provides the reference substances in alphabetical order, and information on physical/chemical characteristics such as molecular weight, chemical class, water solubility, acid/base dissociation constant (pK), boiling point, and octanol-water partition coefficient (log K_{ow}), a measure of lipid solubility. Although test substance concentration and toxicity may be heavily influenced by molecular charge and surface activity (ICCVAM 2006), these attributes were not characterized because this type of information is not readily available. **Appendix F2** also includes the major toxic effects attributed to each chemical, ability to pass the blood:brain barrier (BBB), metabolic activation/inactivation (whether or not it is metabolized, or the identification of the

metabolites), and mechanism of lethality (where known) for each of the reference substances. The remainder of this section summarizes selected characteristics of the reference substances.

3.2.1 Source Databases Represented by the Selected Reference Substances

The primary sources of substances were well represented in the final list of reference substances. **Table 3-3** shows the distribution of reference substances by GHS category from each of the source lists. Forty-two (58%) of the 72 substances were MEIC chemicals (17 of the 42 MEIC chemicals [40%] were also EDIT chemicals), 46 (64%) were involved in human poisonings as reported by TESS, 51 (71%) have been evaluated by the NTP, and 18 (25%) are listed in the EPA's HPV Challenge Program. Some substances were present in more than one database.

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

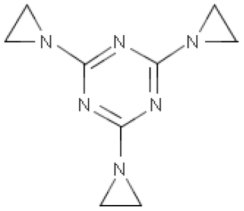
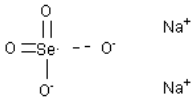
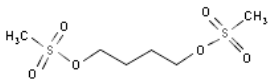
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
$LD_{50} \leq 5$ mg/kg							
Mercury II chloride	1	MEIC, EDIT, RC (outlier), TESS, NTP	Preservative; Manufacturing; Insecticide	271.50	0.22	Inorganic compound; Mercury compound; Chlorine compound	
Triethylenemelamine	1	RC (outlier), NTP	Manufacturing; Insect chemosterilant	204.23	–0.54	Organic compound; Heterocyclic compound	
Sodium selenate	2**	TESS, NTP	Feed additive	188.90	NA	Inorganic compound; Sodium compound; Selenium compound	
Busulfan	2	RC (outlier), NTP	Pharmaceutical (antineoplastic)	246.31	–0.52	Organic compound; Alcohol; Acyclic hydrocarbon; Sulfur compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

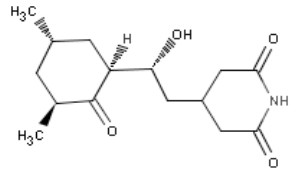
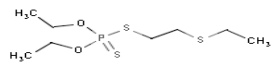
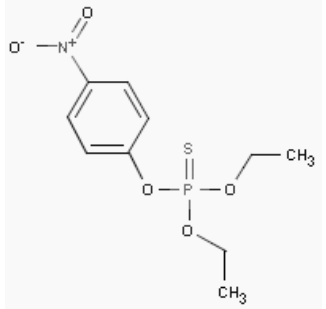
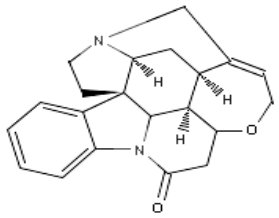
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Cycloheximide	2	RC (outlier), NTP	Antibiotic Fungicide	281.40	0.55	Organic compound; Heterocyclic compound	
Disulfoton	2	RC (outlier), EPA, NTP	Pesticide (insecticide)	274.42	4.02	Organic compound; Organophosphorous compound; Sulfur compound	
Parathion	2	RC (outlier), EPA, NTP	Pesticide (insecticide)	291.28	3.83	Organic compound; Organophosphorous compound; Sulfur compound	
Strychnine	2*	MEIC, TESS, EPA	Pesticide (rodenticide)	334.40	1.93	Organic compound; Heterocyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

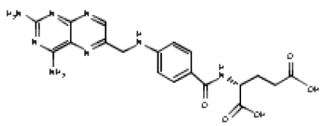
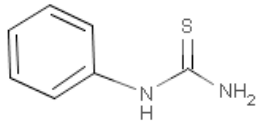
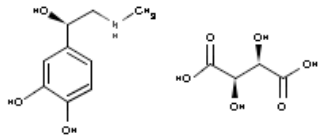
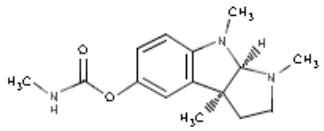
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Aminopterin	3**	RC	Pharmaceutical (antineoplastic); Pesticide (rodenticide)	476.45	NA	Organic compound; Heterocyclic compound	
Phenylthiourea	3	RC (outlier), NTP	Pesticide (rodenticide)	152.20	0.71	Organic compound; Sulfur compound; Urea	
Epinephrine bitartrate	4**	RC (outlier), NTP (HCl salt)	Pharmaceutical (adrenergic)	333.30	–1.52	Organic compound; Alcohol; Amine	
Physostigmine	5*	EHS	Pharmaceutical (anticholinesterase)	275.40	NA	Organic compound; Carboxylic acid; Heterocyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

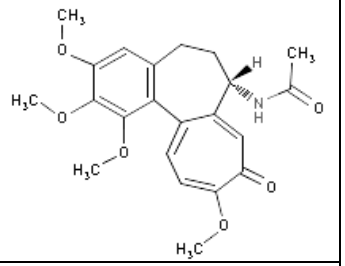
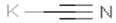
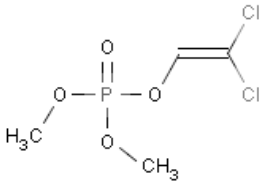
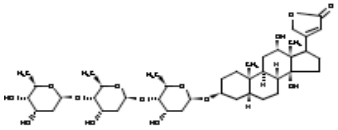
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
5< LD_{50} ≤50 mg/kg							
Colchicine	6**	MEIC, RC, TESS, NTP	Pharmaceutical (gout suppressant)	399.45	1.03	Organic compound; Polycyclic compound	
Potassium cyanide	10	MEIC, EDIT, RC (outlier), TESS	Electroplating	65.12	NA	Inorganic compound; Potassium compound; Nitrogen compound	
Dichlorvos	17*	TESS, EPA, NTP, HPV	Pesticide (insecticide)	220.98	1.43, 1.45	Organic compound; Organophosphorous compound	
Digoxin	18**	MEIC, EDIT, RC (outlier), TESS	Pharmaceutical (antiarrhythmic)	780.90	1.26	Organic compound; Polycyclic compound; Carbohydrate	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

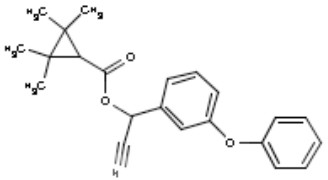
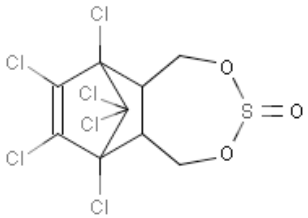
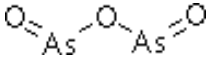
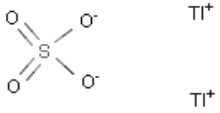
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Fenpropathrin	18*	EPA	Pesticide (insecticide)	349.43	6.0 @ 20° C	Organic compound; Nitrile; Ester; Ether	
Endosulfan	18*	TESS, EPA, NTP	Pesticide (insecticide)	406.91	3.83	Organic compound; Heterocyclic Compound; Sulfur compound	
Arsenic III trioxide	20	MEIC, EDIT, RC, TESS, EPA, NTP	Pesticide (insecticide)	197.80	NA	Inorganic compound; Arsenical	
Thallium I sulfate	29**	MEIC, EDIT, RC (outlier), TESS	Pesticide (rodenticide/insecticide)	504.80	NA	Inorganic compound; Metal; Sulfur compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

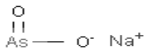
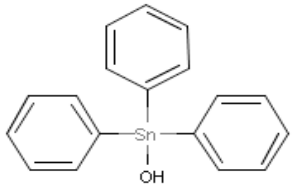
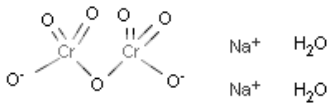
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Sodium arsenite	41*	TESS, NTP	Pesticide (herbicide, insecticide, fungicide)	129.90	NA	Inorganic compound; Arsenical; Sodium compound	
Triphenyltin hydroxide	44	RC, EPA, NTP, HPV	Pesticide (fungicide/insecticide)	367.02	NA	Organic compound; Organometallic compound	
Sodium dichromate dihydrate	50	RC, EPA, GD, NTP	Oxidizing agent	298.00	NA	Inorganic compound; Sodium compound; Chromium compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

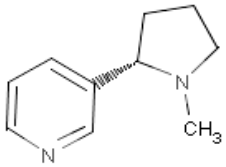
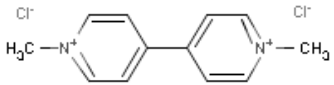
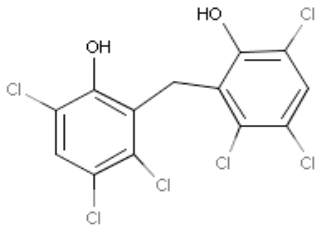
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Nicotine	50	MEIC, EDIT, RC (outlier), TESS, EPA, NTP	Pharmaceutical (stimulant)	162.020	1.17	Organic compound; Heterocyclic compound	
$50 < LD_{50} \leq 300$ mg/kg							
Paraquat	58	MEIC, EDIT, RC (outlier), TESS, EPA	Pesticide (herbicide)	257.20	–4.22 @ pH 7.4	Organic compound; Heterocyclic compound	
Hexachlorophene	61	MEIC, RC, TESS, NTP	Disinfectant	406.91	6.91	Organic compound; Cyclic hydrocarbon; Phenol	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

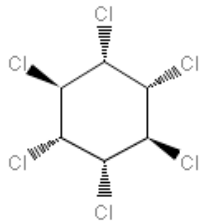
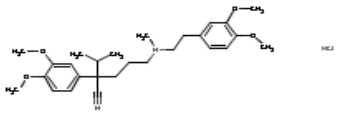
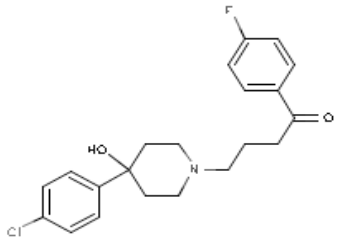
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Lindane	76	MEIC, EDIT, RC (outlier), EPA, NTP	Pesticide (insecticide)	290.80	3.72	Organic compound; Halogenated hydrocarbon	
Cadmium II chloride	88	RC, TESS, GD, NTP	Consumer; Industrial products	183.31	NA	Inorganic compound; Cadmium compound	
Verapamil HCl	108	MEIC, EDIT, RC (outlier), TESS, NTP	Pharmaceutical (antiarrhythmic)	491.08	3.79	Organic compound; Amine	
Haloperidol	128*	MEIC, TESS	Pharmaceutical (antipsychotic)	375.90	3.36	Organic compound; Ketone	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

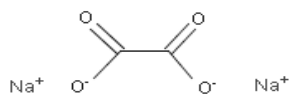
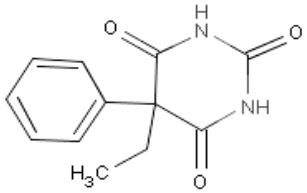
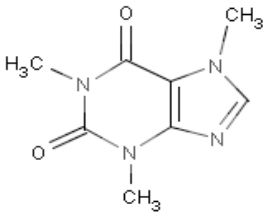
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Sodium oxalate	155	MEIC, EDIT, RC, TESS, NTP	Paints; Cleaners	134.00	NA	Organic compound; Carboxylic acid; Sodium compound	
Phenobarbital	163	MEIC, RC (outlier), TESS, NTP	Pharmaceutical (anticonvulsant)	232.23	1.47	Organic compound; Heterocyclic compound	
Sodium I fluoride	180	MEIC, RC, TESS, EPA, NTP	Electroplating; Water fluoridation	41.99	NA	Inorganic compound; Sodium compound; Fluorine compound	
Caffeine	192	MEIC, RC (outlier), TESS, NTP, HPV	Pharmaceutical (stimulant); Food additive	194.20	–0.07	Organic compound; Heterocyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

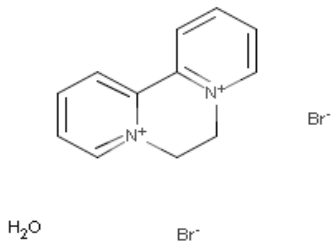
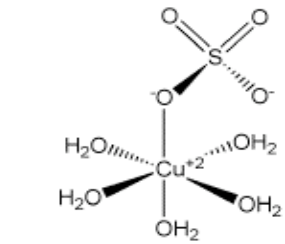
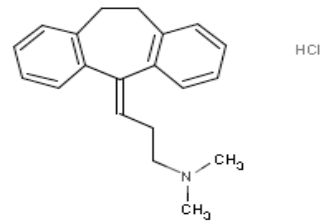
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Diquat dibromide	231	MEIC, RC, TESS	Pesticide (herbicide)	362.10	–3.05	Organic compound; Heterocyclic compound	
Cupric sulfate * 5 H2O	300	MEIC, RC, TESS, EPA, NTP	Pesticide (insecticide/fungicide)	249.70	NA	Inorganic compound; Sulfur compound; Metal	
300< LD_{50} ≤2000 mg/kg							
Amitriptyline HCl	319	MEIC, EDIT, RC, TESS	Pharmaceutical (antidepressant)	313.90	5.04	Organic compound; Polycyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

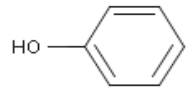
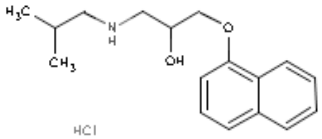
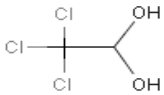
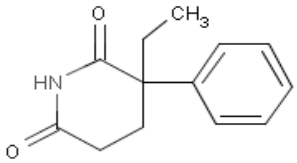
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Phenol	414	MEIC, RC, TESS, EPA, NTP, HPV	Disinfectant	94.11	1.46	Organic compound; Phenol	
Propranolol HCl	470**	MEIC, RC, TESS, GD	Pharmaceutical (antiarrhythmic)	295.80	3.09	Organic compound; Alcohol; Amine; Polycyclic compound	
Chloral hydrate	479	MEIC, RC, TESS, NTP	Pharmaceutical (sedative)	165.40	0.99	Organic compound; Alcohol	
Glutethimide	600	MEIC, RC, TESS	Pharmaceutical (sedative)	217.30	1.9	Organic compound; Heterocyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

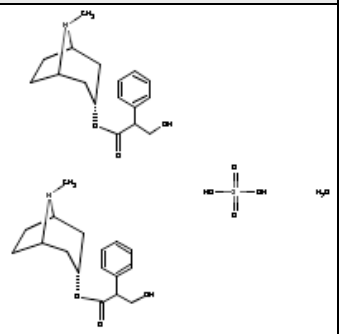
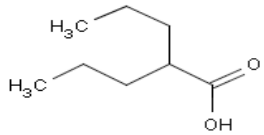
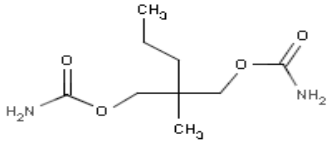
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Atropine sulfate	623	MEIC, EDIT, RC, TESS	Pharmaceutical (antimuscarinic)	694.80	1.83	Organic compound; Heterocyclic compound	
Valproic acid	1695 **	RC, MEIC, TESS, NTP	Pharmaceutical (anticonvulsant)	144.20	2.75	Organic compound; Carboxylic acid; Lipids	
Meprobamate	794*	MEIC, TESS	Pharmaceutical (antidepressant)	218.30	NA	Organic compound; Carboxylic acid	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

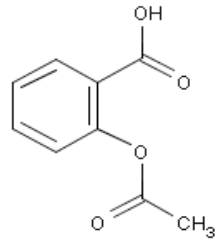
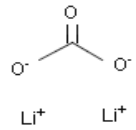
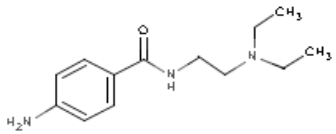
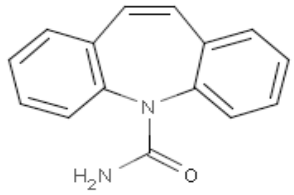
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Acetylsalicylic acid	1000	MEIC, EDIT, RC, TESS, NTP	Pharmaceutical (analgesic)	180.20	1.19	Organic compound; Carboxylic acid; Phenol	
Lithium I carbonate	1187[7]	MEIC, RC, TESS, NTP (Cl salt)	Pharmaceutical (mood stabilizer)	73.89	NA	Inorganic compound; Lithium compound; Alkylies; Carbon compound	
Procainamide	1950*	MEIC, TESS	Pharmaceutical (antiarrythmic)	271.79	NA	Organic compound; Carboxylic acid; Amide	
Carbamazepine	1957*	MEIC, TESS	Pharmaceutical (antiepileptic)	236.30	2.45	Organic compound; Heterocyclic compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

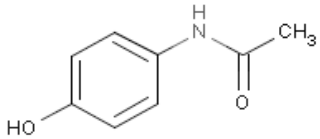
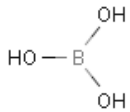
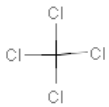
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
2000< LD_{50} ≤5000 mg/kg							
Acetaminophen	2404	MEIC, EDIT, RC, TESS, NTP	Pharmaceutical (analgesic)	151.20	0.8	Organic compound; Amide	
Potassium I chloride	2602	MEIC, RC, TESS, NTP	Pharmaceutical (electrolyte); Manufacturing	74.55	NA	Inorganic compound; Potassium compound; Chlorine compound	K^+ Cl^-
Boric aid	2660*	TESS, EPA, NTP	Pesticide (insecticide)	61.83	NA	Inorganic compound; Boron compound; Acids	
Carbon tetrachloride	2799	MEIC, RC, TESS, NTP, HPV	Solvent	153.82	2.83	Organic compound; Halogenated hydrocarbon	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

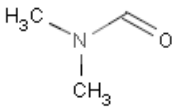
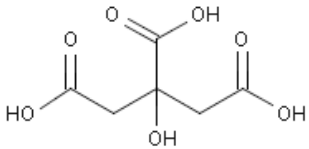
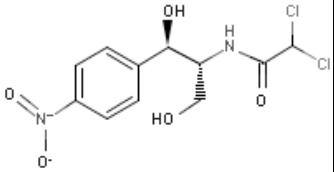
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Dimethylformamide	2800	RC, GD, NTP, HPV	Solvent	73.10	–1.01	Organic compound; Amide; Carboxylic acid	
Sodium chloride	2998	MEIC, EDIT, RC, TESS, EPA, NTP	Pharmaceutical (electrolyte); Food additive	58.44	NA	Inorganic compound; Sodium compound; Chlorine compound	Na^+ Cl^-
Citric Acid	3000*	EPA, NTP, HPV	Food additive	192.10	–1.72	Organic compound; Carboxylic acid	
Chloramphenicol	3393	MEIC, RC, NTP	Pharmaceutical (antibiotic)	323.14	1.14	Organic compound; Alcohol; Cyclic hydrocarbon; Nitro compound	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

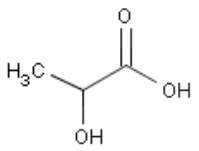
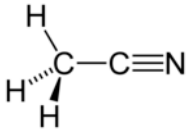
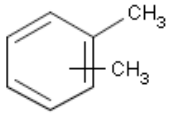
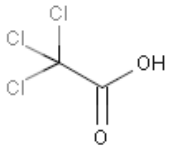
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Lactic acid	3730	RC, NTP, HPV	Food additive	90.08	–0.72	Organic compound; Carboxylic acid	
Acetonitrile	3798	RC, NTP, HPV	Solvent	41.05	–0.34	Organic compound; Nitrile	
Xylene (mixed isomers)	4300	MEIC, RC, TESS, NTP, HPV	Solvent	106.17	3.12 – 3.2	Organic compound; Cyclic hydrocarbon	
Trichloroacetic acid	4999	RC, NTP	Fixative	163.40	1.33	Organic compound; Carboxylic acid	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

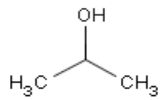
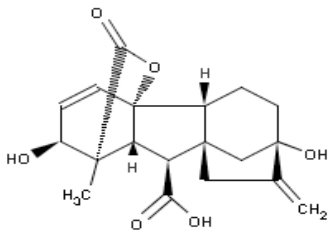
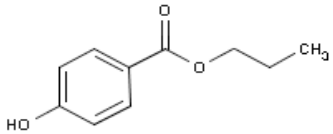
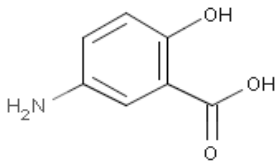
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
LD_{50} >5000 mg/kg							
2-Propanol	5843	MEIC, RC, TESS, EPA, NTP, HPV	Disinfectant	60.10	0.05	Organic compound; Alcohol	
Gibberellic acid	6305	RC, EPA, NTP	Plant growth regulator	346.38	0.24	Organic compound; Polycyclic compound	
Propylparaben	6326**	RC (outlier), NTP	Food additive	180.20	3.04	Organic compound; Carboxylic acid; Phenol	
5-Aminosalicylic acid	7749**	RC (outlier), NTP	Pharmaceutical (antibiotic)	153.10	1.32	Organic compound; Carboxylic acid; Phenol	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

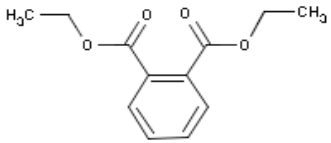
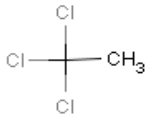
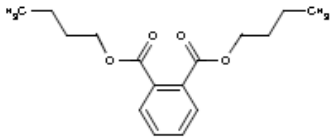
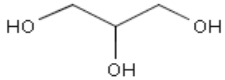
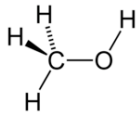
GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Ethylene glycol	8567	MEIC, EDIT, RC, TESS, NTP, HPV	Antifreeze	62.07	–1.36	Organic compound; Alcohol	
Diethyl phthalate	8602	RC (outlier), NTP, HPV	Plasticizer	222.20	2.47	Organic compound; Carboxylic acid	
Sodium hypochlorite	8910[8]	TESS, NTP	Disinfectant	74.44	NA	Inorganic compound; Sodiumcompound; Oxygen compound; Chlorine compound	
1,1,1-Trichloroethane	10298	MEIC, RC, NTP, HPV	Solvent	133.41	2.49	Organic compound; Halogenated hydrocarbon	

Table 3-2 Reference Substances Used in the 3T3 and NHK NRU Test Methods Validation Study - Sorted by Toxicity

GHS Category[1]/Substance	Rodent Oral LD_{50}[2] (mg/kg)	Source[3]	Product/Use[4]	Molecular Weight (g/mol)	log Kow[5]	Chemical Class[6]	Molecular Structure
Dibutyl phthalate	11998	RC (outlier), NTP, HPV	Plasticizer	278.30	4.9	Organic compound; Carboxylic acid	
Glycerol	12691	RC, GD, NTP, HPV	Solvent	92.09	-1.76	Organic compound; Alcohol	
Methanol	13012	MEIC, EDIT, RC, TESS, NTP, HPV	Solvent	32.04	–0.77	Organic compound; Alcohol	
Ethanol	14008	MEIC, RC (outlier), TESS, EPA, NTP, HPV	Solvent	46.07	–0.31	Organic compound; Alcohol	

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); LD_{50}=Dose that produces lethality in 50% of the test animals; K_{ow}=Octanol:water partition coefficient; EDIT=Evaluation-guided Development of New *In vitro* Test Batteries (substances in EDIT program are a subset of the MEIC substance set); EPA=Pesticides registered with the

Environmental Protection Agency; EHS=EPA's Extremely Hazardous Substance list; HPV=High Production Volume chemicals (i.e., those that are imported into or produced in the United States in amounts ≥1,000,000 lbs/year); GD=*Guidance Document* (ICCVAM 2001b); MEIC=Multicentre Evaluation of *In Vitro* Cytotoxicity; NA=Non applicable; NTP=National Toxicology Program; RC=Registry of Cytotoxicity with the chemicals classified as regression outliers shown in parentheses; TESS=Toxic Exposure Surveillance System (Litovitz et al. 2000); HSDB=Hazardous Substances Data Bank; RTECS®=Registry of Toxic Effects of Chemical Substances.

[*]From RTECS® (MDL Information Systems 2002).

[**]Mouse.

[1]GHS category designation for the substance (e.g., LD_{50} <5 mg/kg)

[2]LD_{50} data are from the Registry of Cytotoxicity (Halle 1998, 2003) and are for rats, unless otherwise noted. The LD_{50} values are rounded to the nearest whole number.

[3]Sources used to identify candidate chemicals.

[4]Product/use categories from HSDB (NLM 2002) or RTECS®(MDL Information Systems 2002). Pharmaceutical uses from Gilman et al. (1985) or Thomson PDR® (2004).

[5]From HSDB (NLM 2001, 2002) or Material Safety Data Sheets.

[6]Based on Medical Subject Heading [MeSH®] descriptors (NLM 2005).

[7]Mouse data for lithium sulfate (Halle 1998, 2003).

[8]From HSDB (NLM 2002).

213 The other major source of reference substances was the RC, which contributed 58 (81%) of
214 the 72 chemicals, as shown in **Table 3-4**. Because the RC millimole regression was used to
215 identify outlier substances (see **Section 6.2**), the fit of the RC substances to this regression
216 was relevant (Halle 1998, 2003). Halle (1998, 2003) defined outliers as those chemicals with
217 log IC_{50}-log LD_{50} points that were >0.699 (i.e., log 5) from the RC millimole regression.
218 **Table 3-4** shows the number of RC outliers selected for testing and the corresponding
219 number of outliers in the RC. Although the percentage of outliers in several GHS categories
220 is similar to the percentage in the RC, the total percentage of RC outliers in the set of
221 reference substances (i.e., 38% [22/58]) is greater than the percentage in the RC (i.e., 27%
222 [95/347]). This occurred because the fit to the RC millimole regression was not the major
223 deciding factor during selection of the 72 reference substances.
224
225 Among the 58 RC substances selected for use in the validation study, 22 (38%) were outliers
226 for the RC millimole regression. Toxicity[1] was underpredicted for 17 (77%) of these outlier
227 substances and overpredicted (i.e., predicted LD_{50} was lower than measured *in vivo* LD_{50}) for
228 the remaining five (23%). For the 95 outlier substances in the RC, the number of substances
229 for which toxicity was over- or under-predicted was approximately the same. Toxicity was
230 underpredicted for 49 (52%) outliers and overpredicted for 46 (48%) outliers (Halle 1998,
231 2003). **Figure 3-1** shows the 58 RC chemicals selected for testing, in addition to the 289 RC
232 chemicals that were not selected, and the RC millimole regression. In the figure, the outliers
233 are those points outside the RC prediction interval. For the 58 RC substances selected for
234 testing, the majority (17/22) of the outliers are below the RC millimole regression line.
235

[1] Toxicity is inversely proportional to LD_{50}. High LD_{50} values reflect low toxicity and low LD_{50} values reflect high toxicity

236 **Table 3-3 Distribution of Candidate Substances and Reference Substances by Source[1] and Toxicity Category**
237

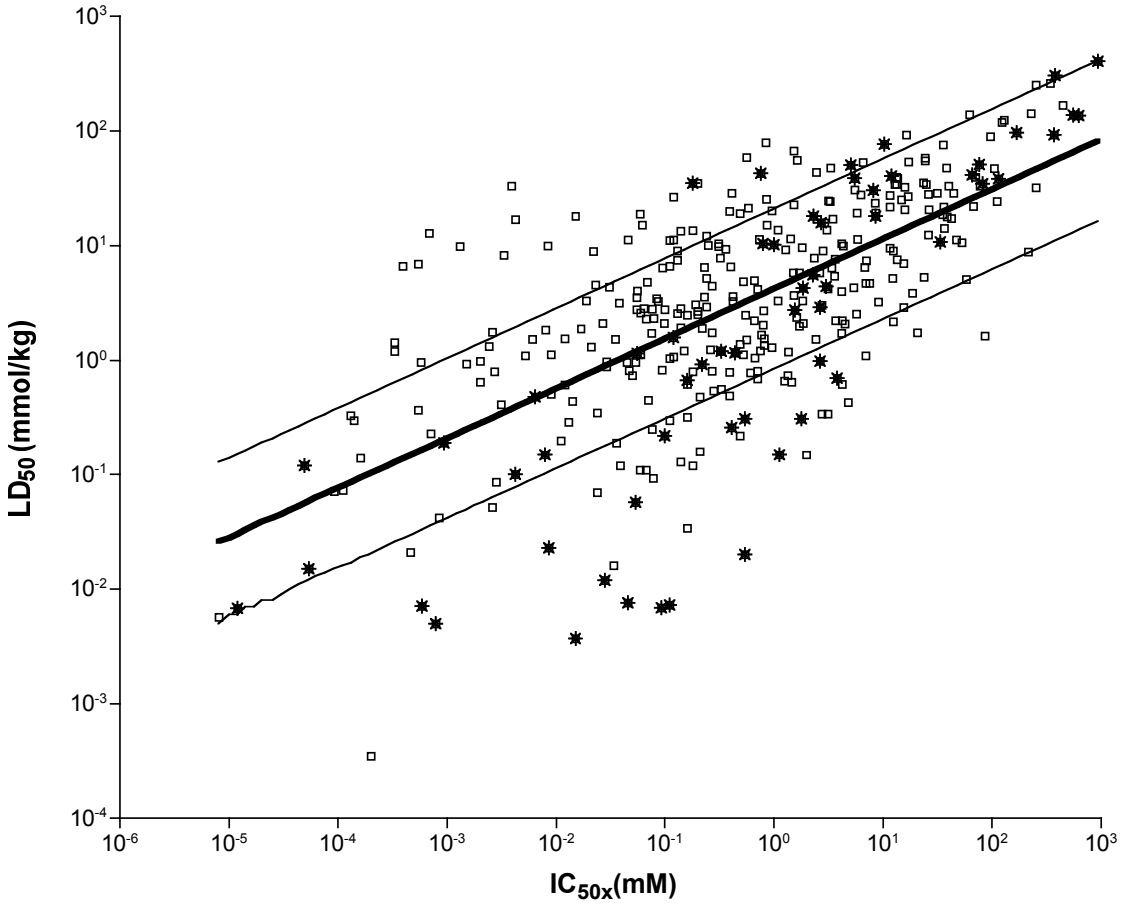
GHS Category (mg/kg)	Reference Substances/ Candidate Substances	MEIC Reference/ MEIC Candidates	EDIT Reference/ EDIT Candidates	TESS Reference/ TESS Candidates	NTP Reference/ NTP Candidates	HPV Reference/ HPV Candidates
$LD_{50} \leq 5$	12/13	2/2	1/1	3/3	5/9	0/0
$5 < LD_{50} \leq 50$	12/15	6/6	5/5	9/10	8/11	2/5
$50 < LD_{50} \leq 300$	12/26	11/17	4/5	11/19	9/18	1/3
$300 < LD_{50} \leq 2000$	12/38	12/29	3/5	12/27	5/23	1/5
$2000 < LD_{50} \leq 5000$	12/12	6/6	2/2	6/6	12/12	6/6
$LD_{50} > 5000$	12/12	5/5	2/2	5/5	12/12	8/8
Total	72/116	42/65	17/20	46/70	51/85	18/27

238 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); LD_{50}=Dose that produces lethality in 50% of the test animals;
239 MEIC=Multicentre Evaluation of *In Vitro* Cytotoxicity**;** EDIT=Evaluation-Guided Development of *In vitro* Tests; TESS=Toxic Exposure Surveillance System; NTP=U.S.
240 National Toxicology Program; HPV=U.S. Environmental Protection Agency (EPA) High Production Volume program.
241 [1]Substances may have been selected from more than one source (see **Table 3-2** and **Appendix F3**).
242
243
244 **Table 3-4 Selected Substances: Distribution of RC Chemicals and RC Outliers[1] by Toxicity Category**
245

GHS Category (mg/kg)	RC Outliers/ Total Chemicals	Candidate and Selected Substances		
		Candidate Substances	RC Reference / RC Candidates	RC Reference Outliers/ RC Reference Chemicals
$LD_{50} \leq 5$	10/11 (91%)	13	9/10	8/9 (89%)
$5 < LD_{50} \leq 50$	15/26 (58%)	15	8/10	4/8 (50%)
$50 < LD_{50} \leq 300$	24/70 (34%)	26	11/18	5/11 (45%)
$300 < LD_{50} \leq 2000$	14/139 (10%)	38	9/29	0/9 (0%)
$2000 < LD_{50} \leq 5000$	12/57 (21%)	12	10/10	0/10 (0%)
$LD_{50} > 5000$	20/44 (45%)	12	11/11	5/11 (45%)
Total	95/347 (27%)	116	58/88	22/58 (38%)

246 Abbreviations: RC=Registry of Cytotoxicity; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); LD_{50}=Dose
247 that produces lethality in 50% of the test animals.
248 [1]Chemicals falling outside the log 5 (i.e., > ±0.699) prediction interval for the RC millimole regression (Halle 1998, 2003).
249

Figure 3-1 The Fifty-Eight (58) Selected RC Reference Substances on the RC Regression



Abbreviations: RC=Registry of Cytotoxicity; LD_{50}=dose that produces lethality in 50% of the test animals; IC_{50}=Test concentration that reduces cell viability by 50%.
The 58 RC chemicals tested in the NICEATM/ECVAM validation study are shown by *. The RC regression, log (LD_{50}) = 0.435 x log (IC_{50x}) + 0.625, is shown by the bold line. The lighter lines show the ± log 5 (i.e., ±0.699) prediction interval (Halle 1998, 2003). The open boxes represent the 289 chemicals not included in the validation study.

3.2.2 Chemical Classes Represented by the Selected Reference Substances

Medical subject heading (MeSH®) descriptors from the NLM were used to determine chemical class designations for the selected substances. Of the 72 reference substances, 57 (79%) were organic and 15 (21%) were inorganic. The number of substances in the organic (79) and inorganic (31) subclasses is greater than the number of substances in each class because some of the substances are classified in more than one subclass. The most commonly represented classes of organic compounds were heterocyclics (14/57, 25%), carboxylic acids

268 (14/57, 25%), and alcohols (10/57, 18%). **Table 3-5** shows the distribution of the substances
269 among the GHS toxicity categories. The 14 heterocyclics were evenly distributed among the
270 first four GHS toxicity categories for LD_{50} ≤2000 mg/kg with the majority of the
271 heterocyclics (11/14) in the categories for LD_{50} <300 mg/kg. The majority of the carboxylic
272 acids (12/14) and alcohols (8/10) had an LD_{50} >300 mg/kg, while the majority of the
273 inorganics (10/15) had an LD_{50} <300 mg/kg.

274 3.2.3 Product/Use Classes Represented by the Selected Reference Substances

275 Product and use information was obtained from HSDB (NLM 2002) or RTECS® (MDL
276 Information Systems 2002). The number of assigned uses (77) is greater than the number of
277 selected substances because some of the substances have more than one use. **Table 3-6**
278 shows the distribution of products and uses of the selected substances according to their GHS
279 categories. Pharmaceutical (27/77; 35%) and pesticide (17/77; 22%) uses were observed
280 most frequently. The toxicity category of 300 < LD_{50} ≤2000 mg/kg had the highest number
281 of pharmaceuticals. Every toxicity category except LD_{50} >5000 mg/kg had at least four
282 substances with pharmaceutical uses. The majority of pesticides (16/17; 94%) had an LD_{50}
283 <300 mg/kg. The next most frequent uses were as solvents (8/77; 10%) and food additives
284 (5/77; 6%); LD_{50} >2000 mg/kg contained most of the substances with solvent (8/8; 100%)
285 and food additive (4/5; 80%) uses.

286 3.2.4 Toxicological Characteristics of the Selected Reference Substances

287 3.2.4.1 *Corrosivity*

288 The intent of the SMT was to prioritize only those substances with low corrosivity because
289 guidelines for acute systemic toxicity testing indicate that corrosive or severely irritating
290 substances need not be tested (OECD 2001a, c, d). The UN and U.S. Department of
291 Transportation Packing Group (DOT PG) classification system was used to classify the
292 corrosivity hazard associated with the candidate substances. However, after substance
293 selection was completed and testing had begun, the SMT learned that the PG classification
294 system was also based on hazards other than corrosivity (e.g., dermal and inhalation toxicity,
295 flammability, etc.). Therefore, the selected substances were not actually prioritized by
296 corrosivity. Subsequent information on the corrosivity of the selected substances was
297 obtained from HSDB (NLM 2004) and the Material Safety Data Sheets (MSDS) provided

298 with the purchased substances. Seven substances that were not identified by the DOT PG
299 classification system had corrosive notations. The MSDS notations for lactic acid, sodium
300 hypochlorite, sodium oxalate, and trichloroacetic acid indicated that these substances should
301 carry a corrosive label. Chloral hydrate, mercury II chloride, and potassium cyanide were
302 noted by HSDB to be corrosive to eyes or skin.

303 3.2.4.2 *Toxicity Targets*

304 As shown in **Appendix F2**, the most common toxicological effects in humans or rodents
305 were neurological (40 substances); 26 cause central nervous system (CNS) depression, seven
306 produce CNS stimulation, four produce CNS affects such as encephalopathy, and three affect
307 the peripheral nervous system. Other common target systems include the liver (17
308 substances), kidney (15 substances), and cardiovascular system (10 substances). No target
309 organ information was available for gibberellic acid. Among the 72 reference substances, 27
310 had more than one toxicity target.
311

312

Table 3-5 Distribution of Chemical Class for the 72 Reference Substances by Toxicity Category

Chemical Class[1]	GHS Acute Oral Toxicity Category (mg/kg)						Total
	$LD_{50} \leq 5$	$5 < LD_{50} \leq 50$	$50 < LD_{50} \leq 300$	$300 < LD_{50} \leq 2000$	$2000 < LD_{50} \leq 5000$	$LD_{50} > 5000$	
Organic							
Carboxylic acid	1	0	1	4	4	4	14
Heterocyclic compound	5	2	4	3	0	0	14
Alcohol	2	0	0	2	1	5	10
Phenol	0	0	1	2	0	2	5
Polycyclic compound	0	2	0	2	0	1	5
Sulfur compound	4	1	0	0	0	0	5
Amine	1	0	1	1	0	0	3
Cyclic hydrocarbon	0	0	1	0	1	1	3
Halogenated hydrocarbon	0	0	1	0	1	1	3
Organophosphorous compound	2	1	0	0	0	0	3
Amide	0	0	0	1	2	0	3
Nitrile	0	1	0	0	1	0	2
Acyclic hydrocarbon	1	0	0	0	0	0	1
Carbohydrate	0	1	0	0	0	0	1
Ester	0	1	0	0	0	0	1
Ether	0	1	0	0	0	0	1
Ketone	0	0	1	0	0	0	1
Lipid	0	0	0	1	0	0	1
Nitro compound	0	0	0	0	1	0	1
Organometallic compound	0	1	0	0	0	0	1
Sodium compound	0	0	1	0	0	0	1
Urea	1	0	0	0	0	0	1
Total Organics	17	11	11	16	11	14	79

Table 3-5 Distribution of Chemical Class for the 72 Reference Substances by Toxicity Category

Chemical Class[1]	GHS Acute Oral Toxicity Category (mg/kg)						Total
	$LD_{50} \leq 5$	$5 < LD_{50} \leq 50$	$50 < LD_{50} \leq 300$	$300 < LD_{50} \leq 2000$	$2000 < LD_{50} \leq 5000$	$LD_{50} > 5000$	
Inorganic							
Sodium compound	1	2	1	0	1	1	6
Chlorine compound	1	0	1	0	2	1	5
Arsenical	0	2	0	0	0	0	2
Metal	0	1	1	0	0	0	2
Potassium compound	0	1	0	0	1	0	2
Sulfur compound	0	1	1	0	0	0	2
Acid	0	0	0	0	1	0	1
Alkalies	0	0	1	0	0	0	1
Boron compound	0	0	0	0	1	0	1
Cadmium compound	0	0	1	0	0	0	1
Carbon compound	0	0	0	1	0	0	1
Chromium compound	0	1	0	0	0	0	1
Fluorine compound	0	0	1	0	0	0	1
Lithium compound	0	0	0	1	0	0	1
Mercury compound	1	0	0	0	0	0	1
Nitrogen compound	0	1	0	0	0	0	1
Oxygen compound	0	0	0	0	0	1	1
Selenium compound	1	0	0	0	0	0	1
Total Inorganic	4	9	7	2	6	3	31

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
[1]Based on the Medical Subject Heading [MeSH®] descriptor (NLM 2005). Some substances are counted more than once because they appear in more than one subclass under the organic or inorganic classes.

316 **Table 3-6 Distribution of Product/Use[1] Class for the 72 Reference Substances by Toxicity Category**

Product/Use Class[1]	GHS Acute Oral Toxicity Category (mg/kg)						Total
	$LD_{50} \leq 5$	$5 < LD_{50} \leq 50$	$50 < LD_{50} \leq 300$	$300 < LD_{50} \leq 2000$	$2000 < LD_{50} \leq 5000$	$LD_{50} > 5000$	
Antibiotic/fungicide	1	0	0	0	0	0	1
Antifreeze	0	0	0	0	0	1	1
Consumer/industrial products	0	0	1	0	0	0	1
Disinfectant	0	0	1	1	0	2	4
Electroplating	0	2	0	0	0	0	2
Fluoridation	0	0	1	0	0	0	1
Feed additive	1	0	0	0	0	0	1
Fixative	0	0	0	0	1	0	1
Food additive	0	0	1	0	3	1	5
Manufacturing	1	0	0	0	1	0	2
Oxidizing agent	0	1	0	0	0	0	1
Paints, cleaners	0	0	1	0	0	0	1
Pesticide	5	7	4	0	1	0	17
Pharmaceutical	4	3	4	11	4	1	27
Plant growth regulator	0	0	0	0	0	1	1
Plasticizer	0	0	0	0	0	2	2
Preservative	1	0	0	0	0	0	1
Solvent	0	0	0	0	4	4	8

317 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
318 [1]Product/use information from Hazardous Substances Data Bank (NLM 2002) or Registry of Toxic Effects of Chemical Substances ([RTECS®], MDL Information Systems 2002).
319 Some substances are counted more than once because they appear in more than one use category.

320 3.2.4.3 *Metabolism*

321 **Table 3-7** shows the 22 reference substances that are known or expected to produce
322 active/toxic metabolites *in vivo*. In contrast, dichlorvos, fenpropathrin, meprobamate,
323 phenylthiourea, and sodium dichromate are rapidly metabolized to less toxic compounds.
324 Because the NHK and 3T3 cells have little (Babich 1991) or no (INVITTOX 1991)
325 metabolic capability, respectively, metabolites of these compounds would not be expected to
326 be present *in vitro*. **Appendix F2** provides for more information on the metabolism
327 (activation/inactivation) of the selected reference substances.
328

329 **Table 3-7 Reference Substances Metabolized to Active Metabolites**
330

Known to Have Active Metabolites				Active Metabolites Expected
Acetaminophen	Carbamazepine	Digoxin	Methanol	Carbon tetrachloride
Acetonitrile	Chloral hydrate	Disulfoton	Parathion	Triethylenemelamine
Acetylsalicylic acid	Cycloheximide	Ethanol	Procainamide HCl	Valproic acid
Amitriptyline HCl	Dibutyl phthalate	Ethylene glycol	Verapamil HCl	
Busulfan	Diethyl phthalate	Glutethimide		

331

332 3.2.5 Selection of Reference Substances for Testing in Phases Ib and II

333 Based on the *Guidance Document* (ICCVAM 2001b) recommendation that 10 to 20
334 substances be tested to qualify candidate *in vitro* cytotoxicity tests for determining starting
335 doses for rodent acute oral toxicity assays, 12 reference substances were chosen from among
336 the 72 reference substances for testing in Phases Ib and II (see **Table 3-8**). The criteria for
337 choosing these reference substances, in order of importance, were:

338 • Two reference substances must be included from each of the five GHS
339 toxicity categories and the unclassified category.
340 • The log LD_{50} (mmol/kg) must be within the prediction interval (±0.699) of the
341 RC millimole regression. The *Guidance Document* (ICCVAM 2001b)
342 recommends that reference substances for evaluating an *in vitro* basal
343 cytotoxicity test to use with the RC millimole regression fit the regression as
344 closely as possible.

- MEIC chemicals must be included. Cytotoxicity data from these phases (and Phase III of this study), and the available human toxicity information for the MEIC chemicals, could be used to build a prediction model for estimating human lethal blood concentrations. The Phase Ib reference substances arsenic trioxide and ethylene glycol are also EDIT chemicals (subset of MEIC chemicals).

If more than two substances in a GHS category met the above criteria, reference substances were selected so that the LD_{50} was as close to the RC millimole regression as possible and/or to represent the full range of toxicity in each GHS category.

Only nine of the 72 reference substances met all three criteria. In the most toxic category (i.e., $LD_{50} \leq 5$ mg/kg), only one RC chemical, aminopterin, was within 0.699 of the RC millimole regression. Sodium selenate was selected as the second reference substance in this category even though its fit to the RC millimole regression was not known. Neither aminopterin nor sodium selenate were MEIC chemicals. For the $50 < LD_{50} \leq 300$ mg/kg category, cadmium chloride was selected over the MEIC chemicals cupric sulfate $5H_2O$, diquat dibromide, sodium oxalate, and hexachlorophene because it fit the RC millimole regression better than the four MEIC chemicals (the observed LD_{50} minus log predicted LD_{50} values from -0.534 to -0.337).

3.2.6 Unsuitable and Challenging Reference Substances

Several reference substances could not be adequately tested for cytotoxicity in 3T3 cells and/or NHKs in from one to all three of the laboratories. The following reference substances did not produce sufficient toxicity at soluble concentrations for calculation of an IC_{50} at the highest concentrations tested under the testing conditions used in the study (see also **Tables 5-2, 5-4,** and **5-5**):

Table 3-8 Reference Substances Tested in Phases Ib and II

Reference Substances	CASRN	RC Reference No.	MEIC Reference No.	Rodent Oral LD_{50}[1] (mg/kg)	Observed – Predicted log LD_{50}[2]
LD_{50} ≤5 mg/kg					
Aminopterin	54-62-6	3	NA	3	-0.652
Sodium selenate	13410-01-0	NA	NA	1.6[3]	NA
5 < LD_{50} ≤50 mg/kg					
Colchicine	64-86-8	6	60	6[4]	-0.593
Arsenic III trioxide	1327-53-3	153	26	20	-0.591
50 < LD_{50} ≤300 mg/kg					
Cadmium II chloride	10108-64-2	81	NA	88	0.011
Sodium I fluoride	7681-49-4	106	14	180	-0.109
300 < LD_{50} ≤2000 mg/kg					
DL-Propranolol HCl	350-60-90	54	23	470[4]	-0.023
Lithium I carbonate	544-13-2	327[4]	20	1187[4,5]	-0.256[4]
2000 < LD_{50} ≤5000 mg/kg					
Potassium I chloride	7447-40-7	346	50	2602	0.085
Chloramphenicol	56-75-7	91	45	3393	0.441
LD_{50} >5000 mg/kg					
2-Propanol	67-63-0	128	10	5843	0.396
Ethylene glycol	107-21-1	360	7	8567	0.321

Abbreviations: CASRN=Chemical Abstracts Service Registry Number; RC=Registry of Cytotoxicity; MEIC=Multicentre Evaluation of *In Vitro* Cytotoxicity; NA=Not applicable (i.e., substances not included in the RC and/or MEIC studies); RTECS®=Registry of Toxic Effects of Chemical Substances.
[1]From the RC (Halle 1998, 2003) unless otherwise indicated. Data are for rats unless otherwise indicated.
[2]Available only for substances included in the RC. This figure characterizes the log LD_{50} deviation from the RC regression. Outliers are > ±0.699 from the regression line.
[3]RTECS® (MDL Information Systems 2002).
[4]Mouse data.
[5]For lithium sulfate.

- Carbon tetrachloride (no 3T3 or NHK NRU IC_{50} data from ECBC, FAL, or IIVS)
- Xylene (no 3T3 or NHK NRU IC_{50} data from ECBC or FAL)
- Methanol (no 3T3 NRU IC_{50} data from ECBC, FAL, or IIVS; no NHK NRU IC_{50} data from ECBC)
- Lithium carbonate (no 3T3 NRU IC_{50} data from FAL or IIVS)
- 1,1,1-Trichloroethane (no 3T3 NRU IC_{50} data from FAL or IIVS; no NHK NRU IC_{50} data from ECBC)
- Valproic acid (no 3T3 NRU IC_{50} data from ECBC or FAL; no NHK NRU IC_{50} data from ECBC, FAL, or IIVS)

395 Other reference substances were difficult to test because of volatility or lack of toxicity, but
396 three acceptable tests could be obtained after a number of trials.

397 • Acetonitrile and 2-propanol were highly volatile and nontoxic, so that even
398 with the use of film plate sealers, from one to seven tests failed the VC and
399 data points test acceptance criteria at each laboratory.
400 • Disulfoton failed at least one test in both test methods at ECBC and FAL
401 because of inadequate toxicity (i.e., an IC_{50} could not be detected) and
402 insolubility. All laboratories reported precipitate in the test plates for 3T3 and
403 NHK NRU tests. IIVS had no failed tests in either test method.
404 • Dibutyl phthalate failed one 3T3 NRU test at ECBC and one NHK NRU test
405 at FAL because of inadequate toxicity and solubility.
406 • Lindane failed one 3T3 NRU test at FAL because of inadequate toxicity and
407 solubility and one because of its volatility.
408 • Parathion failed one test because of inadequate toxicity and solubility in both
409 test methods and one NHK NRU test because of volatility at FAL.
410 • Diethyl phthalate failed one NHK NRU test because of volatility at FAL.
411 • Digoxin (all laboratories), gibberellic acid (ECBC and FAL), and strychnine
412 (ECBC and FAL) failed at least one 3T3 NRU test because of inadequate
413 toxicity and solubility.

414 **3.3 Reference Substance Procurement, Coding, and Distribution**

415 BioReliance collected information from the suppliers of the reference substances on their
416 analytical purity, composition, and stability (see **Appendix F1**), tested the reference
417 substances for solubility, packaged them into 4 g aliquots for shipment to the testing
418 laboratories, and archived two additional samples. All reference substances were given a
419 random number code that was unique for each testing facility to conceal the identities from
420 the testing laboratories. Approximately 100 g of the positive control substance, SLS, was
421 distributed, uncoded, to each laboratory and one additional sample was archived.
422
423 Reference substances were packaged so as to minimize damage during transit, and shipped
424 under appropriate storage conditions and according to the appropriate regulatory
425 transportation procedures. Testing facilities were notified upon shipment in order to prepare

426 for receipt. With the exception of the positive control substance which was shipped directly
427 to the Study Directors, the reference substances were shipped to the test facility Safety
428 Officers. Shipments were accompanied by a sealed information packet containing the
429 appropriate health and safety procedures (i.e., MSDS or equivalent documentation with
430 information regarding the proper protection for handling, procedures for dealing with
431 accidental ingestion or contact with skin or eyes, and for containing and recovering spills),
432 and a code disclosure key. Also provided was a data sheet giving a minimum of essential
433 information needed by the testing laboratory for each reference substance, including color,
434 odor, physical state, weight or volume of sample, specific density for liquid reference
435 substances, and storage instructions. The shipment directed the Safety Officer to:

436 • Notify BioReliance and the SMT upon receipt of reference substances
437 • Retain the health and safety package and provide the coded reference
438 substances and chemical data sheets with minimum essential information to
439 the laboratory Study Director without revealing the identities of the test
440 substances
441 • Notify the SMT if test facility personnel open the health and safety packet at
442 any time, for any reason, during the study
443 • Return the unopened health and safety package to BioReliance after testing is
444 completed

445 3.3.1 Exceptions

446 The Safety Officer for ECBC required the information on reference substance codes before
447 the substances were shipped in order to satisfy the facility's environmental procedures and
448 requirements. The reference substance codes were stored in a classified safe located in the
449 Safety Office which was in a building separate from the cytotoxicity testing laboratory, and
450 were to be opened only by the Safety Officer. The ECBC Safety Officer opened the sealed
451 health and safety packets for lithium carbonate and ethanol upon receipt of those substances
452 because the code information for these substances was not included in the list originally
453 provided. ECBC cytotoxicity testing personnel did not have direct access to the reference
454 substance codes.

455

456 **3.4 Reference Substances Recommended by the *Guidance Document***

457 The *Guidance Document* specifically recommended testing the following 11 substances to
458 validate candidate *in vitro* basal cytotoxicity assays: sodium dichromate dihydrate, cadmium
459 chloride, *p*-phenylenediamine, DL-propranolol HCl, trichlorfon, ibuprofen, nalidixic acid,
460 salicylic acid, antipyrene, dimethylformamide, and glycerol (ICCVAM 2001b). Of these 11
461 substances (see **Appendix F3** and **Section 3.1.2**), five (sodium dichromate dihydrate,
462 cadmium chloride, DL-propranolol HCl, dimethylformamide, and glycerol) were chosen for
463 testing after the candidate substances were prioritized as described in **Section 3.1.3**. The
464 seven that were not selected did not satisfy the selection criteria (e.g., not MEIC chemicals,
465 not identified as high exposure risk in TESS)

466 **3.5 Summary**

467 Seventy-two reference substances were selected for testing in the NICEATM/ECVAM
468 validation study. These substances were selected to represent: (1) the complete range of *in*
469 *vivo* acute oral LD_{50} values; (2) the types of substances regulated by the various regulatory
470 authorities; and (3) those with human toxicity data and/or human exposure potential. To
471 insure that the complete range of toxicity was covered, the GHS (UN 2005) was used to
472 select 12 substances for each acute oral toxicity category and 12 unclassified substances. The
473 set of selected reference substances had the following characteristics:

474 • Thirty-five percent (27/77 uses) were pharmaceuticals, 22% (17/77 uses) were
475 pesticides, 10% (8/77 uses) were solvents, and 6% (5/77 uses) were food
476 additives. The remaining substances were used for a variety of manufacturing
477 and consumer products.
478 • In terms of relevance of the substances to human exposure, 58% (42/72) were
479 included in the MEIC study (substances chosen because of availability of
480 human lethality data), 24% (17/72) of which were included also in the EDIT
481 program (EDIT substances a subset of the MEIC substances), 64% (46/72)
482 had human exposure data reported by TESS, 71% (51/72) had been evaluated
483 by NTP, and 25% (18/72) were on the EPA HPV list.
484 • Eighty-one percent (58/72) of the substances were in the RC and 38% (22/58)
485 of these were outliers with respect to the RC millimole regression. The RC

486 millimole regression underpredicted the toxicity of 77% (17/22) of the outliers
487 and overpredicted the toxicity of 23% (5/22). For the 95 outlier substances in
488 the RC, however, the number of substances for which toxicity was over- or
489 under-predicted was approximately the same (i.e., toxicity was underpredicted
490 for 49 [52%] outliers and overpredicted for 46 [48%] outliers [Halle 1998,
491 2003]).

492 • Seventy-nine percent (57/72) were organic compounds and 21% (15/72) were
493 inorganic. The most commonly represented classes of organic compounds
494 were heterocyclics (25%, 14/57), carboxylic acids (25%, 14/56), and alcohols
495 (18%, 10/57).

496 • Nineteen substances (26%, 19/72,) were known to have active metabolites and
497 three others were expected to have active metabolites based on their chemical
498 structures.

499 • Many of the substances produced toxicity in more than one organ system. The
500 most common target systems were neurological (40 substances), liver (17
501 substances), kidney (15 substances), and cardiovascular (10 substances). No
502 target organ information was available for one substance (gibberellic acid).
503

[This Page Intentionally Left Blank]

4.0 RODENT ACUTE ORAL LD_{50} REFERENCE VALUES USED TO ASSESS THE ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS 4-3

4.1 Methods Used to Obtain Rodent Acute Oral LD_{50} Reference Values 4-3
4.1.1 Identification of Candidate Rodent Acute Oral LD_{50} Reference Data 4-3
4.1.2 Criteria Used to Select Candidate Rodent Acute Oral Data for Determination of LD_{50} Reference Values 4-6

4.2 Final Rodent Acute Oral LD_{50} Reference Values 4-9

4.3 Relevant Toxicity Information for Humans 4-14

4.4 Accuracy and Reliability of the Rodent Acute Oral LD_{50} Reference Values 4-15
4.4.1 Variability Among the Acceptable LD_{50} Values 4-15
4.4.2 Comparison of Rodent Acute Oral LD_{50} Reference Values with the Corresponding RC LD_{50} Values 4-16
4.4.3 Comparison of the Variability Among Acceptable LD_{50} Values to Those Obtained in Other Studies 4-18

4.5 Summary.....4-19

[This Page Intentionally Left Blank]

4.0 RODENT ACUTE ORAL LD_{50} REFERENCE VALUES USED TO ASSESS THE ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS

The procedures and analyses presented in this section were designed to identify the most accurate rodent acute oral LD_{50} values for the 72 reference substances used in the validation study. These values were needed to ensure that the references substances were correctly placed within the different GHS toxicity categories and to provide a data set against which to compare the predicted LD_{50} values estimated using the IC_{50} data obtained from the 3T3 and NHK NRU test methods (see **Section 6**). The predicted LD_{50} values are used to determine the starting dose for rodent acute oral toxicity tests and the more accurate the prediction, the fewer the number of rodents that would be used in an acute oral toxicity test (see **Sections 1.0** and **1.2.2**).

4.1 Methods Used to Obtain Rodent Acute Oral LD_{50} Reference Values

4.1.1 Identification of Candidate Rodent Acute Oral LD_{50} Reference Data

No animal testing was performed to obtain the rodent oral acute LD_{50} reference data for this validation study. To identify reference data for the 72 substances, rat acute oral LD_{50} studies were located using literature searches, secondary references, and electronic database searches. Literature searches were conducted in PubMed (U.S. NLM) and the Institute of Scientific Information (ISI) Web of Science® (Thomson Scientific, Philadelphia, PA) using each chemical name and "lethal dose 50" as search terms. Secondary references included NTP technical reports, Toxicological Profiles from the Agency for Toxic Substances and Disease Registry (ATSDR), Cosmetic Ingredient Reviews by the Cosmetics Industry Council, pesticide handbooks, the Merck Index, and various other summary sources. **Table 4-1** lists the electronic databases searched to locate references for rat oral LD_{50} values. Rat LD_{50} data were preferred because:

- The current acute oral toxicity test guidelines recommend using rats (OECD 2001a, c, d; EPA 2002a)
- The majority of LD_{50} data used in the RC millimole regression were from studies using rats (282 rat data points and 65 mouse data points) (Halle 1998, 2003)
- The majority of acute oral systemic toxicity testing is performed with rats

Table 4-1 Internet-Accessible Databases Searched for LD_{50} Information

Database/Source[1]	Sponsor(s)
Agency for Toxic Substances and Disease Registry (ATSDR)	U.S. Department of Health and Human Services (DHHS)
Center for Drug Evaluation and Research (CDER)	U.S. Food and Drug Administration (FDA)
CHEMFINDER	CambridgeSoft Corporation
Chemical Carcinogenesis Research Information System (CCRIS); National Cancer Institute (NCI) Website	NCI; National Institutes of Health (NIH); DHHS
Chemical Evaluation Search and Retrieval System (CESARS)	Michigan Department of Natural Resources; Ontario Ministry of the Environment; Canadian Centre for Occupational Health and Safety (CCOHS) CHEMpendium™
Chemical Hazard Response (CHRIS)	U.S. Coast Guard
Chemical Ingredients Database	U.S. Environmental Protection Agency (EPA) Office of Pesticide Programs (OPP); California EPA Department of Pesticide Regulation
CHEMINDEX; CHEMINFO	(CCOHS) CHEMpendium™
ChemRTK High Production Volume (HPV) Challenge Program; OPPT Chemical Fact Sheets; Chemical Information Collection and Data Development	EPA Office of Pollution Prevention and Toxics (OPPT)
CIS Chemical Information	World Health Organization (WHO) International Programme on Chemical Safety (IPCS); CCOHS; International Labour Organisation (ILO) Occupational Safety and Health Information Centre (CIS)
Concise International Chemical Assessment Documents (CICADS)	WHO IPCS; CCOHS; ILO; United Nations Environment Programme (UNEP)
Consumer Product Safety Commission Website	U.S. Consumer Product Safety Commission (CPSC)
Deutsches Institut fur Medizinische Dokumentation und Information (DIMDI) [The German Institute for Medical Documentation and Information]; Registry of Cytotoxicity (RC)	Zentralstelle zur Erfassung und Bewertungvon Ersatz- und Erganzungsmethoden zum Tierversuch (ZEBET) [German Centre for the Documentation and Validation of Alternative Methods]
Developmental and Reproductive Toxicology/Environmental Teratology Information Center (DART®/ETIC)	EPA; The National Library of Medicine (NLM); The National Institute of Environmental Health Sciences (NIEHS); National Center for Toxicological Research (NCTR)
Emergency Response Guidebook (ERG 2000)	Transport Canada; U.S. Department of Transportation (DOT); Secretariat of Communications and Transportation of Mexico
Environmental Health Criteria (EHC) monographs; Health and Safety Guides (HSG); International Agency for Research on Cancer (IARC)	WHO IPCS; CCOHS
European Centre for the Validation of Alternative Methods (ECVAM) Scientific Information Service (ECVAM SIS)	European Commission Joint Research Centre
HAZARDTEXT®; MEDITEXT®; INFOTEXT®; SARATEXT®; REPROTEXT®; REPROTOX®	TOMES Plus®, MICROMEDEX, Greenwood Village, CO
Integrated Risk Information System (IRIS)	EPA Office of Research and Development (ORD)
International Chemical Safety Cards (ICSC) IPCS/EC Evaluation of Antidotes Series	WHO IPCS; CCOHS; Commission of the European Union (EU)
International Uniform Chemical Information Database (IUCLID)	European Chemicals Bureau
Joint Expert Committee on Food Additives	WHO IPCS; CCOHS; Food and Agriculture Organization

Table 4-1 Internet-Accessible Databases Searched for LD_{50} Information

Database/Source[1]	Sponsor(s)
(JECFA); Joint Meeting on Pesticide Residues (JMPR); Pesticide Data Sheets (PDS)	(FAO) of the United Nations
Material Safety Data Sheets (MSDS)	Interactive Learning Paradigms, Incorporated
Multicentre Evaluation of In Vitro Cytotoxicity (MEIC)	Scandinavian Society for Cell Toxicology
The National MSDS Repository	MSDSSEARCH, Inc.
National Toxicology Program (NTP) Chemical Health and Safety Database	NIEHS
National Transportation Library	DOT
New Jersey Hazardous Substance Fact Sheets	New Jersey Department of Health and Senior Services
Oil and Hazardous Materials/Technical Assistance Data System (OHM/TADS)	EPA Office of Waste and Water Management
Organisation for Economic Co-operation and Development (OECD) Screening Information Data Sets (SIDS)	IPCS; CCOHS; International Register of Potentially Toxic Chemicals (IRPTC); UNEP
Pesticide Action Network Pesticide Database	Pesticide Action Network North America
Pesticide Product Information System (PPIS)	EPA Office of Pesticide Programs (OPP)
Poisons Information Monographs (PIMs)	IPCS; CCOHS
Registry of Toxic Effects of Chemical Substances (RTECS®);NIOSH Pocket Guide to Chemical Hazards	National Institute for Occupational Safety and Health (NIOSH)
SCORECARD	Environmental Defense
The EXtension TOXicology NETwork (EXTOXNET)	University of California, Davis; Oregon State University; Michigan State University; Cornell University; University of Idaho
The Right-to-Know Network (RTK NET)	Office of Management and Budget Watch; Center for Public Data access
Toxic Chemical Release Inventory (TRI); GENE-TOX	The National Library of Medicine (NLM)
Toxic Substances Control Act Test Submissions (TSCATS)	EPA OPPT
TOXLINE®; Hazardous Substances Data Bank (HSDB); ChemIDplus	NLM (TOXNET)

Abbreviations: LD_{50}=dose lethal to 50% of the animals tested
[1]Includes public and proprietary databases

A total of 195 references containing LD_{50} data retrieved through these searches were reviewed and evaluated. Information regarding the materials, animals, and methods used to derive the 491 LD_{50} values reported by these references were compiled and are provided in **Appendix H1**. **Appendix H2** provides a narrative characterization and evaluation of the LD_{50} values.

4.1.2 Criteria Used to Select Candidate Rodent Acute Oral Data for Determination of LD_{50} Reference Values

This effort was to designed to derive a set of high quality reference oral LD_{50} values from data that were collected using standardized protocols, accompanied by documentation showing that established testing procedures were followed in compliance with national and international GLP guidelines (OECD 1998; FDA 2003; EPA 2003a,b). After a review of the collected data, the SMT determined that a requirement for GLP compliance would eliminate 99% (452 of the 459 values remaining after exclusion of 30 duplicate values and two erroneous values) of the oral LD_{50} values.

The SMT then considered limiting the selection of LD_{50} values to those from studies that used the specifications for animals recommended by the current acute oral toxicity test guidelines. The current guidelines recommend using young adult rats, 8 to 12 weeks of age, of a common laboratory strain (e.g., Sprague-Dawley) and the most sensitive sex (OECD 2001a, c, d; EPA 2002a). Female animals are recommended if there is no information from which to determine the most sensitive sex. A limited number of LD_{50} values were available from animals that fit this description; only 3% (14/459) of the oral LD_{50} values were determined using 8 to 12 week old female laboratory rats. An additional 15 LD_{50} values were obtained from female rats in an appropriate weight range (age not provided in the reference) for that age range (~ 176-250 g according to Charles River [http://www.criver.com], Harlan [http://www.harlan.com/us/index.htm], and Taconic Farms [http://www.taconic.com/anmodels/spragued.htm] websites). Thus, only 6% (29/459) of the acute oral LD_{50} values in the database, covering 21 of the 72 reference substances (29%), were from studies that used the strain, sex, and age of rats recommended by current test guidelines (OECD 2001a; EPA 2002a).

4.1.2.1 *Final Exclusion Criteria*

Because so few studies met the initial criteria (i.e., GLP compliance and use of animals recommended by current acute oral toxicity test guidelines), the database was reviewed and evaluated to derive alternative criteria for the development of reference LD_{50} values. For this evaluation, the SMT looked for commonalities among the data records that, when selected, provided a comparable data set for each chemical. Review of the available data indicated that

the majority of acute oral toxicity tests were conducted by gavage to unanesthetized, young adult laboratory rats of both genders. Thus, the selection process was revised to exclude studies that reflected the following, less typical, materials, animals, and methods in order to compile a homogenous set of reference LD_{50} values for each chemical. The studies excluded were those with:

- Feral rats
- Rats <4 weeks of age
- Anesthetized rats
- Test chemical administered in food or capsule (or non-oral routes, e.g., intraperitoneal [i.p.])
- LD_{50} reported as a range or inequality

Data from feral rats were excluded because the health status and age of these animals was uncertain. All laboratory rat strains/stocks were deemed acceptable on the assumption that they were healthy and provided with adequate care and housing during testing. Data from neonates and weanlings were excluded because their sensitivity to chemical toxicity may differ from that of adults. Four weeks was considered the minimum acceptable age, because rats are typically weaned at approximately three weeks of age (Barrow 2000). Data from feeding experiments or experiments that involved administration of the chemical in capsules were also excluded because gavage is the most common mode of administration for acute oral studies and the rate of gastrointestinal absorption for these other methods is likely to be different (Nebendahl 2000). Because LD_{50} point estimates are required for the prediction model, LD_{50} values reported as ranges or inequalities were unacceptable.

4.1.2.2 *Assumptions Regarding Materials, Animals, and Methods*

The level of detail for describing the materials, animals, and methods for the LD_{50} studies varied greatly. For example, some studies reported only that white rats were used, while others provided complete information on stock/strain, gender, and age of animals. Details on other protocol components such as the number of animals tested per dose group, method of administration, doses administered, clinical signs, and times of death varied as well. In order to use as much of the available data as possible, the following assumptions were made if a study report did not state otherwise:

140 • Rats were young adults of a common laboratory strain
141 • Rats were not anesthetized
142 • Oral route of administration was by gavage

143 4.1.2.3 *Calculation of Reference LD_{50} Values*

144 If a substance had multiple LD_{50} values after the application of the exclusion criteria, the
145 outliers at the 99% level (Dixon and Massey 1981) were excluded. A geometric mean and
146 95% confidence limits were calculated from the remaining values, and used as the reference
147 LD_{50}. A geometric mean was used because it is the antilog of the mean of the logarithm of
148 the values and is less affected than the arithmetic mean by extreme values. The use of a
149 geometric mean also corresponds with the approach used for the RC millimole regression to
150 derive a single IC_{50} value from multiple IC_{50} values (Halle 1998, 2003), and with the
151 approach used to derive the IC_{50} value for each chemical for the *in vitro* - *in vivo* regressions
152 evaluated in the NICEATM/ECVAM validation study (see **Section 6**).
153
154 In addition to the statistical evaluation of outliers, an extreme value, which was not a
155 statistical outlier but was based on biological plausibility, was identified for trichloroacetic
156 acid. This chemical had five reported LD_{50} values ranging from 400-8900 mg/kg after
157 applying the exclusionary criteria. The lowest value (400 mg/kg) was rejected as biologically
158 implausible because up to 1000 mg/kg/day had been used in an oral chronic rodent
159 carcinogenicity study with no, or only minimal, toxicity (EPA 1996).

160 4.1.2.4 *Use of Rat and Mouse Data*

161 If no rat oral LD_{50} values could be found for a reference substance, mouse acute oral LD_{50}
162 values were evaluated using the same approach as was used for rat values. Because an IC_{50}-
163 LD_{50} regression model using only rat data was preferable, the three reference substances (i.e.,
164 epinephrine bitartrate, colchicine, and propylparaben) for which mouse values only were
165 available were not used for the evaluations of accuracy (**Section 6**) or animal reduction
166 (**Section 10**).
167
168

4.2 Final Rodent Acute Oral LD_{50} Reference Values

After the application of the exclusionary criteria, there were 385 acceptable rodent acute oral LD_{50} values from which to calculate reference LD_{50} values. **Table 4-2** shows the reference LD_{50} value for each substance in descending order of toxicity, presented both as mg/kg and as mmol/kg. Data are presented as mmol/kg in order to be consistent with the RC approach. The RC regression used units of mmol/kg for the LD_{50} and mM for the IC_{50} (see **Section 1.1.3**). Also shown for each substance are the 95% confidence limits around the geometric mean, the ratio of the maximum to the minimum acceptable value, the number of LD_{50} values used to calculate the reference value, the number of LD_{50} values available (not including duplicate values or erroneous values), and the LD_{50} value initially used for hazard classification of the reference substance (see **Table 3-2**).

Table 4-2 lists the reference substances grouped by GHS acute oral toxicity category (UN 2005) using the reference LD_{50} values that were derived as described above. The initial categorization for this study, which used the LD_{50} values in the far right column of **Table 4-2** (i.e., values reported in **Table 3-2**, which come from the RC unless otherwise specified), placed 12 substances in each toxicity category. **Table 4-3** compares the number of substances in each GHS toxicity category based on their reference LD_{50} values with the number in each category based on the initial LD_{50} values. The initial and reference LD_{50} values placed 53 (74%) of the substances in the same GHS category. Nineteen substances (26%) were reclassified based on the reference LD_{50} values (this value is the sum of the numbers in the discordant cells in **Table 4-3)**. Compared with the initial LD_{50} value, the reference LD_{50} value was higher for 18 (25%) and lower for only one (1%) of the substances.

Table 4-2 Rodent Acute Oral Reference LD_{50} Values Listed by GHS Category[1]

GHS Category[1]/ Reference Substance	Reference Acute Oral LD_{50}[2,3] (mg/kg)	95% Confidence Interval[4] (mg/kg)	Reference Acute Oral LD_{50} Range[5] (mg/kg)	Reference Acute Oral LD_{50}[2] (mmol/kg)	95% Confidence Interval[4] (mmol/kg)	Maximum: Minimum Value[6]	N	Initial Rodent Acute Oral LD_{50}[3,7] (mg/kg)
LD_{50} ≤5 mg/kg (N =7)								
Cycloheximide	2	NC	1-2.5	0.00711	NC	2.5	3	2
Phenylthiourea	3	NC	3	0.0197	NC	NC	1	3
Sodium selenate	3	NC	1.6-5.98	0.0159	NC	3.7	2	2[8]
Epinephrine bitartrate	4 (mouse)	NC	4	0.0196	NC	NC	1	4 (mouse)
Triethylenemelamine	4	1-25	1-13	0.0120	0.0037-0.12	13.0	4	1
Physostigmine	5	NC	5	0.0182	NC	NC	1	5[8]
Disulfoton	5	2-10	2.3-12.6	0.0182	0.009-0.036	5.5	6	2
5< LD_{50} ≤50 mg/kg (N =12)								
Parathion	6	3-12	1.8-30	0.0209	0.010-0.041	16.7	10	2
Strychnine	6	NC	2.35-16.2	0.0188	NC	6.9	3	2[8]
Aminopterin	7	NC	7	0.016	NC	NC	1	3 (mouse)
Potassium cyanide	7	5-10	5-10	0.111	0.077-0.15	2.0	7	10
Busulfan	12	NC	1.9-29	0.049	0.008-0.38	15.3	4	2
Colchicine	15 (mouse)	NC	5.886-29	0.0375	NC	4.9	3	6 (mouse)
Thallium I sulfate	25	NC	25	0.0495	NC	NC	1	29 (mouse)
Arsenic III trioxide	25	10-64	13-81.5	0.127	0.050-0.32	6.3	5	20
Endosulfan	28	NC	18-43	0.068	NC	2.4	2	18[8]
Digoxin	28	NC	28	0.0362	NC	NC	1	18 (mouse)
Mercury II chloride	40	27-60	12-92	0.148	0.010-0.22	7.7	10	1
Sodium arsenite	44	36-53	36-53	0.336	0.28-0.40	1.5	5	41[8]
50< LD_{50} ≤300 mg/kg (N =12)								
Sodium dichromate dihydrate	51	44-58	34.17-64.5	0.193	0.17-0.22	1.9	11	50
Dichlorvos	59	40-88	17-97.5	0.266	0.18-0.40	5.7	9	17[8]
Nicotine	70	68-72	68-71	0.430	0.42-0.44	1.0	4	50
Fenpropathrin	76	57-100	48.5-164	0.217	0.16-0.29	3.4	9	18[8]
Hexachlorophene	82	68-98	56-215	0.202	0.17-0.24	3.8	19	61
Paraquat	93	65-132	57-115	0.498	0.35-0.71	2.0	5	58
Lindane	100	78-129	88-125	0.344	0.27-0.44	1.4	4	76
Verapamil HCl	111	NC	108-114	0.226	NC	1.1	2	108
Sodium I fluoride	127	92-175	64-279	3.020	2.19-4.16	4.4	12	180

Table 4-2 Rodent Acute Oral Reference LD_{50} Values Listed by GHS Category[1]

GHS Category[1]/ Reference Substance	Reference Acute Oral LD_{50}[2,3] (mg/kg)	95% Confidence Interval[4] (mg/kg)	Reference Acute Oral LD_{50} Range[5] (mg/kg)	Reference Acute Oral LD_{50}[2] (mmol/kg)	95% Confidence Interval[4] (mmol/kg)	Maximum: Minimum Value[6]	N	Initial Rodent Acute Oral LD_{50}[3,7] (mg/kg)
Cadmium II chloride	135	88-208	88-211	0.738	0.48-1.14	2.4	5	88
Diquat dibromide	160	NC	121-231	0.466	NC	1.9	3	231
Phenobarbital	224	NC	162-318	0.966	NC	2.0	3	163
300< LD_{50} ≤2000 mg/kg (N =16)								
Caffeine	310	256-374	192-483	1.59	1.32-1.93	2.5	10	192
Triphenyltin hydroxide	329	208-520	46.4-1200	0.896	0.57-1.42	25.9	15	44
Haloperidol	330	NC	128-850	0.877	NC	6.6	2	128[8]
Amitriptyline HCl	348	NC	320-380	1.18	NC	1.2	2	319
Propranolol HCl	466	NC	466	1.575	NC	NC	1	470 (mouse)
Cupric sulfate □ 5 H_2O	474	269-836	236.2-960	1.90	1.08-3.35	4.1	6	300
Phenol	548	434-692	317-1500	5.82	4.82-7.68	4.7	14	414
Lithium carbonate	590	479-728	525-710	7.98	6.5-9.9	1.4	4	1187 (mouse; sulfate salt)
Glutethimide	600	NC	600	2.76	NC	NC	1	600
Sodium oxalate	633	NC	558-707	4.724	NC	1.3	2[11]	155 (mouse)[9]
Chloral hydrate	638	391-1040	479-863	3.86	2.36-6.29	1.8	4	479
Atropine sulfate	819	641-1045	600-1136	1.21	0.95-1.54	1.9	7	623
Valproic acid	995	NC	670-1480	6.91	NC	2.2	2	1695 (mouse)
Meprobamate	1387	1291-1489	1286-1522	6.35	5.92-6.82	1.2	6	794[8]
Acetylsalicylic acid	1506	1224-1854	616-2840	8.36	6.8-10.3	4.6	14[11]	1000
Procainamide HCl	1950	NC	1950	8.286	NC	NC	1	1950[8]
2000< LD_{50} ≤5000 mg/kg (N =11)								
Acetaminophen	2163	NC	1944-2404	14.3	NC	1.2	2	2404
Potassium I chloride	2799	NC	2600-3020	37.6	NC	1.2	2	2602
Carbamazepine	2805	NC	1957-4025	11.9	NC	2.1	2	1957[8]
Boric aid	3426	2617-4486	2660-5140	55.4	42.3-72.6	1.9	6	2660[8]
5-Aminosalicylic acid	3429	NC	2800-4200	22.4	NC	1.5	2	7749 (mouse)
Chloramphenicol	3491	NC	2500-5000	10.8	NC	2.0	3	3393
Acetonitrile	3598	2951-4375	1320-8120	87.6	71.9-107	6.2	26	3798
Lactic acid	3639	NC	3543-3730	40.3	NC	1.1	2	3730
Carbon tetrachloride	3783	3024-4732	2350-10054	24.6	20-31	4.3	15	2799
Sodium chloride	4046	2917-5623	3000-6140	69.3	50-96	2.0	5	2998

Table 4-2 Rodent Acute Oral Reference LD_{50} Values Listed by GHS Category[1]

GHS Category[1]/ Reference Substance	Reference Acute Oral LD_{50}[2,3] (mg/kg)	95% Confidence Interval[4] (mg/kg)	Reference Acute Oral LD_{50} Range[5] (mg/kg)	Reference Acute Oral LD_{50}[2] (mmol/kg)	95% Confidence Interval[4] (mmol/kg)	Maximum: Minimum Value[6]	N	Initial Rodent Acute Oral LD_{50}[3,7] (mg/kg)
Xylene	4667	1294-16827	1537-8620	43.9	12-158	5.6	4	4300
LD_{50} >5000 mg/kg (N =14)								
2-Propanol	5105	4624-5636	4500-5840	84.9	77-94	1.3	6	5843
Trichloroacetic acid	5229	2745-9961	3320-8900	32.0	16.8-61.0	2.7	4	4999
Dimethylformamide	5309	3548-7925	2800-7182	72.6	49-108	2.6	6	2800
Citric Acid	5929	NC	3000-11700	30.9	NC	3.9	2	3000[8]
Gibberellic acid	6040	NC	5780-6300	17.4	NC	1.1	2	6305
Propylparaben	6332 (mouse)	NC	6332	35.1	NC	NC	1	6326 (mouse)
Ethylene glycol	7161	6266-8204	4000-9900	115.4	101-132	2.5	16	8567
Methanol	8710	6223-12218	5628-12880	272	194-381	2.3	6	13012
Dibutyl phthalate	8892	6180-12794	7499-12436	31.9	22-46	1.7	4	11998
Diethyl phthalate	9311	NC	8600-10100	41.9	NC	1.2	2	8602
Sodium hypochlorite	10328	NC	8200-13000	62.8	NC	1.6	2	8910[10]
Ethanol	11324	8610-14894	7060-17775	245.7	187-323	2.5	8	14008
1,1,1-Trichloroethane	12078	10000-14588	9600-16000	90.5	75-109	1.7	6	10298
Glycerol	19770	10495-37154	12600-27650	215	114-403	2.2	4	12691

Abbreviations: LD_{50}=dose lethal to 50% of the animals tested; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); N=Number of acceptable values used for geometric mean; NC=Not calculated.

[1]Categorized using the reference oral LD_{50}.

[2]Based on a geometric mean of acceptable LD_{50} values from adult laboratory rats unless otherwise specified.

[3]Values rounded to the nearest whole number.

[4]For the geometric mean of the acceptable LD_{50} values, NC is used for substances with three acceptable values or less, which was considered too few for calculation of a valid confidence interval.

[5]Range of acceptable oral LD_{50} values.

[6]Ratio of minimum acceptable LD_{50} to maximum acceptable LD_{50}.

[7]Values rounded to the nearest whole number. Values are from the RC unless otherwise specified; rat data unless otherwise specified.

[8]RTECS® (MDL Information Systems 2002).

[9]RC reference for rat oral LD_{50} of 155 mg/kg is Shrivastava et al. (1992), which references Klinger and Kersten (1961). Klinger and Kersten (1961) indicate the value was determined by intraperitoneal administration to mice.

[10]HSDB (NLM 2002).

[11]An erroneous value obtained from the literature was not included.

208 **Table 4-3 GHS Category Matches for the Rodent Acute Oral LD_{50} Initial and Reference Values**
209

Initial LD_{50} (mg/kg)[1]	Reference LD_{50} (mg/kg)						Total	Category Match	Reference LD_{50} Lower	Reference LD_{50} Higher
	$LD_{50} \leq 5$	$5 < LD_{50} \leq 50$	$50 < LD_{50} \leq 300$	$300 < LD_{50} \leq 2000$	$2000 < LD_{50} \leq 5000$	$LD_{50} > 5000$				
$LD_{50} \leq 5$	7	5	0	0	0	0	12	58%	0%	42% (5)
$5 < LD_{50} \leq 50$	0	7	4	1	0	0	12	58%	0%	42% (5)
$50 < LD_{50} \leq 300$	0	0	8	4	0	0	12	67%	0%	33% (4)
$300 < LD_{50} \leq 2000$	0	0	0	11	1	0	12	92%	0%	8% (1)
$2000 < LD_{50} \leq 5000$	0	0	0	0	9	3	12	75%	0%	25% (3)
$LD_{50} > 5000$	0	0	0	0	1	11	12	92%	8%	0% (0)
Total	7	12	12	16	11	14	72	74%	1%	25% (18)

210 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); LD_{50}=dose lethal to 50% of
211 animals tested.
212 Note: Shaded cells show the number of chemicals for which both LD_{50} categories agree.
213 [1]Initial LD_{50} values were used for reference substance selection and were obtained from the RC (Halle 1998, 2003), RTECS® (MDL
214 Information Systems 2002), and HSDB (NLM 2002) (see **Table 3-2**).
215
216
217

218 Nineteen reference substances (i.e., the sum of the numbers in the mismatching cells in
219 **Table 4-3**) were reclassified because of the reference LD_{50} values. Five substances originally
220 assigned to the most toxic, LD_{50} ≤5 mg/kg category, (i.e., aminopterin, mercury chloride,
221 busulfan, parathion, and strychnine) were moved to the next, less toxic, category (5< LD_{50}
222 ≤50 mg/kg) based on their reference LD_{50} values. In the 5< LD_{50} ≤50 mg/kg category, four
223 substances (dichlorvos, fenpropathrin, sodium dichromate dihydrate, and nicotine) moved to
224 the less toxic 50< LD_{50} ≤300 mg/kg category, and one (triphenyltin hydroxide) moved two
225 categories to 300< LD_{50} ≤2000 mg/kg. In the 50< LD_{50} ≤300 category, four substances
226 (haloperidol, caffeine, copper sulfate pentahydrate, and sodium oxalate) moved to a lower
227 toxicity category (300< LD_{50} ≤2000 mg/kg). Only carbamazepine moved from the 300<
228 LD_{50} ≤2000 mg/kg category to the 2000< LD_{50} ≤5000 mg/kg category. In the 2000< LD_{50}
229 ≤5000 mg/kg category, citric acid, trichloroacetic acid and dimethylformamide moved to the
230 next lower toxicity category (LD_{50} >5000 mg/kg). In the LD_{50} >5000 mg/kg category, 5-
231 aminosalicylic acid moved to the higher toxicity, 2000< LD_{50} ≤5000 mg/kg category. This
232 was the only substance that moved to a more toxic category

233 **4.3 Relevant Toxicity Information for Humans**

234 The relevance of rodent acute systemic LD_{50} data to human lethal blood concentrations was
235 assessed by the MEIC program (Ekwall et al. 1998b), which used mouse and rat oral LD_{50}
236 data from RTECS® (Ekwall et al. 1998a). Mean lethal doses in humans were collected
237 primarily from handbooks containing human clinical toxicity information (Ekwall et al.
238 1998a) supplemented, when necessary, by an in-house compendium from the Swedish
239 Poisons Information Centre. Ekwall et al. (1998b) calculated least squares linear regressions
240 for the prediction of the mean human lethal blood concentrations by rat and/or mouse oral
241 LD_{50} data for the 50 MEIC substances using units of log mol/kg. They reported a correlation
242 of R^2 =0.607 for the rat oral LD_{50} prediction of mean human lethal blood concentrations and
243 R^2 =0.653 for the mouse oral LD_{50} prediction of mean human lethal blood concentrations. It
244 is important for comparisons of MEIC data with rodent LD_{50} values to note that the MEIC
245 human values are not lethal doses, and therefore not equivalent to LD_{50} values. Many of the
246 values (if not the majority) are blood concentrations that were associated with morbidity or
247 mortality, and usually do not reflect the actual dose consumed by the patient. These are not
248 necessarily the peak blood concentrations, but only the concentrations at the time of

249 ascertainment, which could have ranged from immediately after onset of medical treatment
250 to post-mortem. The MEIC organizers readily admitted that they could not relate the blood
251 concentrations to the administered dose.
252
253 The relevance of the NRU data collected in the NICEATM/ECVAM validation study to the
254 prediction of human acute toxicity will be addressed elsewhere by ECVAM in a separate
255 evaluation.

256 4.4 Accuracy and Reliability of the Rodent Acute Oral LD_{50} Reference Values

257 Accuracy (concordance) is the closeness of agreement between a test method result and an
258 accepted reference value (in this case to the rodent acute oral LD_{50} measurement) (ICCVAM
259 2003). Because there are insufficient data to permit a comparison between rodent and human
260 lethal doses, the accuracy of rodent acute oral LD_{50} values for predicting the oral LD_{50} in
261 humans cannot be determined. Acute toxicity testing in rodents leads to a relative ranking of
262 the toxicity of chemicals for regulatory purposes, with the default assumption that the rodent
263 values and ranking are predictive of the human values and ranking.
264
265 The among laboratory reproducibility of the reference LD_{50} values determined in this section
266 may be judged by evaluating the range of acceptable LD_{50} values for each reference
267 substance and by comparing the values (and their variability) with the variability of LD_{50}
268 values derived from controlled acute oral toxicity studies.

269 4.4.1 Variability Among the Acceptable LD_{50} Values

270 The variability among the acceptable rodent acute oral LD_{50} values used to calculate the
271 reference LD_{50} value for each reference substance was assessed by calculating the ratio of the
272 maximum to the minimum value (see **Table 4-2**). For the 62 reference substances with more
273 than one acceptable LD_{50} value, the average maximum:minimum ratio ranged from 1.1 to
274 25.9, with a mean of 4.3 and a median of 2.2. The maximum:minimum ratios were greater
275 than 10 for four substances: triethylenemelamine, parathion, busulfan, and triphenyltin
276 hydroxide. The low LD_{50} values for triethylenemelamine, busulfan, and parathion may have
277 contributed to the high maximum:minimum ratios. The four LD_{50} values for

278 triethylenemelamine ranged from 1 to 13 mg/kg, the four values for busulfan ranged from 1.9
279 to 29 mg/kg, and the 10 values for parathion ranged from 1.8 to 30 mg/kg.
280
281 **Table 4-4** shows the maximum:minimum LD_{50} ratios by toxicity category. The more toxic
282 substances (i.e., $LD_{50} \leq 50$ mg/kg) tended to have higher maximum:minimum ratios than
283 substances with lower toxicity (i.e., $LD_{50} > 50$ mg/kg). This is anticipated because small day-
284 to-day, or laboratory-to-laboratory variations in weighing and dosing the lower
285 concentrations would have a higher impact on the chemicals being administered in low doses
286 than those being administered in the high dose range.
287
288 **Table 4-4 Maximum:Minimum LD_{50} Ratios by GHS Toxicity Category**

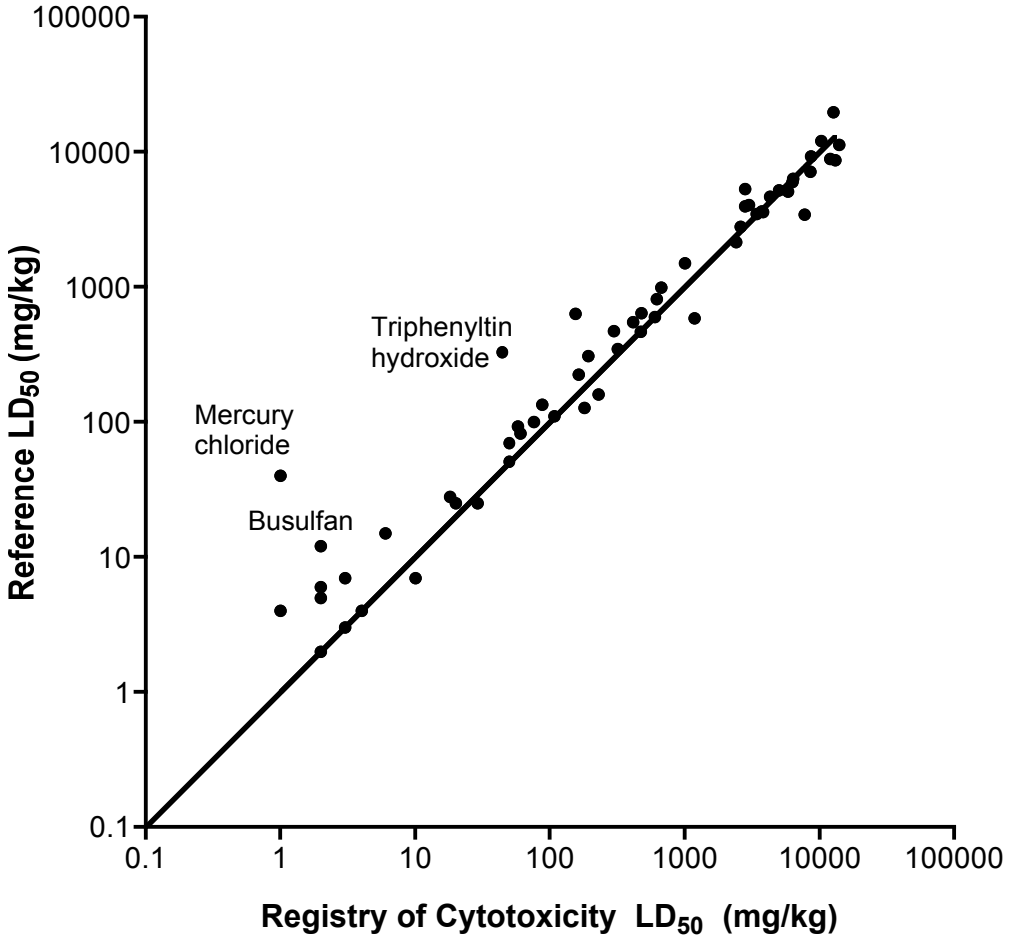
GHS Category (LD_{50} in mg/kg)	Mean Maximum:Minimum LD_{50} Ratio	Median Maximum:Minimum LD_{50} Ratio	Range of Maximum:Minimum LD_{50} Ratio	N
$LD_{50} \leq 5$	6.2	4.6	2.5 – 13.0	4
$5 < LD_{50} \leq 50$	7.1	6.3	2.0 - 16.7	9
$50 < LD_{50} \leq 300$	2.4	1.9	1.1 - 5.7	12
$300 < LD_{50} \leq 2000$	4.6	2.2	1.2 - 25.9	13
$2000 < LD_{50} \leq 5000$	2.6	2.0	1.2- 22.3	11
$LD_{50} > 5000$	2.3	2.3	1.1 - 3.9	13

289 Abbreviations: LD_{50}=dose lethal to 50% of animals tested; GHS-Globally Harmonized System of Classification and
290 Labelling of Chemicals (UN 2005); N=number of chemicals with more than one acceptable LD_{50} value after application of
291 the exclusion criteria described in **Section 4.1.2**.
292

293 4.4.2 Comparison of Rodent Acute Oral LD_{50} Reference Values with the Corresponding
294 RC LD_{50} Values

295 The correspondence of the rodent acute oral LD_{50} reference values with the RC LD_{50} values
296 for the 58 reference substances in common with the RC are shown on a log scale in **Figure 4-**
297 **1**. Not surprisingly, a Spearman correlation analysis for the two sets of log transformed
298 values yielded a significant correlation ($p < 0.0001$) with a correlation coefficient, r_s, of 0.97.
299 **Figure 4-1** shows that the LD_{50} reference values tended to be higher than the RC LD_{50}
300 values. One factor in this difference is that the majority of LD_{50} values used in the RC were
301 from the 1983/84 RTECS®, which contains the lowest LD_{50} value found for a particular
302 chemical without regard to the available methodological information, without consideration
303 of whether it is an outlier with respect to the other available values, and without scientific

304 review before publication. Thus, because the reference LD_{50} values are based on the
305 geometric mean from multiple studies, it is not surprising that these values tended to be
306 higher than the single values in the RC database.
307
308 **Figure 4-1 Correlation of LD_{50} Values for the 58 RC Chemicals With the Reference**
309 **LD_{50} Values**



310
311 Abbreviations: LD_{50}=dose lethal to 50% of animals tested; RC=Registry of Cytotoxicity.
312 The diagonal line shows the 1:1 relationship.
313
314
315 When comparing the reference LD_{50} values to the RC values, the substances with the largest
316 differences were busulfan, triphenyltin hydroxide, and mercury chloride **(**see **Figure 4-1)**.
317 • The LD_{50} reference value for busulfan was six times that of the RC value (12
318 mg/kg vs. 1.9 mg/kg). The RC value (from 1983/84 RTECS®) was from a
319 paper by Schmahl and Osswald (1970) in which they cited a rat oral LD_{50} of

1.86 mg/kg. The literature also contained rat oral LD_{50} values of 28 and 29 mg/kg for male and female Sprague-Dawley rats, respectively (Matsuno et al. 1971).

- The LD_{50} reference value for triphenyltin hydroxide was 7.5 times the RC LD_{50} (329 mg/kg vs. 44 mg/kg). The 15 LD_{50} values used to determine the reference value included the RC value, and had a wide range, 44-1200 mg/kg. Because of the large variation in the data, which was evenly distributed throughout the range neither the highest nor the lowest values were outliers.
- The LD_{50} reference value for mercury chloride was 40 mg/kg, while the RC value was 1 mg/kg. The RC value was from a summary document that reported the rat oral LD_{50} as a range of 1-5 mg/kg (Worthing and Walker 1991). Because it was reported as a range, it was excluded from the calculation of the reference value (see **Section 4.1.2.1**). The remaining 11 values ranged from 12 to 160 mg/kg. The highest value (160 mg/kg) was considered an outlier when compared to the other 10 values and therefore excluded from the reference value calculation.

4.4.3 Comparison of the Variability Among Acceptable LD_{50} Values to Those Obtained in Other Studies

The variation seen here for 62 reference substances is not atypical, considering the results of other studies that examined the variation among rodent acute oral LD_{50} values derived for the same substance. For example, Weil and Wright (1967) showed that LD_{50} values varied by as much as five-fold for the 10 substances tested in eight laboratories using exactly the same protocol. Another international study involving 65 participating laboratories in eight countries that did not control the LD_{50} protocols among laboratories, reported maximum:minimum ratios from 3.6 to 11.3 (with LD_{50} values ranging from 44 to 5420 mg/kg) for five substances (Hunter et al. 1979). The chemicals tested, and the LD_{50} ranges were:

- PCP[1] 44-523 mg/kg
- Sodium salicylate 800-4150 mg/kg

[1] Compound undefined in the publication.

- Aniline 350-1280 mg/kg
- Acetanilide 805-5420 mg/kg
- Cadmium chloride 70-513 mg/kg

The results of a follow-on study in which the same substances were tested by 100 laboratories in 13 countries showed that adherence to a specific protocol reduced the range of maximum:minimum LD_{50} ratios from 3.6 to 11.3 to 2.4 to 8.4 (Zbinden and Flury-Roversi 1981).

Although the LD_{50} data collected from the literature for the NICEATM/ECVAM validation study used various rat strains, sexes, observation durations, and calculation methods for estimating the LD_{50}, the variation in LD_{50} values for individual substances was similar to the data of the earlier cited studies. The current study found four of the 62 substances with multiple LD_{50} values had maximum:minimum LD_{50} values higher than that reported by Hunter et al. (1979) (i.e., >11.3), and three of those were in the highest toxicity category. Hunter et al. (1979) also observed that the largest variation was associated with the more highly toxic substances.

4.5 Summary

To enable the comparison of *in vitro* NRU data with rodent acute oral toxicity data, LD_{50} reference values for the 72 reference substances were calculated using data obtained from the literature, database searches, and secondary references. Rat acute oral LD_{50} values were preferred, but mouse acute oral LD_{50} values were collected for three substances with no available or acceptable rat data. The 491 LD_{50} values that were retrieved comprised 485 rat LD_{50} values and six mouse values. It was not possible to identify a high quality data set produced under GLP guidelines because only 3% of the data records were in GLP compliance. Instead, as described in **Section 4.1.2.1**, a homogenous set of LD_{50} values for each substance was identified by applying specific exclusion criteria related to the materials, animals, and methods used for each study.

After analysis of the acceptable values for outliers, the remaining 385 values were used to derive rodent acute oral LD_{50} reference values by calculation of a geometric mean of the

values for each substance. As a result of this procedure, the LD_{50} reference values for 19 of the 72 reference substances were sufficiently different from the values that were used in the RC and other summary sources, so that they were reclassified into different GHS oral toxicity categories.

Because there is no reference standard against which to evaluate the accuracy of the rodent acute oral toxicity test, the reliability of the LD_{50} reference values was assessed by comparison to other evaluations of the performance of this test method. The maximum:minimum ratio of the acceptable values for the 62 reference substances that had more than one LD_{50} value ranged from 1.1 to 25.9, and the ratios for four of the substances were greater than one order of magnitude.

5.0 3T3 AND NHK NRU TEST METHOD DATA AND RESULTS 5-3

5.1 Study Timeline and Participating Laboratories.....5-3
5.1.1 Statements of Work (SOW) and Protocols.....5-3
5.1.2 Study Timeline.....5-4
5.1.3 Participating Laboratories 5-4

5.2 Coded Reference Substances and GLP Guidelines 5-6
5.2.1 Coded Reference Substances.....5-6
5.2.2 Lot-to-Lot Consistency of Reference Substances.....5-6
5.2.3 Adherence to GLP Guidelines.....5-6

5.3 3T3 and NHK NRU Test Method Protocols 5-7
5.3.1 Phase Ia: Laboratory Evaluation Phase.....5-7
5.3.2 Phase Ib: Laboratory Evaluation Phase.....5-8
5.3.3 Phase II: Laboratory Qualification Phase.....5-9
5.3.4 Phase III: Main Validation Phase 5-11

5.4 Data Used to Evaluate Test Method Accuracy and Reliability.....5-14
5.4.1 PC Data.....5-14
5.4.2 Reference Substance Data 5-17

5.5 Statistical Approaches to the Evaluation of 3T3 and NHK Data.....5-31
5.5.1 Statistical Analyses for Phase Ia.....5-31
5.5.2 Statistical Analyses for Phase Ib.....5-33
5.5.3 Statistical Analyses for Phase II 5-34
5.5.4 Statistical Analyses for Phase III 5-35
5.5.5 Summary of the Data Used for Statistical Analyses.....5-39

5.6 Summary of NRU Test Results 5-39

5.7 Availability of Data.....5-46

5.8 Solubility Test Results 5-47
5.8.1 Solubility Data.....5-47
5.8.2 Solubility and Volatility Effects in the Cytotoxicity Tests 5-52

5.9 Summary.....5-53

[This Page Intentionally Left Blank]

5.0 3T3 AND NHK NRU TEST METHOD DATA AND RESULTS

This section summarizes the IC_{50} results generated by testing 72 coded reference substances (see **Section 3**) in the 3T3 and NHK NRU test method protocols. These IC_{50} values were used to evaluate the accuracy (also known as concordance - see **Section 6**) of the two *in vitro* cytotoxicity test methods for predicting *in vivo* GHS acute oral toxicity categories and their reliability (intra- and inter-laboratory reproducibility - see **Section 7**). The individual test data for the passing and failing tests are provided in **Appendix I** for the reference substances and the PC. The raw data for each test (in EXCEL® and PRISM® files) are available upon request from NICEATM on compact disk(s), as are the laboratory reports. Requests can be made 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.

Section 5.1 discusses the timeline for the validation study, the study participants, and their roles in the study. **Section 5.2** documents the use of coded reference substances and the GLP compliance by the participating laboratories. **Section 5.3** discusses the protocol revisions that were made during the study and the effect the revisions had on the results. **Section 5.4** presents the IC_{50} data collected during each phase to assess the reliability and accuracy (relevance) of the NRU methods. **Section 5.5** presents the statistical analyses performed. **Section 5.6** summarizes the results of IC_{50} comparisons of the 3T3 and NHK methods. **Section 5.7** offers information about the availability of all the data (e.g., raw OD data from all tests, laboratory reports), and **Section 5.8** presents the solubility test results for the reference substances from all laboratories.

5.1 Study Timeline and Participating Laboratories

5.1.1 Statements of Work (SOW) and Protocols

The SMT provided the laboratories with SOWs for each test method prior to initiation of testing (see **Appendix G**), and proposed dates for completion of the various aspects of the study (e.g., transfer of data, provision of reports). The SOWs defined the following:

- Project objectives
- Management and key personnel

- Required facilities, equipment, and supplies
- Quality assurance requirements
- Test phases and schedules
- Products (e.g., reports) required
- Report preparation

The SOW for BioReliance contained all of the above requirements, and also included requirements for:

- Reference substance acquisition, coding, preparation, and distribution
- Solubility testing

The SMT, in consultation with the laboratories, prepared Test Method Protocols for each phase of the study. Cytotoxicity testing in each phase of the validation study was initiated in each laboratory when the SMT received a signed protocol specific for that phase from the Study Director. Solubility testing for the Phases I and II substances was performed prior to cytotoxicity testing for those substances; most of the solubility testing for the Phase III substances was performed toward the end of Phase II and during the early part of Phase III.

5.1.2 Study Timeline

The actual timeline of the study is shown in **Table 5-1**. The SMT modified the original timeline presented in the SOWs because of a number of factors, such as, protocol revisions, side studies, difficulties with acquisition of medium, etc.

5.1.3 Participating Laboratories

- BioReliance Corporation
  14920 Broschart Road
  Rockville, Maryland 20850-3349
  Study Director: Dr. Martin Wenk
- U.S. Army Edgewood Chemical Biological Center (ECBC)
  Molecular Engineering Team
  Aberdeen Proving Ground, MD 21010
  Study Director: Dr. Cheng Cao

116 • Institute for *In Vitro* Sciences (IIVS)
117 21 Firstfield Road Suite 220
118 Gaithersburg, MD 20878
119 Study Director: Mr. Hans Raabe
120 • Fund for the Replacement of Animals in Medical Experiments Alternatives
121 Laboratory (FAL)
122 Queens Medical Centre, University of Nottingham
123 Nottingham NG7 2UH
124 United Kingdom
125 Study Director: Dr. Richard Clothier
126

127 **Table 5-1 Validation Study Timetable**
128

Event	BioReliance	ECBC	FAL	IIVS
Receipt of SOW from SMT	Jun 2002	Jun 2002	Jun 2002	Jun 2002
Procurement of Test Substances	Jul 2002 - Jan 2003	NA	NA	NA
Solubility Testing Completed	Jul 2002 - Jan 2003	Dec 2003	Dec 2003	Jan 2004
Distribution of Reference Substances Phase Ia Phase Ib Phase II Phase III	 Jul 2002 Sep 2002 Nov 2002 Feb - Mar 2003	NA	NA	NA
Initiation of Phase Ia	NA	Aug 2002	Aug 2002	Aug 2002
Completion of Phase Ia	NA	Nov 2002	Nov 2002	Oct 2002
Initiation of Phase Ib	NA	Dec 2002	Dec 2002	Dec 2002
Completion of Phase Ib	NA	May 2003	May 2003	May 2003
Initiation of Phase II	NA	Jun 2003	Jun 2003	Jun 2003
Completion of Phase II	NA	Nov 2003	Nov 2003	Nov 2003
Initiation of Phase III	NA	Dec 2003	Dec 2003	Dec 2003
Completion of Phase III	NA	Dec 2004	Dec 2004	Jan 2005

129 Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical
130 Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; SOW=Statement of Work; SMT=Study
131 Management Team; NA=Not applicable.
132 Note: BioReliance distributed the reference substances and performed solubility testing. ECBC, FAL, and IIVS tested the
133 reference substances for solubility and *in vitro* cytotoxicity.
134

5.2 Coded Reference Substances and GLP Guidelines

5.2.1 Coded Reference Substances

BioReliance acquired 73 substances (72 reference substances and one PC substance) from reputable commercial sources (see **Appendix F1**). All but eight of the reference substances were >99% pure (see **Section 8.1.2.1**). BioReliance coded each substance with a unique, random identification number when repackaging them into smaller units for distribution to the laboratories. These units were given an additional code unique to the respective cytotoxicity laboratories, so that they could be provided in a blinded fashion (see **Section 3.4** for distribution procedures). The coded substance units were packaged and shipped such that their identities were concealed; however, all laboratories knew the identity of the positive control. The SMT revealed the codes for each phase after all laboratories had submitted their data and reports for that phase. The laboratories periodically required additional aliquots of reference substance, and BioReliance provided these aliquots from the original stock of reference substance in the same manner that the original aliquots were provided.

5.2.2 Lot-to-Lot Consistency of Reference Substances

Each substance was purchased as a single lot, and each laboratory received aliquots from this same lot throughout the validation study. The reference substance suppliers provided certificates of analysis for each lot, along with the MSDS documents containing substance, physical, and safety and handling information.

5.2.3 Adherence to GLP Guidelines

BioReliance, ECBC, and IIVS, followed GLP procedures for all testing, with the exception of tests designed to resolve technical challenges (e.g., formation of NR crystals; use of film plate sealers for volatile substances; slow growth of cells). The laboratories submitted all data to their respective quality assurance units (as per GLP requirements) and copies of the data were submitted to NICEATM. FAL followed most of the GLP guidelines, but did not employ independent quality assurance reviews of laboratory procedures or documentation. The Study Director for FAL performed all data reviews and provided copies to NICEATM. Hard copy printouts and electronic versions of all data are available at NICEATM.

5.3 3T3 and NHK NRU Test Method Protocols

The protocols for the 3T3 and NHK NRU test methods used during Phase III laboratory testing were the result of modifications and revisions to the *Guidance Document* (ICCVAM 2001b) protocols, the optimization of the protocols used in the laboratory evaluation Phases Ia and Ib, and the laboratory qualification phase (Phase II) (see **Section 2.6**). **Figure 1-2** provides an outline of the study phases, and identifies where repeated observations were carried out to permit protocol evaluation and comparison. **Sections 2.2** and **2.3** address the similarities and differences between the 3T3 and NHK protocols. The remaining subsections in **Section 5.3** address the modifications to the protocols used in each phase, and how those modifications affected each data set.

5.3.1 Phase Ia: Laboratory Evaluation Phase

During Phase Ia, each laboratory established an historical database for the PC substance, SLS. No reference substances were tested in this phase. Ten concentration-response tests were performed using SLS and no more than two tests were performed/day. The resulting data were used to calculate the acceptable response limits for the SLS IC_{50} for use during Phase Ib testing.

Section 2.6.1 summarizes issues that occurred during Phase I and addresses protocol changes made after the initiation of Phase Ia. The specific changes to the protocols for both cell systems are summarized below, along with the impact these changes had on the test data. Changes made in the protocols during Phase Ia were incorporated into the Phase Ib protocols.

5.3.1.1 *Protocol Changes and the Effect on the Data*

- *NR Dye Crystals:* Reduced the NR dye concentration for both cell types. No subsequent tests failed because of NR crystal formation. The background OD values decreased and this was not interpreted as a negative effect on the data.
- *3T3 Cell Growth*: Modified cell culture conditions for 3T3 cells to improve cell growth characteristics. No apparent effect on the data was detected.
- *NHK Cell Growth (96-well plates):* Removed the cell culture refeeding step performed prior to reference substance addition. Although the OD values for

192 the vehicle controls became higher, the SLS IC_{50} results were similar whether
193 or not the cells were re-fed.

194 • *NHK Cell Growth (in culture flasks)*: FAL coated their culture flasks with
195 fibronectin-collagen prior to seeding thawed cells. This may have affected the
196 SLS data from FAL because it had the highest SLS IC_{50} values of the three
197 laboratories (7.45 μg/mL vs. 4.03 μg/mL for ECBC and 3.68 μg/mL for
198 IIVS). The fibronectin-collagen coating procedure was eliminated, and
199 subsequent SLS data and IC_{50} results from FAL were comparable to the data
200 from the other two laboratories.

201 • *OD Limits*: Eliminated the VC OD range as a test acceptance criterion. The
202 SMT decided to accept tests that had VC ODs outside the originally preset
203 range if all other test acceptance criteria were met. Test data were not
204 adversely affected by relaxing this criterion.

205 • *Dilution Factor*: The SMT accepted data generated using dilution factors
206 other than the recommended 1.47 for definitive tests if all other test
207 acceptance criteria were met. The use of smaller dilution factors generally
208 increased the number of data points between 10 - 90% viability, and the
209 precision of the IC_{50} calculation was improved.

210 5.3.2 Phase Ib: Laboratory Evaluation Phase

211 Phase Ib was designed to determine whether the protocol revisions following Phase Ia were
212 effective in improving intra- and inter-laboratory reproducibility, and to determine whether
213 the laboratories could obtain reproducible results when testing coded reference substances of
214 various toxicities. Three coded reference substances representing the full range of toxicity
215 were tested: arsenic trioxide (high toxicity: $5 < LD_{50} \leq 50$ mg/kg), propranolol HCl (medium
216 toxicity: $300 < LD_{50} \leq 2000$ mg/kg), and ethylene glycol (low toxicity: $LD_{50} > 5000$ mg/kg)
217 (see **Section 3.3.5** for the selection of substances to be tested in Phases Ib and II). Because
218 Phase Ib was part of the laboratory evaluation phase, the SMT decided that three substances
219 would be sufficient, and that it was not necessary to represent all GHS acute oral toxicity
220 categories. Each substance was tested in all laboratories at least once in a range finding
221 experiment, and then in three, acceptable definitive tests performed on three different days.

222

223 **Section 2.6.2** summarizes the technical challenges that arose during this phase and addresses
224 protocol changes made after initiation of Phase Ib. The specific changes made in the 3T3 and
225 NHK protocols, along with the effect the changes had on the test data, are summarized
226 below.

227 5.3.2.1 *Protocol Changes and the Effect on the Data*

228 • *NR Dye Crystals*: Reduced the concentration of NR in the 3T3 method. The
229 OD values and SLS IC_{50} results were similar in four exploratory experiments
230 regardless of the NR concentration or NRU incubation time. The elimination
231 of NR crystals reduced the background OD values without affecting the
232 sensitivity of the procedure.

233 • *VC OD Range*: Used new VC OD ranges for guidance (e.g., as target values to
234 assess cell growth), rather than as a test acceptance criterion, for the remainder
235 of the study. This increased the number of tests that met the acceptance
236 criteria. Relative toxicities did not change. The test data were not adversely
237 affected by the removal of this criterion.

238 5.3.3 <u>Phase II: Laboratory Qualification Phase</u>

239 The results from Phase II were used to determine whether the protocol revisions from Phase
240 Ib were effective in improving intra- and inter-laboratory reproducibility, and whether the
241 laboratories could obtain reproducible results when testing a larger set of substances covering
242 a wider range of physical/substance characteristics and toxicities. Nine coded reference
243 substances were tested: aminopterin, cadmium chloride, chloramphenicol, colchicine, lithium
244 carbonate, potassium chloride, 2-propanol, sodium fluoride, and sodium selenate. These
245 substances (with the exception of sodium selenate) are included in the RC, and were selected
246 because they fit the RC millimole regression line (i.e., they were within the acceptance
247 intervals established by Halle [1998, 2003]). The RC is a database of acute oral LD_{50} values
248 for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays
249 using multiple cell lines and cytotoxicity endpoints for substances with known molecular
250 weights (Halle 1998, 2003). Sodium selenate was selected because of its high toxicity,
251 despite the fact that it was not in the RC, because there were no other substances in the

252 highest GHS acute oral toxicity category, other than aminopterin, that were within the RC
253 millimole regression acceptance intervals. Each laboratory tested each substance at least once
254 in a range finding experiment, and then in three acceptable definitive tests performed on
255 different days.
256
257 **Section 2.6.2** summarizes the technical issues that arose during this phase and the protocol
258 changes made prior to Phase II. The specific changes made in the 3T3 and NHK NRU
259 protocols, along with the effect the changes had on the test data, are summarized below.

260 5.3.3.1 *Protocol Changes and the Effect on the Data*

261 • *Blank Wells*: Added reference substance to blank wells of the test plate to
262 determine if reference substance affected (i.e., increased OD values) compared
263 to medium-filled blank wells. There was no apparent effect on the test data as
264 there were no noticeable differences in OD values between blanks with culture
265 medium or culture medium and reference substance.
266 • *VC OD Range*: Eliminated the VC OD range as an acceptance criterion. There
267 was no apparent effect on test data from not restricting the OD values to a pre-
268 set range.
269 • *Harmonization of Laboratory Techniques*: Made revisions to the Phase II
270 protocols as a result of the harmonization training by the testing laboratories
271 (see **Section 2.6.2.6**). There was no apparent effect on the test data from IIVS
272 and ECBC, but there was an improvement in the FAL data quality (e.g., fewer
273 lost OD values due to cell seeding errors, more uniform OD values for six
274 replicate wells per reference substance).
275 • *3T3 Cell Seeding Density*: Added a range of cell seeding densities to be used
276 by the laboratories. This optimized the cell confluence at the end of chemical
277 exposure and no apparent effects on the data were detected because of this
278 modification.
279 • *NHK Cell Growth from Cryopreserved Stock Cells*: Eliminated the use of
280 fibronectin-collagen coating of 80-cm^2 flasks for the initial propagation of
281 NHK cells. By doing this, FAL achieved better cell growth, lower IC_{50} values

for the PC, and better agreement of the mean SLS IC_{50} values with those of the other laboratories.

- *Volatile Substances*: Added the use of a CO_2 permeable plate sealer to control volatility (as identified by cross contamination of the control wells). The use of plate sealers for volatile substances was incorporated into the Phase III protocols.
- *R^2 Acceptance Criterion*: Relaxed the R^2 criterion for the fit of the dose-response data to the Hill function. Some tests that did not meet the original criterion were accepted by the SMT after determining that even though the curve fit was not optimum, it adequately conveyed the toxicity of the substance (i.e., an IC_{50} could be calculated with an adequate number of toxicity points between 0 and 100% viability).
- *Unusual Concentration-Response*: Revised the Hill function calculation to address substances that produced a concentration-response in which toxicity plateaued before reaching 0% viability. This modification allowed for a curve fit to the Hill function for such substances, and thus a better estimation of their IC_{50} values.
- *PC IC_{50} Range*: Expanded the SLS IC_{50} acceptable range, which resulted in additional tests in Phase II being acceptable. Expanding the PC range reduced the number of reference substance retests, and thereby qualified additional definitive tests as acceptable because they would not fail simply because the PC was out of the pre-set range.

5.3.4 Phase III: Main Validation Phase

The purpose of Phase III was to generate high quality *in vitro* cytotoxicity data using the 3T3 and NHK NRU test methods with protocols that were optimized based on the experience and results in Phases I and II. Sixty coded reference substances were tested; 46 of these were RC substances that covered a broad range of toxicity. The reference substances in Phase III spanned all five GHS toxicity categories and unclassified substances. Each substance was tested in each laboratory at least once in a range finding experiment, and then in three acceptable definitive tests performed on different days.

Section 2.6.4 addresses protocol changes made before the initiation of Phase III. The specific changes made in the 3T3 and NHK protocols, along with the effect the changes had on the test data, are summarized below.

5.3.4.1 *Protocol Changes and the Effect on the Data*

- *Prequalification of NHK Culture Medium*: Included a protocol for prequalifying NHK culture medium and supplements. This prevented the participating laboratories from using medium and supplements that did not support adequate growth of the cells.
- *Stopping Rule for Testing*: Added this rule for reference substances that were insoluble (i.e., <200 μg/mL) and/or did not produce sufficient cytotoxicity for the calculation of an IC_{50}. This rule allowed testing to end for substances that produced no IC_{50} data after three definitive tests. Substances for which an IC_{50} was not produced by one or more laboratories are presented in **Table 5-2**. Carbon tetrachloride did not produce an IC_{50} in any of the laboratories in either the 3T3 or the NHK NRU test methods, and methanol did not produce an IC_{50} in the 3T3 NRU test method.
- *Acceptable Range for Dose-Response Data Points:* Modified the test acceptance criterion for the number of data points required on the toxicity curve. The criterion was changed from requiring a minimum of two points (at least one >0% and ≤50% viability, and at least one >50% and <100% viability) to one point >0% and <100% viability, if the smallest practical dilution factor (i.e., 1.21) was used, and all other test acceptance criteria were met. This reduced the number of failed experiments for substances with very steep concentration-response curves, without reducing the quality of the IC_{50} data. For the 3T3 NRU test method, diquat dibromide (1/9 definitive tests), epinephrine bitartrate (2/9 definitive tests), and 1,1,1-trichloroethane (2/8 definitive tests) had such steep dose-responses that some acceptable tests met these revised criteria. None of the NHK NRU tests needed the revised criteria.
- *R^2 Acceptance Criterion*: Rescinded the R^2 criterion for the fit of the Hill function. The SMT determined that the R^2 criterion was best used to characterize the shape of the concentration-response curve rather than to

344 establish a criterion for test acceptability. This reduced the number of failed
345 experiments without affecting the calculation of the IC_{50} values as long as an
346 adequate number of toxicity points between 0 and 100% viability were
347 obtained.

348 • *PC Acceptance Criteria*: Modified the PC acceptance criterion for Hill
349 function fit.

350 • *Hill Function Analysis*: Altered the PRISM® template for the Hill function
351 analysis to perform calculations for IC_x values in two ways: (1) constraining
352 Bottom parameter to zero, and (2) fitting the Bottom parameter. As a result of
353 the changes and efforts by the laboratories to use dilution schemes that
354 captured the entire concentration-response range, very few tests in Phase III
355 had R^2 <0.9.

356 • *Biphasic Dose-Response in Range Finder Test*: Provided guidance for
357 proceeding with definitive testing when a biphasic dose-response was
358 obtained in the range-finder test. The definitive test was to focus on the lowest
359 concentrations that produced responses around 50% viability (See **Section**
360 **2.6.3.2**).

361

362 **Table 5-2 Reference Substances Affected by Stopping Rule[1]**

363

Reference Substance	Testing Stopped -- No IC_{50} Data					
	3T3 NRU Test Method			NHK NRU Test Method		
	ECBC	FAL	IIVS	ECBC	FAL	IIVS
Carbon tetrachloride	X	X	X	X	X	X
Disulfoton		X				
Gibberellic acid		X				
Methanol	X	X	X	X		
1,1,1-Trichloroethane	X				X	X
Valproic acid			X			
Xylene	X	X		X	X	

364 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=neutral red uptake;
365 ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments
366 Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
367 [1]Substances that did not provide sufficient cytotoxicity for the calculation of an IC_{50} in one or more laboratories (identified
368 by X).
369

370

5.4 Data Used to Evaluate Test Method Accuracy and Reliability

This section first presents the acceptable PC data and IC_{50} results from each laboratory for each phase of the validation study, and then presents the reference substance IC_{50} results and Hill Slopes from each phase. The individual test data for both passing and failing tests are provided in **Appendix I** for the PC and reference substances. Accuracy (concordance for the prediction of GHS acute oral toxicity category) and reliability assessments are provided in **Sections 6** and **7**, respectively.

5.4.1 PC Data

A summary of the acceptable SLS data IC_{50} results used to calculate quality control acceptance limits for each test method in each laboratory are provided in **Table 5-3**. The SLS IC_{50} results were used to calculate acceptable limits for each laboratory to use in subsequent study phases. One of the test acceptance criteria for each reference substance test was that the associated SLS IC_{50} must be within the acceptance limits. The individual test data for both passing and failing PC tests are provided in **Appendix I3** for the 3T3 and in **Appendix I4** for the NHK methods.

Table 5-3 Positive Control (PC)[1] Data and IC_{50} Results by Study Phase

Study Phase	ECBC				FAL				IIVS			
	Mean IC_{50} (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N	Mean IC_{50} (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N	Mean IC_{50} (µg/mL)	Standard Deviation (µg/mL)	Acceptance Limits	N
3T3 NRU												
Ia[2]	38.3	4.71	28.8 – 47.7	15	42.3	8.56	25.2 – 59.5	25	40.9	3.19	34.5 – 47.3	12
Ib[3]	41.3	5.99	26.4 – 56.3	12	43.2	4.68	31.5 – 54.9	17	42.1	3.40	33.6 – 50.6	13
II[4]	41.2	4.20	30.8 – 51.6	29	45.9	7.50	27.2 – 64.7	36	40.6	3.50	31.8 – 49.3	21
III[5]	41.6	3.41	NA	65	41.1	6.23	NA	26	41.5	3.74	NA	22
NHK NRU												
Ia[2]	4.03	1.32	1.40 – 6.67	15	7.45	3.07	1.34 – 13.6	18	3.68	0.555	2.57 – 4.79	30
Ib[3]	3.65	0.98	1.22 – 6.10	11	5.35	2.32	0[6] – 11.1	15	3.57	0.59	2.10 – 5.04	17
II[4]	3.59	1.41	0.07 – 7.11	22	3.20	1.05	0.57 – 5.82	15	3.78	0.73	1.94 – 5.61	26
III[5]	3.03	0.75	NA	57	3.45	0.90	NA	35	3.12	0.53	NA	20

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; N=number of acceptable tests; NA=Not applicable

[1]PC was sodium lauryl sulfate (SLS).

[2]Values generated from Phase Ia data were used as acceptance criteria for Phase Ib tests; Acceptance limits = Mean ±2 X standard deviation.

[3]Values generated from Phases Ia and Ib data were used as acceptance criteria for Phase II tests; Acceptance limits = Mean ±2.5 X standard deviation.

[4]Values generated from Phases Ia, Ib, and II data were used as acceptance criteria for Phase III tests; Acceptance limits = Mean ±2.5 X standard deviation.

[5]Values generated from Phase III test data.

[6]Calculation of lower limits yielded a negative value, so that lower limit was set at 0 and later revised to 0.1 µg/mL.

398 5.4.1.1 *Phase Ib PC Data Acceptance Limits*

399 The SLS IC_{50} acceptance limits for Phase Ib testing were calculated using the Phase Ia data.
400 The data sets from each laboratory were examined for outliers using the method of Dixon
401 and Massey (1981), but none were identified. The acceptance limits for the SLS IC_{50} values
402 for each laboratory and test method were the mean ±2 SD.

403 5.4.1.2 *Phase II PC Data Acceptance Limits*

404 The IC_{50} values from the Phase Ia and Ib SLS tests were used to calculate laboratory-specific
405 and test method-specific quality control acceptance limits for Phase II. Phase Ib tests that had
406 SLS IC_{50} values outside of the acceptance limits were considered acceptable if they met all
407 other test acceptance criteria. For any day during which there was more than one SLS test
408 (for any one method and laboratory), the IC_{50} values were averaged to better reflect day-to-
409 day variation and avoid overweighting the overall mean with multiple values from a single
410 day. Outliers at the 99% level were removed and the remaining values were used to calculate
411 the mean ±2.5 SD acceptance limits. The acceptance limits were expanded from 2 SD in
412 Phase Ib to 2.5 SD for Phase II to allow for the fact that the SDs decrease as more data are
413 collected.

414 5.4.1.3 *Phase III PC Data Acceptance Limits*

415 The IC_{50} values from the Phase I and II SLS tests were used to calculate laboratory-specific
416 and method-specific quality control acceptance limits for Phase III data. The SLS IC_{50} values
417 outside the acceptance limits were considered acceptable if the tests met all other acceptance
418 criteria. For any day for which there was more than one SLS test (for any one method and
419 laboratory), the IC_{50} values were averaged to better reflect day-to-day variation and avoid
420 overweighting the overall mean with multiples values from a single day. ANOVA was used
421 to compare the Phase Ia, Ib, and II data within each laboratory to determine whether the SLS
422 IC_{50} for each method and laboratory was changing over the course of the study. For PC data
423 that were not significantly different from phase to phase at $p < 0.05$, the IC_{50} values were used
424 to calculate the mean ±2.5 SD as the acceptance limits for Phase III. The only significant
425 differences in SLS values seen between study phases ($p < 0.0002$) were the FAL results for
426 NHK. This difference was attributed to the changes in cell culture practices between Phases

427 Ib and II (see **Section 5.3.3**). Thus, only the Phase II SLS IC_{50} values were used to calculate
428 the acceptance limits for Phase III NHK data at FAL.

429 5.4.2 Reference Substance Data

430 Reference substance data and results from the individual 3T3 and NHK tests (both acceptable
431 and unacceptable) from each laboratory are presented in **Appendices I1** and **I2**. **Tables 5-4**
432 and **5-5** summarize the IC_{50} and Hill Slope data from the acceptable 3T3 and NHK tests,
433 respectively, for each reference substance and laboratory. The Hill Slope data are provided
434 for supplemental information on the concentration-response characteristics for each reference
435 substance, but were not used for reliability or accuracy analyses. These tables are organized
436 alphabetically by substance name and provide substance class (based on the NLM Medical
437 Subject Heading [MeSH index]), arithmetic mean IC_{50} and SD for each laboratory, arithmetic
438 mean Hill Slope and SD for each laboratory, and the number of tests used to produce the
439 mean values. **Figure 5-1** graphically presents the 3T3 IC_{50} data from **Table 5-4**, and **Figure**
440 **5-2** presents the NHK IC_{50} data from **Table 5-5**. The reference substances in **Figures 5-1** and
441 **5-2** are ordered by ascending IC_{50} (lowest value [most toxic] to highest value [least toxic])
442 using the 3T3 IC_{50} values from IIVS (the lead laboratory for the study). This allows a simple
443 comparison of each reference substance value from each laboratory. **Table 5-6** provides the
444 numerical key to the reference substances in **Figures 5-1** and **5-2**.

445

Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]
Acetaminophen	Amide	III	40.8	9.12	3	-1.53	0.354	66.2	23.0	3	-1.23	0.503	43.4	11.4	3	-1.55	0.165
Acetonitrile	Nitrile	III	6433	129	3	-2.29	0.648	9690	5634	3	-1.55	0.196	9330	1217	3	-2.63	0.245
Acetylsalicylic acid	Carboxylic Acid; Phenol	III	646	61.5	3	-1.75	0.473	1234	298	3	-1.99	0.393	401	62.0	3	-1.31	0.167
Aminopterin	Heterocyclic	II	0.005	0.001	3	-2.00	0.395	0.012	0.005	3	-3.36	1.59	0.005	0.001	3	-1.46	0.198
5-Aminosalicylic acid	Carboxylic Acid; Phenol	III	1467	203	3	-1.82	0.267	2070	334	3	-2.33	0.809	1557	179	3	-1.64	0.326
Amitriptyline HCl	Polycyclic	III	6.03	1.38	3	-2.47	0.668	7.86	2.20	3	-2.98	0.446	7.81	1.38	3	-4.48	0.916
Arsenic III Trioxide	Arsenical	Ib	2.41	0.782	4	-1.94	0.204	1.04	0.070	4	-3.02	2.09	4.09	2.23	3	-1.62	0.285
Atropine sulfate	Heterocyclic	III	54.1	29.6	3	-1.32	0.480	133	41.1	3	-2.20	0.695	70.0	5.7	3	-1.27	0.165
Boric acid	Boron compound; Acid	III	1497	484	3	-1.14	0.039	3987	693	3	-1.86	0.654	1202	581	3	-1.71	0.677
Busulfan	Alcohol; Sulfur compound; Acyclic hydrocarbon	III	40.4	19.3	3	-0.515	0.003	321	180	3	-1.14	0.802	43.7	1.77	3	-0.627	0.164
Cadmium II chloride	Cadmium compound; Chlorine compound	II	0.480	0.066	3	-1.85	0.529	0.400	0.129	3	-3.05	0.743	0.817	0.427	3	-2.45	0.449
Caffeine	Heterocyclic	III	133	13.3	3	-1.11	0.097	157	81.7	3	-0.866	0.250	191	14.4	3	-1.27	0.077
Carbamazepine	Heterocyclic	III	83.0	12.0	3	-1.94	0.539	152	56.9	3	-3.50	1.27	91.8	11.0	3	-2.34	0.307
Carbon tetrachloride	Halogenated hydrocarbon	III	NA	NA	-	NA	NA	NA	NA	-	NA	NA	NA	NA	-	NA	NA
Chloral hydrate	Alcohol	III	151	15.6	3	-1.73	0.172	241	25.1	3	-2.16	0.597	170	19.9	3	-1.68	0.084
Chloramphenicol	Alcohol; Nitro compound; Cyclic hydrocarbon	II	55.3	12.4	4	-0.779	0.057	273	82.2	4	-1.16	0.249	156	27.9	3	-0.952	0.036
Citric acid	Carboxylic acid	III	473	138	3	-1.89	0.423	1148	143	4	-3.68	0.407	865	160	3	-2.51	0.530

Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]
Colchicine	Polycyclic	II	0.021	0.002	4	-1.69	0.049	0.093	0.042	3	-1.61	1.80	0.028	0.0003	3	-1.69	0.255
Cupric sulfate pentahydrate	Sulfur compound; Metal	III	82.7	3.18	3	-4.85	0.700	123	54.0	4	-17.7	15.5	5.72	1.75	3	-5.71	1.14
Cycloheximide	Heterocyclic	III	0.125	0.057	3	-1.19	0.167	0.647	0.451	3	-1.53	0.128	0.109	0.025	3	-0.937	0.158
Dibutyl phthalate	Carboxylic acid	III	23.5	3.98	3	-3.37	1.27	191	94.5	4	-0.965	0.140	20.7	1.37	3	-2.62	0.283
Dichlorvos	Organophos-phorous	III	9.83	3.42	3	-1.32	0.297	32.8	2.07	3	-3.42	1.00	18.3	2.09	3	-2.13	0.439
Diethyl phthalate	Carboxylic acid	III	85.5	29.0	3	-1.11	0.340	147	37.8	3	-2.03	0.422	106	25.3	3	-2.35	0.824
Digoxin	Polycyclic; Carbohydrate	III	351	137	3	-2.11	2.05	892	319	3	-3.26	2.21	317	67.9	2	-3.04	1.52
Dimethyl-formamide	Amide; Carboxylic acid	III	5343	515	3	-1.96	0.087	5483	517	3	-1.80	0.143	4900	183	3	-1.87	0.102
Diquat dibromide monohydrate	Heterocyclic	III	3.87	0.887	3	-1.59	0.197	36.1	35.5	3	-11.5	10.1	5.39	1.36	3	-3.00	0.784
Disulfoton	Organophos-phorous; Sulfur compound	III	137	74.9	3	-2.06	1.88	11200	NA	1	-1.22	NA	60.4	52.5	3	-2.23	1.08
Endosulfan	Heterocyclic Sulfur compound	III	5.27	3.01	3	-0.669	0.243	15.2	11.9	4	-0.762	0.221	3.61	1.53	3	-0.871	0.636
Epinephrine bitartrate	Alcohol; Amine	III	51.5	6.16	3	-5.99	3.08	63.4	6.63	3	-45.1	32.0	63.4	1.91	3	-4.74	1.51
Ethanol	Alcohol	III	5360	1754	3	-1.33	0.104	8420	1205	3	-1.88	0.128	6413	345	3	-1.99	0.372
Ethylene glycol	Alcohol	Ib	18325	1658	4	-3.79	4.08	31650	7453	4	-1.70	0.166	25900	3081	3	-1.67	0.079
Fenpropathrin	Nitrile; Ester; Ether	III	22.6	2.41	3	-2.54	0.350	42.4	26.8	4	-1.44	0.645	16.7	2.03	3	-2.53	0.495
Gibberellic acid	Polycyclic	III	8027	908	3	-1.95	0.678	NA	NA	-	NA	NA	7657	745	3	-1.66	0.087
Glutethimide	Heterocyclic	III	167	7.00	3	-1.3	0.045	284	20.7	3	-1.47	0.131	125	9.25	4	-1.20	0.163
Glycerol	Alcohol	III	20000	2987	3	-2.02	0.273	38878	28238	4	-2.27	1.29	27833	10882	3	-1.87	0.306
Haloperidol	Ketone	III	5.32	0.649	3	-2.34	0.445	7.99	0.655	3	-4.99	0.378	5.47	0.654	3	-1.86	0.048
Hexachlorophene	Cyclic hydrocarbon	III	5.02	2.41	3	-1.62	0.189	5.35	1.75	3	-1.17	0.322	3.06	0.289	3	-1.66	0.217

Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]
	Phenol																
Lactic acid	Carboxylic acid	III	2943	315	3	-4.13	1.54	3487	561	3	-6.62	3.23	2790	259	3	-3.64	1.09
Lindane	Halogenated hydrocarbon	III	125	119	3	-0.737	0.231	266	94.8	4	-1.26	1.283	90.4	111	5	-1.46	0.262
Lithium I carbonate	Alkalies; Inorganic carbon; Lithium compound	II	564	67.6	3	-1.59	0.313	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Meprobamate	Carboxylic acid	III	353	49.7	3	-1.16	0.438	877	128	4	-1.32	0.270	386	9.02	3	-1.12	0.133
Mercury II chloride	Mercury compound; Chlorine compound	III	3.45	0.177	3	-4.18	0.988	5.99	1.87	3	-4.34	1.11	3.51	0.120	3	-4.16	1.31
Methanol	Alcohol	III	NA	NA	-	NA	NA	NA	NA	-	NA	NA	NA	NA	-	NA	NA
Nicotine	Heterocyclic	III	272	65.3	3	-1.58	0.357	412	136	3	-12.0	6.99	450	54.7	3	-49.6	70.9
Paraquat	Heterocyclic	III	21.3	7.29	3	-1.32	0.341	24.9	16.5	3-	-4.10	3.13	23.7	15.2	3	-1.92	0.581
Parathion	Organophos-phorous; Sulfur compound	III	22.7	12.1	3	-1.89	1.33	141	98.7	4	-1.62	0.520	22.0	4.94	3	-1.55	0.562
Phenobarbital	Heterocyclic	III	634	134	3	-1.43	0.177	726	255	3	-1.84	0.851	476	111	4	-1.67	0.418
Phenol	Phenol	III	50.2	10.9	3	-1.46	0.318	104	24.8	3	-1.55	0.205	58.1	6.78	3	-1.41	0.259
Phenylthiourea	Sulfur compound; Urea	III	30.1	19.8	3	-0.781	0.218	239	65.8	3	-0.890	0.206	89.0	21.9	3	-1.40	0.127
Physostigmine	Carboxylic acid; Heterocyclic	III	28.2	14.9	3	-1.51	0.595	37.8	1.93	3	-7.22	1.04	20.4	6.71	4	-1.70	0.157
Potassium I chloride	Potassium compound; Chlorine compound	II	3352	468	4	-3.32	1.17	3842	1198	5	-4.31	2.27	3710	417	3	-2.87	0.147
Potassium cyanide	Potassium compound; Nitrogen	III	15.3	3.76	3	-1.48	0.677	159	81.9	3	-1.03	0.152	18.9	0.950	3	-3.43	0.488

Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]
	compound																
Procainamide HCl	Carboxylic acid; Amide	III	400	15.3	3	-12.4	1.91	431	4.73	3	-45.6	18.4	497	39.3	3	-19.9	13.1
2-Propanol	Alcohol	II	2610	240	2	-1.80	0.001	3970	139	3	-1.65	0.241	4110	161	3	-1.93	0.160
Propranolol HCl	Alcohol	Ib	13.6	4.37	4	-2.54	0.627	13.5	6.85	4	-3.31	2.53	17.6	3.78	3	-3.45	1.44
Propylparaben	Carboxylic acid; Phenol	III	20.9	3.33	3	-1.23	0.259	51.8	14.8	3	-1.45	0.442	17.1	2.10	3	-1.24	0.245
Sodium arsenite	Sodium compound; Arsenical	III	0.496	0.028	3	-1.43	0.087	1.44	0.819	3	-3.79	1.22	0.683	0.117	3	-1.90	0.535
Sodium chloride	Sodium compound; Chlorine compound	III	4790	233	3	-1.55	0.182	4625	611	4	-2.67	0.620	4877	457	3	-2.03	0.366
Sodium dichromate dihydrate	Sodium compound; Chromium compound	III	0.603	0.087	3	-1.64	0.136	0.657	0.244	3	-5.01	1.51	0.547	0.092	3	-1.93	0.194
Sodium I fluoride	Sodium compound; Fluorine compound	II	61.3	5.55	3	-5.06	1.50	96.1	17.7	3	-4.40	0.971	82.0	5.81	3	-2.73	0.850
Sodium hypochlorite	Sodium compound Oxygen compound; Chlorine compound	III	823	108	3	-2.57	1.12	805	367	3	-4.13	3.05	2005	872	4	-3.20	0.279
Sodium oxalate	Sodium compound; Carboxylic acid	III	42.0	17.3	3	-1.83	0.380	31.0	8.66	3	-3.11	0.367	49.5	26.3	4	-2.32	0.592
Sodium selenate	Sodium compound; Selenium compound	II	12.7	1.62	3	-1.59	0.217	54.2	10.4	3	-3.76	0.968	36.5	5.23	3	-1.65	0.112
Strychnine	Heterocyclic	III	389	80.9	3	-2.51	0.728	124	20.3	3	-5.85	0.922	83.5	5.35	3	-6.49	2.12
Thallium I sulfate	Sulfur compound;	III	2.81	0.671	3	-1.02	0.201	13.4	10.4	4	-0.714	0.302	6.27	1.75	3	-0.752	0.081

Table 5-4 3T3 NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2] (IC_{50})	N	Hill Slope[3]	SD[4]
	Metal																
Trichloroacetic acid	Carboxylic acid	III	762	99.1	3	-1.66	0.118	1220	72.1	3	-2.22	0.089	801	114	3	-1.77	0.130
1,1,1-Trichloro-ethane	Halogenated hydrocarbon	III	41100	NA	1	-2.38	NA	21250	2357	3	-31.5	32.1	9827	180	3	-21.8	8.47
Triethylene-melamine	Heterocyclic	III	0.086	0.009	3	-0.567	0.018	1.45	0.265	3	-1.88	1.04	0.169	0.049	3	-0.615	0.138
Triphenyltin hydroxide	Organo-metallic compound	III	0.026	0.004	3	-1.66	0.257	0.026	0.021	3	-4.78	3.37	0.015	0.008	3	-1.46	0.149
Valproic acid	Carboxylic acid; Lipids	III	547	67.1	3	-2.24	0.742	1807	175	3	-4.07	0.766	574	NA	1	-1.24	NA
Verapamil HCl	Amine	III	32.2	5.82	3	-4.43	1.362	34.6	1.72	3	-29.1	18.6	38.9	4.20	3	-5.00	0.935
Xylene	Cyclic hydrocarbon	III	NA	NA	-	NA	NA	NA	NA	-	NA	NA	724	87.1	3	-1.91	0.473

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; SD=Standard deviation; N=Number of data points; NA=Not available (i.e., IC_{50} values or Hill Slope values could not be generated [see notes in **Appendix I** for more information])

[1]Arithmetic mean.

[2]Standard deviation of IC_{50}.

[3]Arithmetic Mean of Hill Slope values.

[4]Standard deviation of Hill Slope values.

[5]Chemical class assigned is based on the classification of the National Library of Medicine's Medical Subject Heading (MeSH), http://www.nlm.nih.gov/mesh/meshhome.html.

454

Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]
Acetaminophen	Amide	III	558	80.7	3	-1.09	0.108	447	83.7	3	-1.09	0.646	571	79.0	3	-1.20	0.154
Acetonitrile	Nitrile	III	10868	7824	4	-2.61	0.424	10153	1960	4	-5.95	3.34	9290	413	3	-2.79	0.306
Acetylsalicylic acid	Carboxylic Acid; Phenol	III	631	19.9	3	-1.94	0.367	694	98.3	3	-1.85	0.324	514	79.1	3	-1.97	0.083
Aminopterin	Heterocyclic	II	889	182	3	-2.03	0.375	545	42.2	3	-1.27	0.225	611	70.7	2	-1.72	0.547
5-Aminosalicylic acid	Carboxylic Acid; Phenol	III	29.9	6.52	3	-3.45	0.806	78.2	42.3	3	-7.96	6.90	48.8	7.90	3	-3.66	0.629
Amitriptyline HCl	Polycyclic	III	10.8	3.34	3	-1.79	0.236	7.57	5.43	3	-1.43	0.479	10.9	1.04	3	-2.27	0.278
Arsenic III Trioxide	Arsenical	Ib	7.77	2.54	4	-2.67	0.470	2.55	1.92	6	-1.78	1.14	20.9	6.4	3	-2.02	0.338
Atropine sulfate	Heterocyclic	III	85.4	10.5	3	-1.26	0.307	104	88.2	3	-2.90	3.48	83.2	21.0	3	-1.21	0.101
Boric acid	Boron compound; Acid	III	440	138	3	-1.19	0.233	517	378	3	-0.752	0.117	464	11	3	-1.33	0.194
Busulfan	Alcohol; Sulfur compound; Acyclic hydrocarbon	III	253	68.2	3	-0.783	0.323	268	193	3	-1.50	0.357	313	37.2	3	-1.66	0.459
Cadmium II chloride	Cadmium compound; Chlorine compound	II	2.20	0.823	5	-4.01	1.25	1.88	1.22	3	-3.36	3.14	1.86	0.151	3	-4.65	1.38
Caffeine	Heterocyclic	III	817	256	3	-1.44	0.504	591	186	3	-1.06	0.499	574	7.81	3	-1.28	0.117
Carbamazepine	Heterocyclic	III	66.1	8.4	3	-1.15	0.307	253	325	3	-2.57	2.53	63.9	5.27	3	-1.34	0.444
Carbon tetrachloride	Halogenated hydrocarbon	III	NA	NA	-	NA	NA	NA	NA	-	NA	NA	NA	NA	-	NA	NA
Chloral hydrate	Alcohol	III	140	34.2	3	-1.55	0.378	159	50.1	3	-1.33	0.105	112	1.73	3	-1.42	0.123
Chloramphenicol	Alcohol; Nitro compound; Cyclic hydrocarbon	II	318	142	3	-1.51	0.794	414	182	4	-1.16	0.091	367	79.7	3	-0.917	0.249
Citric acid	Carboxylic acid	III	526	82.4	3	-1.62	0.158	312	51.6	4	-1.25	0.249	433	22.3	3	-1.62	0.080

Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]
Colchicine	Polycyclic	II	0.005	0.002	3	-2.15	1.39	0.008	0.001	3	-3.16	1.96	0.008	0.002	3	-13. 8	11.0
Cupric sulfate pentahydrate	Sulfur compound; Metal	III	190	19.6	3	-6.16	3.16	195	12.5	3	-3.85	0.328	207	7.09	3	-5.69	0.871
Cycloheximide	Heterocyclic	III	0.053	0.012	3	-1.24	0.152	0.120	0.094	3	-0.850	0.388	0.071	0.013	3	-1.54	0.178
Dibutyl phthalate	Carboxylic acid	III	28.3	7.64	3	-1.40	0.295	47.4	34.3	3	-1.02	0.352	22.0	1.32	3	-1.33	0.197
Dichlorvos	Organophos-phorous	III	8.56	2.28	3	-1.17	0.147	12.4	3.74	3	-2.29	2.33	12.2	0.416	3	-1.50	0.214
Diethyl phthalate	Carboxylic acid	III	174	14.4	3	-2.21	0.358	71.5	67.3	3	-1.67	0.637	189	33.1	3	-1.97	0.242
Digoxin	Polycyclic; Carbohydrate	III	0.0054	0.0007	3	-2.00	0.127	0.0001	0.00002	3	-1.38	0.684	0.004	0.0003	3	-4.59	1.73
Dimethyl-formamide	Amide; Carboxylic acid	III	9353	155	3	-3.67	0.273	7817	100	3	-2.85	0.590	6397	202	3	-3.00	0.161
Diquat dibromide monohydrate	Heterocyclic	III	3.59	0.825	3	-1.44	0.051	6.77	3.73	4	-1.38	0.488	3.84	0.313	3	-1.10	0.139
Disulfoton	Organophos-phorous; Sulfur compound	III	140	27.0	3	-1.65	1.15	808	213	3	-0.841	0.452	186	59.2	3	-0.836	0.209
Endosulfan	Heterocyclic Sulfur compound	III	3.44	0.573	3	-1.68	0.438	1.42	0.701	4	-1.19	0.369	2.19	0.437	3	-2.20	0.242
Epinephrine bitartrate	Alcohol; Amine	III	115	10.8	3	-7.37	2.10	81.7	28.4	3	-8.39	5.81	75.0	12.2	3	-4.90	2.81
Ethanol	Alcohol	III	8290	390	3	-2.13	0.035	12013	2286	3	-1.82	0.635	10250	867	3	-2.29	0.185
Ethylene glycol	Alcohol	Ib	38000	4681	3	-3.22	0.650	49800	4371	3	-3.02	0.188	40000	5341	4	-2.56	0.444
Fenpropathrin	Nitrile; Ester; Ether	III	3.73	1.01	3	-1.42	0.486	2.23	0.616	3	-4.37	4.45	1.82	0.310	3	-1.78	0.617
Gibberellic acid	Polycyclic	III	2850	402	3	-2.45	0.372	2940	276	3	-5.90	2.69	2807	121	3	-3.30	1.104
Glutethimide	Heterocyclic	III	187	64.3	3	-1.47	0.616	170	24.1	3	-1.29	0.145	176	27.5	3	-1.54	0.237
Glycerol	Alcohol	III	34267	15399	3	-3.32	1.97	18023	8334	3	-1.62	0.521	29033	4596	3	-2.69	0.511
Haloperidol	Ketone	III	3.69	1.01	3	-0.964	0.206	3.72	1.81	3	-0.732	0.097	3.29	1.15	3	-0.840	0.100

Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]
Hexachlorophene	Cyclic hydrocarbon Phenol	III	0.027	0.004	3	-2.21	0.301	0.046	0.020	3	-2.91	0.662	0.021	0.002	3	-2.36	0.059
Lactic acid	Carboxylic acid	III	1290	52.9	3	-2.36	0.306	1320	60.8	3	-3.25	0.328	1313	138	3	-3.23	0.408
Lindane	Halogenated hydrocarbon	III	19.1	3.14	3	-3.02	0.969	23.2	7.09	3	-2.24	0.315	15.6	2.4	3	-2.61	0.265
Lithium I carbonate	Alkalies; Inorganic carbon; Lithium compound	II	411	119	3	-1.95	0.456	486	95.7	3	-1.78	1.31	535	31.6	3	-2.64	0.164
Meprobamate	Carboxylic acid	III	761	116	3	-1.90	0.695	163	189	3	-0.806	0.206	624	84.2	3	-2.04	0.170
Mercury II chloride	Mercury compound; Chlorine compound	III	6.87	1.04	3	-16.3	4.95	5.4	1.02	3	-17.8	13.1	5.35	0.09	3	-17.8	3.31
Methanol	Alcohol	III	NA	NA	-	NA	NA	1133	213	3	-1.79	0.874	2100	226	3	-1.86	0.297
Nicotine	Heterocyclic	III	94.3	24.7	3	-0.654	0.092	134	78.4	3	-0.668	0.077	112	27.7	3	-0.733	0.047
Paraquat	Heterocyclic	III	48.3	6.03	3	-1.04	0.158	96.6	37.2	3	-1.34	0.326	53.4	5.52	3	-1.47	0.034
Parathion	Organophos-phorous; Sulfur compound	III	34.0	10.0	3	-1.60	0.640	31.2	11.9	3	-1.18	0.200	29.0	8.34	3	-1.85	0.956
Phenobarbital	Heterocyclic	III	693	180	3	-1.10	0.214	360	95.5	3	-0.976	0.229	381	69.9	3	-1.68	0.353
Phenol	Phenol	III	59.1	21.4	3	-0.919	0.084	93.2	5.97	3	-1.15	0.209	80.8	5.12	3	-0.915	0.029
Phenylthiourea	Sulfur compound; Urea	III	363	58	3	-1.55	0.726	401	83.6	3	-3.49	1.91	272	71.7	3	-1.00	0.053
Physostigmine	Carboxylic acid; Heterocyclic	III	164	5.51	3	-3.05	0.552	212	238	3	-3.81	2.44	139	8.74	3	-2.97	0.135
Potassium I chloride	Potassium compound; Chlorine compound	II	2560	432	3	-2.23	0.383	2287	631	3	-1.09	0.163	1990	161	3	-2.05	0.165

Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]
Potassium cyanide	Potassium compound; Nitrogen compound	III	29.3	6.9	3	-1.21	0.241	89.0	100	3	-1.10	0.319	16.9	2.21	3	-1.37	0.154
Procainamide HCl	Carboxylic acid; Amide	III	1480	200	3	-3.56	0.813	1787	221	3	-4.22	1.57	2027	229	3	-4.42	0.459
2-Propanol	Alcohol	II	5263	583	3	-2.01	0.173	4273	1139	3	-2.31	0.211	7087	480	3	-3.01	0.406
Propranolol HCl	Alcohol	Ib	38.3	4.54	3	-3.44	0.559	43.8	2.52	3	-2.72	1.461	28.6	3.28	4	-2.09	0.413
Propylparaben	Carboxylic acid; Phenol	III	18.1	2.42	3	-1.18	0.122	18.6	2.84	3	-1.58	0.399	13.8	1.21	3	-1.20	0.065
Sodium arsenite	Sodium compound; Arsenical	III	0.79	0.248	3	-1.69	0.222	0.336	0.187	3	-1.54	0.317	0.470	0.066	3	-1.96	0.197
Sodium chloride	Sodium compound; Chlorine compound	III	3583	263	3	-2.43	0.153	1118	1388	3	-1.96	0.371	3470	300	3	-2.47	0.208
Sodium dichromate dihydrate	Sodium compound; Chromium compound	III	0.784	0.113	3	-2.35	0.282	0.851	0.302	4	-3.52	1.49	0.576	0.100	3	-2.32	0.199
Sodium I fluoride	Sodium compound; Fluorine compound	II	48.7	6.92	3	-2.50	0.263	39.7	9.61	3	-2.60	1.04	53.7	6.82	4	-2.71	0.150
Sodium hypochlorite	Sodium compound Oxygen compound; Chlorine compound	III	1863	581	3	-5.19	1.14	1243	576	3	-2.78	1.27	1633	180	3	-3.86	0.211
Sodium oxalate	Sodium compound; Carboxylic acid	III	355	54.9	3	-4.00	1.99	350	147	4	-6.10	6.40	360	94.6	3	-3.13	0.555
Sodium selenate	Sodium compound; Selenium compound	II	7.47	0.861	3	-1.78	0.529	16.1	9.55	3	-3.07	0.456	10.0	1.33	3	-1.75	0.226

Table 5-5 NHK NRU Test Method IC_{50} and Hill Slope Data by Laboratory

Substance	Chemical Class[5]	Phase in which Tested	ECBC					FAL					IIVS				
			IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]	IC_{50}[1] µg/mL	SD[2]	N	Hill Slope[3]	SD[4]
Strychnine	Heterocyclic	III	100	76.6	4	-1.30	0.729	52.5	28.0	3	-1.60	0.260	55.1	3.43	3	-1.47	0.466
Thallium I sulfate	Sulfur compound; Metal	III	0.198	0.100	3	-2.08	1.01	0.153	0.031	3	-2.64	0.639	0.127	0.020	3	-2.90	0.338
Trichloroacetic acid	Carboxylic acid	III	348	63.5	3	-1.36	0.241	541	150	3	-1.34	0.411	394	50.8	3	-1.48	0.103
1,1,1-Trichloro-ethane	Halogenated hydrocarbon	III	8137	591	3	-14.0	6.08	NA	NA	-	NA	NA	NA	NA	-	NA	NA
Triethylene-melamine	Heterocyclic	III	1.69	0.950	3	-0.838	0.076	2.03	0.471	3	-1.37	0.471	2.13	0.480	3	-1.95	0.369
Triphenyltin hydroxide	Organo-metallic compound	III	0.021	0.007	3	-2.46	0.698	0.007	0.007	3	-3.55	1.68	0.011	0.003	3	-3.34	0.396
Valproic acid	Carboxylic acid; Lipids	III	468	116	3	-1.31	0.252	702	160	3	-1.83	0.455	430	71.5	3	-1.24	0.115
Verapamil HCl	Amine	III	60.5	13.6	3	-1.72	0.238	79.4	33.9	3	-1.88	0.915	66.2	5.57	3	-2.53	0.221
Xylene	Cyclic hydrocarbon	III	NA	NA	-	NA	NA	NA	NA	-	NA	NA	486	185	3	-2.88	1.99

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; SD=Standard deviation; N=Number of data points; NA=Not available (i.e., IC_{50} values or Hill Slope values could not be generated [see notes in **Appendix I** for more information])

[1]Arithmetic mean.

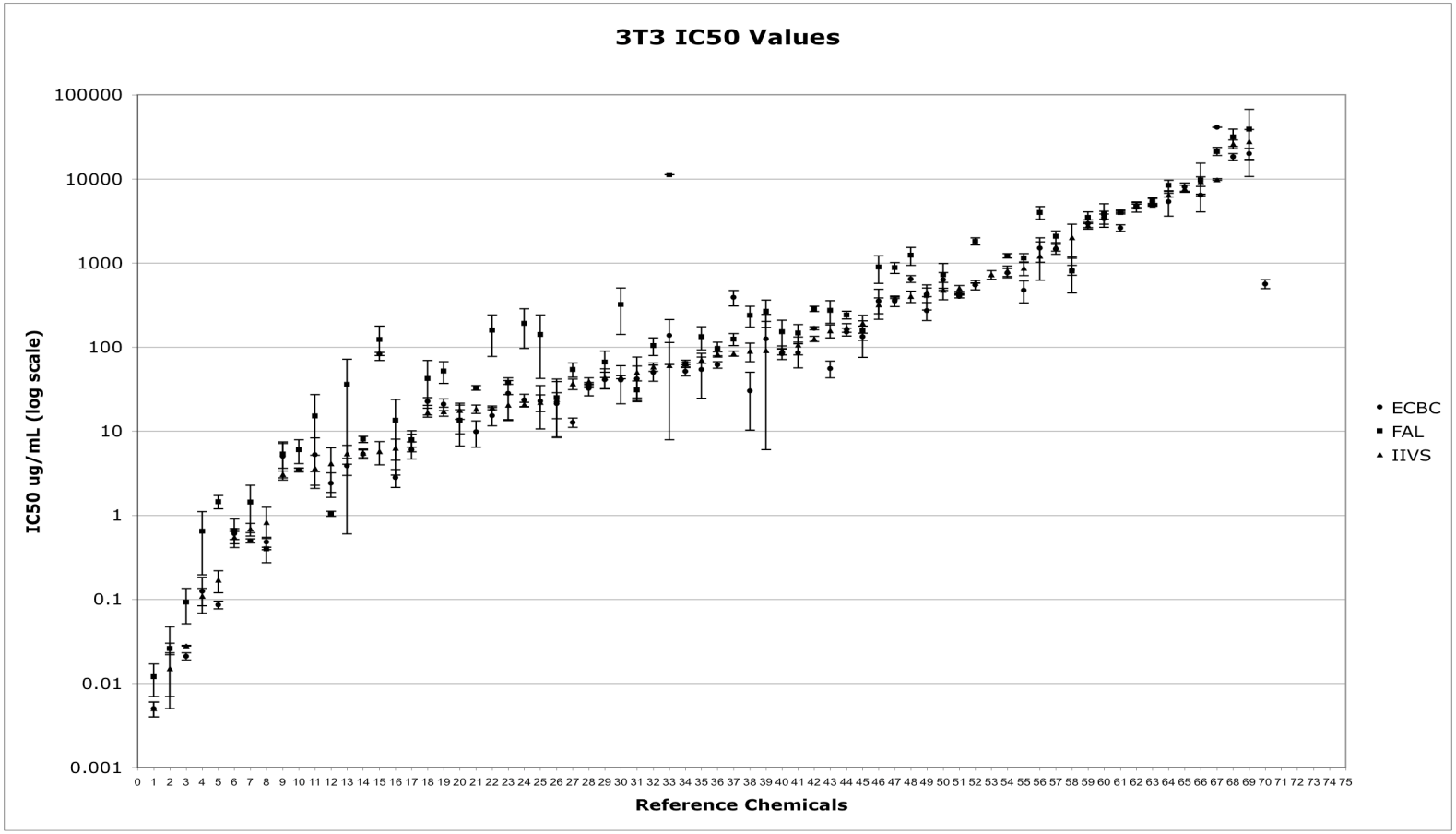
[2]Standard deviation of IC_{50}.

[3]Arithmetic Mean of Hill Slope values.

[4]Standard deviation of Hill Slope values.

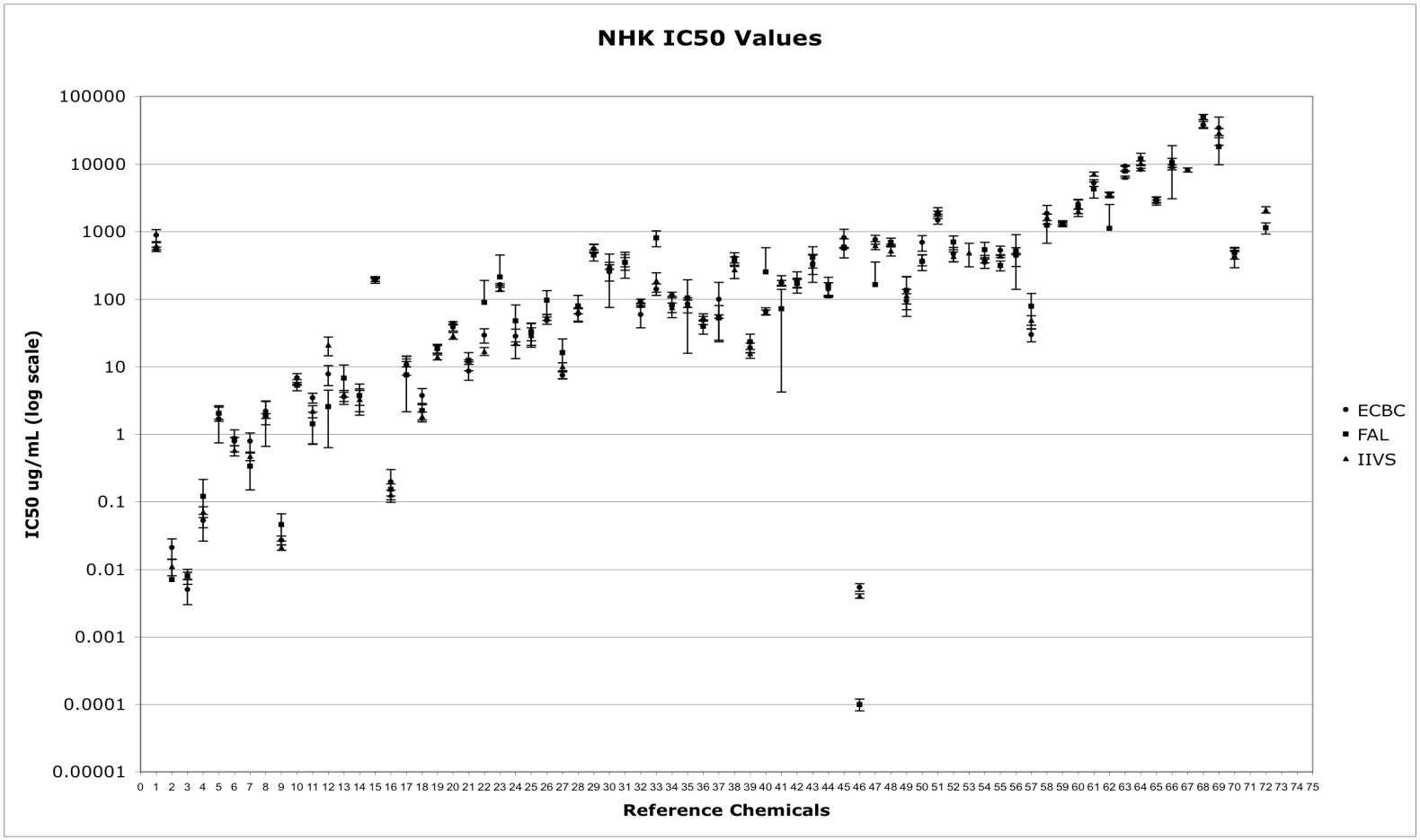
[5]Chemical class assigned is based on the classification of the National Library of Medicine's Medical Subject Heading (MeSH), http://www.nlm.nih.gov/mesh/meshhome.html.

Figure 5-1 Reference Substance IC_{50} Results and Data for the 3T3 NRU Test Method by Laboratory



Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
Points show the mean arithmetic IC_{50} (µg/mL) for each reference substance from each laboratory. Error bars show the standard deviation. Data were sorted in ascending order of 3T3 IC_{50} values from IIVS (lead laboratory in the validation study). **Table 5-6** provides the numerical key for reference substance identification.

Figure 5-2 Reference Substance IC_{50} Results and Data for the NHK NRU Test Method by Laboratory



Abbreviations: ECBC=Edgewood Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.

Points show the mean arithmetic IC_{50} (µg/mL) for each reference substance from each laboratory. Error bars show the standard deviation. Data were sorted in ascending order of 3T3 IC_{50} values from IIVS (lead laboratory in the validation study). **Table 5-6** provides the numerical key for reference substance identification.

551 **Table 5-6 Key to Validation Study Reference Substances[1]**

No	Reference Substance	No	Reference Substance	No	Reference Substance	No	Reference Substance
1	Aminopterin	19	Propylparaben	37	Strychnine	55	Citric acid
2	Triphenyltin hydroxide	20	Propranolol HCl	38	Phenylthiourea	56	Boric acid
3	Colchicine	21	Dichlorvos	39	Lindane	57	5-Aminosalicylic acid
4	Cycloheximide	22	Potassium cyanide	40	Carbamazepine	58	Sodium hypochlorite
5	Triethylenemelamine	23	Physostigmine	41	Diethyl phthalate	59	Lactic acid
6	Sodium dichromate dihydrate	24	Dibutyl phthalate	42	Glutethimide	60	Potassium I chloride
7	Sodium arsenite	25	Parathion	43	Chloramphenicol	61	2-Propanol
8	Cadmium II chloride	26	Paraquat	44	Chloral hydrate	62	Sodium chloride
9	Hexachlorophene	27	Sodium selenate	45	Caffeine	63	Dimethylformamide
10	Mercury II chloride	28	Verapamil HCl	46	Digoxin	64	Ethanol
11	Endosulfan	29	Acetaminophen	47	Meprobamate	65	Gibberellic acid
12	Arsenic III trioxide	30	Busulfan	48	Acetylsalicylic acid	66	Acetonitrile
13	Diquat dibromide monohydrate	31	Sodium oxalate	49	Nicotine	67	1,1,1-Trichloroethane
14	Haloperidol	32	Phenol	50	Phenobarbital	68	Ethylene glycol
15	Cupric sulfate pentahydrate	33	Disulfoton	51	Procainamide HCl	69	Glycerol
16	Thallium I sulfate	34	Epinephrine bitartrate	52	Valproic acid	70	Lithium I carbonate
17	Amitriptyline HCl	35	Atropine sulfate	53	Xylene	71	Carbon tetrachloride
18	Fenpropathrin	36	Sodium I fluoride	54	Trichloroacetic acid	72	Methanol

552 Abbreviations: No=Number.
553 [1]As used in **Figures 5-1** and **5-2**.
554
555

556 Because of their low toxicity and/or low solubility, some substances were not sufficiently
557 toxic for calculation of an IC_{50} value. For the 3T3 NRU test method, no IC_{50} values were
558 obtained for carbon tetrachloride or methanol in any laboratory (see **Table 5-4**). ECBC was
559 the only laboratory that obtained IC_{50} values for lithium carbonate, and IIVS was the only
560 laboratory that obtained IC_{50} values for xylene. Only one acceptable test (and IC_{50} value) was
561 obtained for disulfoton at FAL, for 1,1,1-trichloroethane at ECBC, and for valproic acid at
562 IIVS. FAL did not achieve sufficient toxicity for the calculation of an IC_{50} for gibberellic
563 acid in any 3T3 NRU tests performed. For the NHK NRU test method (see **Table 5-5**), there

564 was insufficient toxicity in all tests in all laboratories for a calculation of an IC_{50} for carbon
565 tetrachloride. Only one laboratory achieved sufficient toxicity for the calculation of an IC_{50}
566 for 1,1,1-trichloroethane (ECBC) and xylene (IIVS). One laboratory, ECBC, failed to
567 achieve sufficient toxicity for the calculation of an IC_{50} for methanol. All of these substances,
568 with the exception of methanol, produced precipitate in the cell culture medium. The solvent
569 used for methanol was DMSO, and because the amount of DMSO that could be used in the
570 cell culture was limited to 0.5%, the amount of DMSO that could be used to dissolve
571 methanol was also limited. The differences among laboratories regarding their ability to
572 attain a high enough concentration to achieve an IC_{50} for some substances may be due to the
573 differing perceptions of the laboratory personnel regarding whether or not the substance was
574 sufficiently dissolved, or differences in the techniques used to dissolve the substances.

575 **5.5 Statistical Approaches to the Evaluation of 3T3 and NHK Data**

576 The statistical approaches used for data evaluation are reviewed in the following sections for
577 each phase of the validation study. **Section 2.2.3** discussed the endpoint measurements for
578 the 3T3 and NHK test methods. The OD values of each of six replicate wells ([minimum of
579 four] in the 96-well plate) per test concentration (eight concentrations/reference substance or
580 PC) were used to determine relative cell viability in relation to the mean VC OD on the same
581 plate. The cell viability values calculated for the replicate wells for each concentration were
582 used to determine the concentration-response curve (percent viability vs. concentration) for
583 each test. The IC_{50} value was determined from fitting the curve to a Hill function.

584 5.5.1 Statistical Analyses for Phase Ia

585 The laboratories reported the IC_{50} results for SLS in μg/mL. The SMT used the results from
586 the acceptable tests to calculate means and SDs for each method at each laboratory.

587 5.5.1.1 *Outlier Determination for Replicate Well Concentration Data*

588 A test for outliers at the 99% level (Dixon and Massey 1981) was used to determine the
589 presence of outlier OD values among the six replicate wells for each reference substance
590 concentration. The SMT applied the outlier test to the Phase Ia data when extreme values
591 were noted. Outliers were excluded from the data set, and the IC_{50} was recalculated. The raw
592 data files include all data provided by the laboratories, including the excluded outlier OD

593 values. Because the protocol required a minimum of four acceptable test wells per reference
594 substance concentration, no more than two wells of the six replicates could be excluded.

595 5.5.1.2 *Curve Fit Criteria*

596 After the completion of Phase Ia testing, a curve fit criterion was implemented for test
597 acceptance following a visual review of the fit of the OD data to the Hill function curve. The
598 SMT considered the fit of the concentration-response curve to the Hill function to be
599 acceptable when $R^2 > 0.9$. A fit of $R^2 < 0.8$ was considered unacceptable and the data from
600 that test were rejected. Curves with a fit of $0.8 < R^2 < 0.9$ were evaluated visually for
601 goodness of fit and accepted if the SMT concluded that there were sufficient data points
602 between 0 and 100% cytotoxicity, and a reasonable shape to the curve, to calculate a
603 reasonably accurate IC_{50} value. Each test with a curve fit in this range was analyzed on a
604 case-by-case basis, and no standard pass/fail criterion was developed. [Note: The use of a
605 curve fit criterion was reevaluated in Phases Ib and II, and was eliminated as a test
606 acceptance criterion for Phase III test results. An R^2 value ≥ 0.85 was maintained as a test
607 acceptance criterion for the PC because its fit to the Hill function was well characterized.]

608 5.5.1.3 *Reproducibility Analyses for PC IC_{50} Values*

609 To evaluate reproducibility of the IC_{50} values for the PC for each test method, within and
610 between the laboratories, the SMT considered the American Society of Testing and Materials
611 (ASTM) Standard E691-99, *Standard Practice for Conducting an Interlaboratory Study to*
612 *Determine the Precision of a Test Method* (ASTM 1999). This method uses two statistics, *h*
613 and *k*, to judge the consistency of means and variances between laboratories. However, a
614 minimum of six laboratories is required for this type of analysis and the SMT decided that it
615 could not be appropriately applied to three laboratories. The variability of the PC IC_{50} results
616 obtained from each test and laboratory was assessed using CV analysis and one-way
617 ANOVA. Dividing the SD by the arithmetic mean IC_{50} value, and multiplying by 100
618 produced the CV. CV values were calculated for the acceptable tests within each laboratory
619 to determine intralaboratory reproducibility. To compare the variation among laboratories,
620 the CV was calculated using the arithmetic mean IC_{50} values from each of the three
621 laboratories. Although no criterion for an acceptable CV was determined for this study,
622 ECVAM recently used CV <30% as an acceptable range for both intra- and inter-laboratory
623 reproducibility (Zuang et al. 2002; Fentem et al. 2001). Although CV <30% was intended to

624 reflect an acceptable maximum for normal biological variability, the range was not supported
625 by data.
626
627 For the ANOVA, IC_{50} values were first converted to mM units and then log-transformed to
628 obtain normal distributions. One-way ANOVA was performed with SAS PROC GLM
629 software (SAS Institute 1999; see **Appendix R1** for example SAS code). A significance level
630 of $p < 0.01$ was used to test results between the laboratories in order to be conservative with
631 respect to identifying laboratory differences.

632 5.5.2 Statistical Analyses of Phase Ib Data

633 5.5.2.1 *Outlier Determination for Replicate Well Concentration Data*
634 For consistency of replicate well concentration data, the SMT applied the same outlier test
635 used for the Phase Ia data (Dixon and Massey 1981) when extreme OD values were noted. If
636 the extreme value was an outlier at the 99% level, it was excluded from the data set, and the
637 IC_{50} was recalculated. All data are available in the data files provided by the laboratories,
638 including the excluded outlier OD values.

639 5.5.2.2 *Reproducibility Analyses of the Reference Substance IC_{50} Values*
640 One-way ANOVA and CV analyses were used to assess method reproducibility within and
641 among laboratories. For the ANOVA, the IC_{50} values were first converted to mM units and
642 then log-transformed to obtain normal distributions. One-way ANOVA was performed with
643 SAS PROC GLM (SAS Institute 1999; see **Appendix R1** for example SAS code). A
644 significance level of $p < 0.01$ was used to test results between the laboratories in order to be
645 conservative with respect to identifying laboratory differences. When the ANOVA detected
646 significant differences among the laboratories, contrast analyses were performed to
647 determine which laboratory was different from the others. These analyses compared the
648 results of each laboratory with those of the other two laboratories. A significant difference in
649 response among the laboratories was indicated by $p < 0.01$.
650
651 CV values were calculated for each reference substance by dividing the SD by the arithmetic
652 mean IC_{50} value and multiplying by 100. CV values were calculated for the acceptable tests
653 in each laboratory to determine intralaboratory reproducibility. To compare the variation

654 among laboratories, the CV was calculated using the arithmetic mean IC_{50} values from each
655 of the three laboratories.

656 5.5.3 Statistical Analyses of Phase II

657 5.5.3.1 *Outlier Determination for Replicate Well Concentration Data*
658 The Dixon and Massey (1981) outlier test was incorporated into the EXCEL® templates to
659 assess the consistency of replicate well data for each reference substance concentration.
660 Outliers at the 99% level were highlighted and the Study Director was offered the option of
661 removing the value from subsequent calculations (e.g., mean OD of the six replicates; %
662 viability; IC_{50}).

663 5.5.3.2 *Reproducibility Analyses of Reference Substance IC_{50} Values*
664 The intra- and inter-laboratory reproducibility of the IC_{50} values were assessed using the
665 acceptable tests to calculate the mean IC_{50}, SD, and CV for each substance, method, and
666 laboratory, as described in **Section 5.5.2.2**. One-way ANOVAs were performed as described
667 in **Section 5.5.2.2**. As an additional approach to the assessment of interlaboratory
668 reproducibility for each test substance, the maximum:minimum IC_{50} ratios (i.e., the
669 maximum arithmetic mean laboratory IC_{50} value compared to the minimum arithmetic mean
670 laboratory IC_{50} value) were calculated . This approach is similar to the calculation of
671 maximum:minimum LD_{50} ratios for examining reproducibility of reference LD_{50} values (see
672 **Section 4.4.1**).

673 5.5.3.3 *Comparison of 3T3 and NHK Test Results with the RC Millimole Regression*
674 To compare the 3T3 and NHK test results for the reference substances to those of the RC
675 millimole regression, each IC_{50} value was transformed to mM units for the calculation of
676 geometric mean IC_{50} values. The use of geometric means corresponded with the approach
677 used to obtain single IC_{50} values from multiple IC_{50} values for the RC millimole regression
678 (Halle 1998, 2003). The log geometric mean IC_{50} values (in mM) of the 11 RC substances
679 tested during Phases Ib and II (see **Table 3-8**) were used with the log RC LD_{50} values, after
680 transformation to log mmol/kg units (see **Appendices J1** and **J2**), to calculate least squares
681 linear regressions for the data from each test method and laboratory. Each of these
682 method/laboratory regressions was compared to the RC millimole regression using an F test
683 with SAS PROC REG (SAS Institute 1999; see **Appendix R2** for example SAS code). An F

684 test with a significance level of $p < 0.01$ was used to determine whether the joint comparison
685 of slope and intercept indicated that the method/laboratory regressions were significantly
686 different from the RC millimole regression.
687
688 As an alternate analysis, a least squares linear regression using IC_{50} and LD_{50} values from the
689 RC was constructed for the 11 RC substances (the "RC-11" regression) tested in Phases Ib
690 and II. Each of these method/laboratory regressions was compared to the RC-11 regression
691 using an F test with SAS PROC GLM (SAS Institute 1999; see **Appendix R2** for example
692 SAS code) at a significance level of $p < 0.01$. This was used to determine whether the
693 comparisons of slope and intercept indicated that the laboratory regressions were
694 significantly different from the RC-11 regressions.

695 5.5.4 Statistical Analyses of Phase III

696 5.5.4.1 *Outlier Determination for Replicate Well Concentration Data*
697 The laboratories used the Dixon and Massey (1981) outlier test at the 99% level that was
698 incorporated into the EXCEL® templates to test for outlier values among replicate well data
699 at the different reference substance concentrations. The Study Director had the option of
700 excluding the outliers from the data set, which were highlighted by the template, and
701 subsequent calculations. All data are available in the data files provided by the laboratories,
702 including the outlier OD values.

703 5.5.4.2 *Reproducibility Analyses of the PC IC_{50} Data*
704 A number of analyses were performed to determine whether the SLS IC_{50} values were
705 reproducible across study phases. The SLS IC_{50} values used to access variability were
706 different from those shown in **Table 5-3**. To get an assessment of the true variation of SLS
707 IC_{50} values, the reproducibility analyses included additional IC_{50} values from SLS tests that
708 did not meet the IC_{50} acceptance limits (see **Table 5-3**) for each laboratory and study phase if
709 they passed all other test acceptance criteria. If more than one SLS test was performed on a
710 single day (for any test method and laboratory), the IC_{50} values were averaged to determine a
711 single IC_{50} for the day. This prevented multiple data values from a single day from overly
712 influencing the mean for each phase. CV analyses were performed as described in **Section**

713 **5.5.1** using the arithmetic mean SLS IC_{50} values for each method, laboratory, and study
714 phase.
715
716 For the remaining analyses of reproducibility, the IC_{50} values were first log-transformed to
717 obtain normal distributions. One-way ANOVAs were performed with SAS PROC GLM
718 (SAS Institute 1999; see **Appendix R1** for example SAS code) for each method using study
719 phase and laboratory as individual variables. A significance level of $p < 0.01$ was used to test
720 for a statistical difference among the laboratory and/or phase results.
721
722 To determine whether there was a linear time trend for the SLS IC_{50} data, linear regression
723 analyses using a least squares method were performed for each laboratory and method using
724 SAS PROC REG (SAS Institute 1999). Time was expressed as an index for each test. The
725 index number of each SLS test reflected its order of testing without respect to the time
726 lapsing between tests. For example, the first SLS test was assigned a time index of 1 and the
727 second SLS test was assigned a time index of 2 whether it occurred the day after the first test
728 or one week after the first test. The slopes of the linear regressions were judged to be
729 statistically significant at $p < 0.05$, which indicated that the IC_{50} had changed significantly
730 over time.

731 5.5.4.3 *Reproducibility Analyses of the Reference Substance IC_{50} Values*

732 CV and one-way ANOVA analyses were performed to assess the intra- and inter-laboratory
733 reproducibility of the Phase III reference substance data, as described in **Section 5.5.2.2**.
734 Interlaboratory reproducibility was also assessed for each test substance by calculating ratios
735 of the maximum mean laboratory IC_{50} to the minimum mean laboratory IC_{50}, as described in
736 **Section 5.5.3.2**. An additional evaluation to determine whether normalizing the reference
737 substance IC_{50} to the SLS IC_{50} would reduce interlaboratory variability was performed using
738 five substances (for each test method) for which the ANOVAs indicated significant
739 interlaboratory differences. The reference substance IC_{50} values were normalized to the SLS
740 IC_{50} by calculating the reference substance IC_{50}:SLS IC_{50} ratio. CVs were calculated for each
741 substance using the mean ratios from each laboratory. To determine whether this
742 normalization reduced variability among the laboratories, the CVs for the substance IC_{50}:SLS
743 IC_{50} ratios were compared to the CVs for the substance IC_{50}.

744

745 In addition, the geometric mean IC_{50} values were used to calculate least squares linear
746 regression models after log transforming the data. Linear regressions were fit for each
747 method and laboratory using the log-transformed reference LD_{50} values from **Table 4-2 (**in
748 mmol/kg), with log IC_{50} in mM. To detect differences among the linear regressions in each
749 laboratory, two models were fit for each method. The first was a full model that included
750 effects for laboratory and interactions, and generated a regression line for each substance in
751 each laboratory, by test method. The second model, which was considered to be a reduced
752 model, assumed that one model fit all the laboratories. A goodness of fit F test was
753 performed to compare the full and reduced models for each method. A significance level of p
754 <0.01 was used to test whether the regressions among laboratories were significantly
755 different from one another. The following criteria were established for selection of data for
756 use in the regression analyses for each test method:

757 - The substance was included in the RC
758 - All three laboratories reported IC_{50} values
759 - There was an associated rat oral reference LD_{50} value (see **Table 4-2**)

760

761 There were 47 reference substances that fit these criteria for the 3T3 and 51 test substances
762 that fit the criteria for the NHK test methods.

763 5.5.4.4 *Comparison of 3T3 and NHK Results with the RC Millimole Regression*

764 To determine whether the IC_{50} values determined in the validation study were significantly
765 different from the RC values, the laboratory-specific regression values for each method were
766 combined using the geometric means of the laboratory-specific geometric mean IC_{50} values
767 in mM and the reference LD_{50} in mmol/kg. Thus, there was one regression analysis with
768 pooled laboratory data for the 3T3 NRU test method and another regression analysis (also
769 with pooled data) for the NHK NRU test method. A third linear regression was calculated
770 using the log transformed IC_{50} and LD_{50} values from the RC. The IC_{50} values and LD_{50}
771 values were log-transformed for the regression calculations. The following criteria were
772 established for the selection of substances to be used for the regression analyses:

773 - The substance was included in the RC

774 • All three laboratories reported IC_{50} values for the 3T3 and NHK NRU test
775 methods
776 • There was an associated rat oral reference LD_{50} value (see **Table 4-2**)
777
778 Forty-seven substances met these criteria. Two models were fit for each test method to detect
779 differences between the NRU regression and the 47 RC substance regression. The first
780 regression model was a full model that included effects for the RC and the NRU regression,
781 and generated one regression line each for the RC and the NRU test method. The second
782 (reduced) model assumed that a single model fit the combined RC and NRU IC_{50} data. The
783 RC regression for the 47 reference substances was compared to the combined laboratory
784 regression for each NRU test method using an F test to simultaneously compare slopes and
785 intercepts. The NRU regressions were statistically different from the RC regressions at
786 $p < 0.01$.
787
788 To assess the accuracy of the NRU methods and the associated IC_{50}-LD_{50} regressions, a
789 predicted LD_{50} was calculated for each reference substance using its laboratory geometric
790 mean IC_{50} in two analyses:
791 • The RC rat-only millimole regression calculated from the 282 RC substances
792 with rat LD_{50} values, using units of mM for the IC_{50} and mmol/kg for the LD_{50}
793 (see **Section 6.2**)
794 • The RC rat-only weight regression calculated from the 282 RC substances
795 with rat LD_{50} values, using units of μg/mL for the IC_{50} and mg/kg for the
796 LD_{50} (see **Section 6.2**)
797
798 The LD_{50} values predicted from the regression analyses were used to predict GHS acute oral
799 toxicity categories (see **Section 6.3**). The accuracy of the predictions was determined by
800 calculating the proportion of substances for which the predicted GHS toxicity category
801 matched the GHS toxicity category. The LD_{50} predictions from these regression models were
802 also used to determine starting doses for acute systemic toxicity test simulations for the
803 purpose calculating animal use and savings that would be achieved using the NRU test

methods. The simulation modeling methods, and results from the UDP and ATC methods, are described in **Section 10**.

5.5.5 Summary of the Data Used for Statistical Analyses

Table 5-7 summarizes the number of substances that were tested and the number of substances used for the various analyses performed to determine the accuracy and reliability of the *in vitro* NRU test methods.

Table 5-7 Datasets Used for Validation Study Analyses[1]

Use	3T3 NRU Test Method[1]	NHK NRU Test Method[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	62	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
Comparison of reproducibility of IC_{50} values with reproducibility of LD_{50} values	53	57	Substances with IC_{50} values from all laboratories and more than one acceptable rat oral LD_{50} value

Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake.
[1]Number of substances.

5.6 Summary of NRU Test Results

Table 5-8 shows the 3T3 and NHK IC_{50} values as geometric means of the geometric mean laboratory values, as a basis to compare the 3T3 and NHK NRU IC_{50} values for each

820 reference substance. The substances in **Table 5-8** are organized by ascending 3T3 NRU IC_{50}
821 values (as was done for **Figures 5-1** and **5-2**). For each method, the table provides the
822 geometric mean IC_{50} (combined across laboratories) in μg/mL, the ratio of the geometric
823 mean IC_{50} to the SLS IC_{50}, and the 3T3 IC_{50}:NHK IC_{50} ratios. Geometric means were used
824 for this comparison because they were used for both the IC_{50} and LD_{50} regression analyses
825 (see **Sections 5.5.3.3**, **5.5.4.3,** and **5.5.4.4**). The 3T3 and NHK NRU IC_{50} values were
826 compared using the ratios of their geometric means. The IC_{50} values for each reference
827 substance were also compared to the IC_{50} for SLS using the ratio of reference substance
828 geometric mean IC_{50} to SLS geometric mean IC_{50}.

829

Table 5-8 Comparison of 3T3 and NHK NRU IC_{50} Geometric Means

Reference Substance	3T3 NRU		NHK NRU		
	Geometric Mean[1] IC_{50} (µg/mL)	Ratio Geometric Mean IC_{50} to SLS IC_{50}	Geometric Mean[1] IC_{50} (µg/mL)	Ratio Geometric Mean IC_{50} to SLS IC_{50}	IC_{50} Ratios 3T3:NHK
Carbon tetrachloride	NA	NA	NA	NA	NA
Methanol	NA	NA	1529[3]	383.2	NA
Aminopterin	0.006	0.0001	669	167.7	0.00001
Triphenyltin hydroxide	0.017	0.0004	0.01	0.003	1.7
Colchicine	0.034	0.001	0.007	0.002	4.9
Cycloheximide	0.187	0.004	0.073	0.02	2.6
Triethylenemelamine	0.272	0.007	1.85	0.5	0.1
Cadmium II chloride	0.518	0.01	1.84	0.5	0.3
Sodium dichromate dihydrate	0.587	0.01	0.721	0.2	0.8
Sodium arsenite	0.759	0.02	0.477	0.1	1.6
Arsenic trioxide	1.96	0.05	5.26	1.3	0.4
Mercury II chloride	4.12	0.1	5.8	1.5	0.7
Hexachlorophene	4.19	0.1	0.029	0.01	144.5
Thallium I sulfate	5.74	0.1	0.152	0.04	37.8
Haloperidol	6.13	0.1	3.36	0.8	1.8
Endosulfan	6.35	0.2	2.13	0.5	3.0
Amitriptyline HCl	7.05	0.2	8.96	2.2	0.8
Diquat dibromide monohydrate	8.04	0.2	4.48	1.1	1.8
Propranolol	13.9	0.3	35.3	8.8	0.4
Dichlorvos	17.7	0.4	10.7	2.7	1.7
Paraquat	20.1	0.5	61.6	15.4	0.3
Fenpropathrin	24.2	0.6	2.43	0.6	10.0
Physostigmine	25.8	0.6	88.5	22.2	0.3
Propylparaben	26.1	0.6	16.6	4.2	1.6
Sodium selenate	29	0.7	10.2	2.6	2.8
Potassium cyanide	34.6	0.8	29	7.3	1.2
Verapamil HCl	34.9	0.8	66.5	16.7	0.5
Parathion	37.4	0.9	30.3	7.6	1.2
Sodium oxalate	37.7	0.9	337	84.5	0.1
Sodium lauryl sulfate (SLS)*	***41.7***	***1.0***	***3.99***	***1.0***	***10.5***
Cupric sulfate pentahydrate	42.1	1.0	197	49.4	0.2
Acetaminophen	47.7	1.1	518	129.8	0.1
Dibutyl phthalate	49.7	1.2	28.7	7.2	1.7
Epinephrine bitartrate	59	1.4	87.4	21.9	0.7
Phenol	66.3	1.6	75	18.8	0.9
Atropine sulfate	76	1.8	81.8	20.5	0.9
Busulfan	77.7	1.9	260	65.2	0.3
Sodium I fluoride	78	1.9	49.8	12.5	1.6
Phenylthiourea	79	1.9	336	84.2	0.2

Table 5-8 Comparison of 3T3 and NHK NRU IC_{50} Geometric Means

Reference Substance	3T3 NRU		NHK NRU		
	Geometric Mean[1] IC_{50} (µg/mL)	Ratio Geometric Mean IC_{50} to SLS IC_{50}	Geometric Mean[1] IC_{50} (µg/mL)	Ratio Geometric Mean IC_{50} to SLS IC_{50}	IC_{50} Ratios 3T3:NHK
Carbamazepine	103	2.5	83.2	20.9	1.2
Diethyl phthalate	107	2.6	120	30.1	0.9
Lindane	108	2.6	18.7	4.7	5.8
Chloramphenicol	128	3.1	348	87.2	0.4
Disulfoton	133	3.2	270	67.7	0.5
Caffeine	153	3.7	638	159.9	0.2
Strychnine	158	3.8	62.5	15.7	2.5
Glutethimide	174	4.2	174	43.6	1.0
Chloral hydrate	183	4.4	133	33.3	1.4
Nicotine	361	8.7	107	26.8	3.4
Procainamide HCl	441	10.6	1741	436.3	0.3
Digoxin	466	11.2	0.001	0.0003	466000.0
Meprobamate	519	12.4	357	89.5	1.5
Lithium I carbonate	562[2]	13.5	468	117.3	1.2
Phenobarbital	573	13.7	448	112.3	1.3
Acetylsalicylic acid	676	16.2	605	151.6	1.1
Xylene	721[2]	17.3	466[2]	116.8	1.5
Citric acid	796	19.1	400	100.3	2.0
Trichloroacetic acid	902	21.6	413	103.5	2.2
Valproic acid	916	22.0	512	128.3	1.8
Sodium hypochlorite	1103	26.5	1502	376.4	0.7
5-Aminosalicylic acid	1667	40.0	46.7	11.7	35.7
Boric acid	1850	44.4	421	105.5	4.4
Lactic acid	3044	73.0	1304	326.8	2.3
Potassium I chloride	3551	85.2	2237	560.7	1.6
2-Propanol	3618	86.8	5364	1344.4	0.7
Sodium chloride	4730	113.4	1997	500.5	2.4
Dimethylformamide	5224	125.3	7760	1944.9	0.7
Ethanol	6523	156.4	10018	2510.8	0.7
Gibberellic acid	7810[3]	187.3	2856	715.8	2.7
Acetonitrile	7951	190.7	9528	2388.0	0.8
1,1,1-Trichloroethane	17248	413.6	8122[2]	2035.6	2.1
Ethylene glycol	24317	583.1	41852	10489.2	0.6
Glycerol	24655	591.2	24730	6198.0	1.0

Abbreviations: SLS=Sodium lauryl sulfate; NA=Not available; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes.
Reference substances ordered by 3T3 NRU IC_{50} values
[1]Geometric mean IC_{50} of the laboratory geometric mean values.
[2]Data available from only one laboratory.
[3]Data available from only two laboratories.
*Acceptable positive control (SLS) values from all study phases: N=293 (3T3); N=281 (NHK).

838 **Table 5-8** shows that there are nine reference substances for which the 3T3 and NHK NRU
839 IC_{50} values differ by at least one order of magnitude (i.e., 3T3 IC_{50}:NHK IC_{50} ≤0.1 or ≥10):
840 aminopterin, triethylenemelamine, hexachlorophene, thallium sulfate, fenpropathrin, sodium
841 oxalate, acetaminophen, digoxin, and 5-aminosalicylic acid. The IC_{50} values for SLS, also
842 differed by slightly more than one order of magnitude in the two NRU test methods (41.7
843 μg/mL for 3T3 and 3.99 μg/mL for NHK). One test method was not more consistently
844 sensitive (i.e., produced lower IC_{50} values) than the other for these nine reference substances.
845 The 3T3 NRU test method was more sensitive than the NHK NRU test method for four of
846 the nine substances: aminopterin, triethylenemelamine, sodium oxalate, and acetaminophen.
847 The NHK NRU test method was more sensitive than the 3T3 NRU test method for five
848 substances: hexachlorophene, thallium sulfate, fenpropathrin, digoxin, and 5-aminosalicylic
849 acid. Despite the normalization procedure, the reference substance IC_{50}:SLS IC_{50} ratios for
850 the two methods were still greater by at least one order of magnitude for six of the nine
851 substances (aminopterin, triethylenemelamine, hexachlorophene, sodium oxalate,
852 acetaminophen, and digoxin) and the order of magnitude difference increased for all six
853 substances. A number of factors could potentially be responsible for these differences
854 between the 3T3 and NHK NRU IC_{50} values:

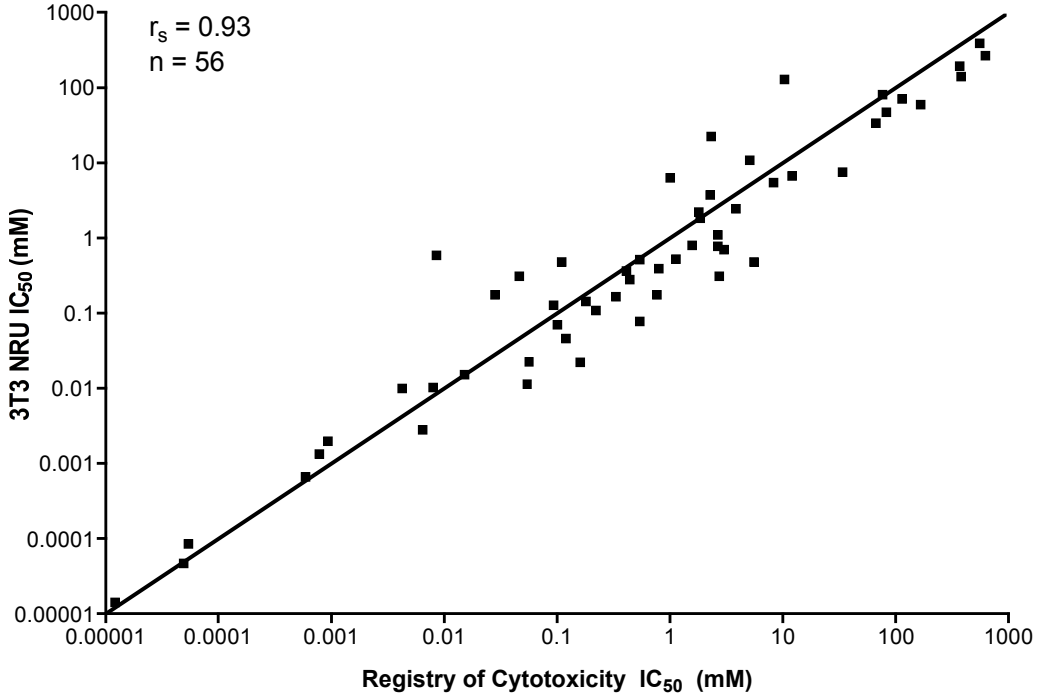
855 • Cell culture conditions (i. e., the 3T3 treatment medium contains serum while
856 the NHK treatment medium does not; differences in cell density in the
857 treatment medium)
858 • Differences in sensitivity between the fibroblast cell line and primary
859 keratinocytes
860 • Differences in sensitivity between human and mouse cells
861 • Differences in metabolic activity between the cell types
862
863 These factors may affect the results for some substances more than others. For example, a
864 substance that binds to serum proteins would be less available to the 3T3 cells (which have
865 serum in their growth medium) than to NHK cells (which are grown without serum). No
866 additional testing was performed to investigate the differences between the 3T3 and NHK
867 NRU IC_{50} values.
868

869 Two substances, digoxin and aminopterin, have IC_{50} values that differ by five orders of
870 magnitude between the two NRU test methods. Digoxin was much more toxic to the NHK
871 cells and aminopterin was more toxic to the 3T3 cells. Both substances are known substrates
872 for organic anionic transporters (OAT) (ICCVAM 2006). Such transporters are important for
873 *in vivo* toxicity responses in terms of the ability of challenge substances to be absorbed, reach
874 target tissues, accumulate, or be excreted. The differential susceptibilities of the 3T3 and
875 NHK cells may be explained by differential functioning of OAT between the cell types.
876 Although species and tissue differences in OAT have been reported (Sekine et al. 2000;
877 Miyazaki et al. 2004), the reason for these differential sensitivities is not known.
878
879 The 3T3 IC_{50}:NHK IC_{50} ratios shown in **Table 5-8** were used to determine the frequency
880 distributions shown in **Table 5-9.** These distributions indicate that the 3T3 and NHK NRU
881 IC_{50} values were within one order of magnitude of each other for 85% of the reference
882 substances (obtained by adding 38.9% and 45.8% for the $0.1 < IC_{50}$ ratio ≤ 1 and $1 < IC_{50}$
883 ratio <10 ranges). Ninety-three percent of the reference substances have 3T3 and NHK NRU
884 IC_{50} values within two orders of magnitude of each other (obtained by adding 4.2% each for
885 the $10 \leq IC_{50}$ ratio ≤ 100 and $0 < IC_{50}$ ratio ≤ 0.1 ranges to the 85% above).
886
887 **Table 5-9 Frequency of 3T3:NHK IC_{50} Ratios[1] for Reference Substances**
888

3T3:NHK IC_{50} Ratio Range	Number of Substances	% of Substances
IC_{50} Ratio <0.00001	1	1.4
$0 < IC_{50}$ Ratio ≤ 0.1	3	4.2
$0.1 < IC_{50}$ Ratio ≤ 1	28	38.9
$1 < IC_{50}$ Ratio <10	33	45.8
$10 \leq IC_{50}$ Ratio <100	3	4.2
$100 \leq IC_{50}$ Ratio <1000	1	1.4
IC_{50} Ratio ≥ 1000	1	1.4
Not Available	2	2.8

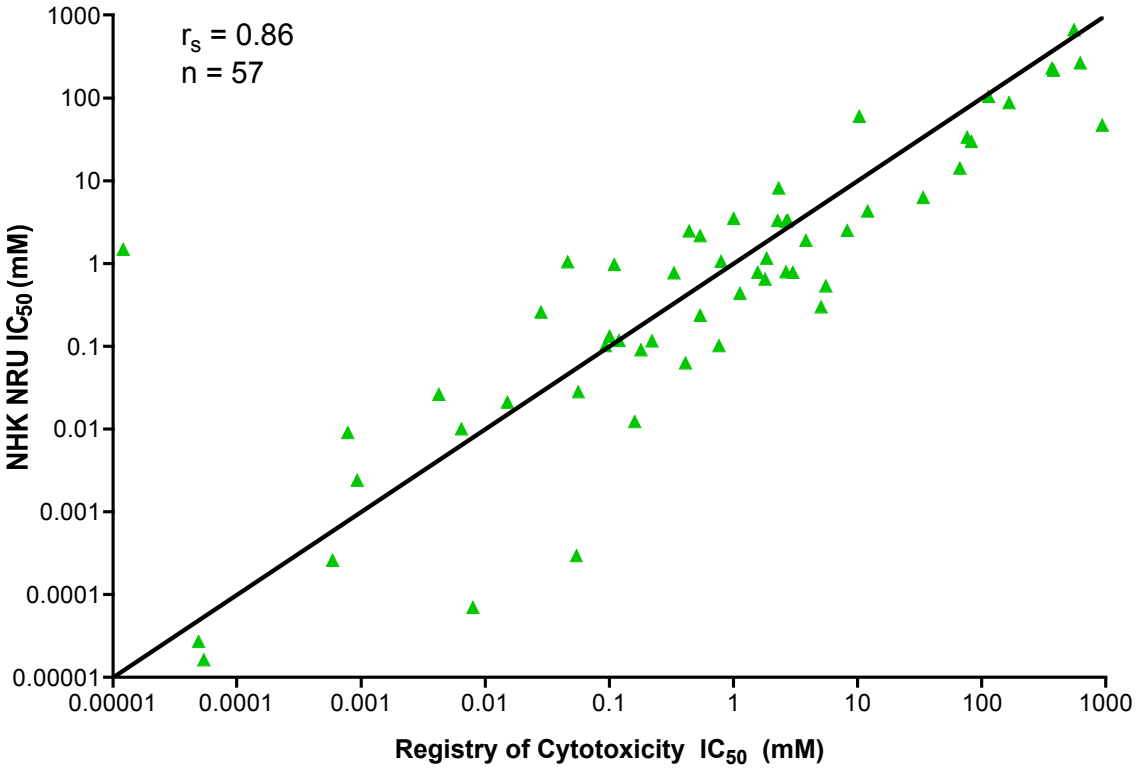
889 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes.
890 Note: Compiled using reference substance data from **Table 5-7**.
891
892 Correlations of the mean IC_{50} values for the reference substances common to the RC
893 database with the IC_{50} values (i.e., geometric mean of IC_{50} values obtained from the literature
894 for various basal cytotoxicity endpoints and cell types) from the RC (Halle 1998, 2003) are

895 shown in **Figure 5-3** (3T3 values) and **Figure 5-4** (NHK values). Although the validation
896 study tested 58 RC substances in common with the RC, IC_{50} values were obtained for 56
897 substances using the 3T3 NRU test method and 57 substances using the NHK NRU test
898 method. Spearman correlation analyses of the log-transformed IC_{50} data (in mM) indicated
899 that the NRU IC_{50} values were significantly correlated with the RC IC_{50x} values ($p<0.001$, for
900 both the 3T3 and NHK NRU test methods). The Spearman correlation coefficient, r_s, was
901 0.93 for the 3T3 values and 0.86 for the NHK values.
902
903 **Figure 5-3 RC IC_{50} Values vs 3T3 NRU IC_{50} Values for the 56 Substances in**
904 **Common**



905
906 Abbreviations: n=Number of substances; r_s=Spearman correlation coefficient; mM=Millimolar
907 The diagonal line indicates the predicted values for a 1:1 correspondence. No IC_{50} values were obtained for
908 carbon tetrachloride or methanol because of insufficient toxicity. The Registry of Cytotoxicity IC_{50} values are
909 geometric means of IC_{50} values obtained from the literature for various basal cytotoxicity endpoints and cell
910 types.
911

Figure 5-4 RC IC_{50} Values vs NHK NRU IC_{50} Values for the 57 Substances in Common



Abbreviations: n=Number of substances; r_s=Spearman correlation coefficient; mM=Millimolar.
The diagonal line indicates the predicted values for a 1:1 correspondence. No IC_{50} values were obtained for methanol because of insufficient toxicity . The Registry of Cytotoxicity IC_{50} values are geometric means of IC_{50} values obtained from the literature for various basal cytotoxicity endpoints and cell types.

5.7 Availability of Data

All data were provided to the SMT as electronic files and paper copies. The laboratories also maintained copies of all raw data and the electronic files. The individual test data and IC_{50} results for both passing and failing tests are provided in **Appendix I** for the reference substances and the PC.

5.8 Solubility Test Results

A solubility protocol (see **Section 2-7** and **Appendix B3**) designed to identify the solvent that would provide the highest concentration of a reference substance for *in vitro* testing was evaluated. Each laboratory performed solubility tests on all reference substances. However, to avoid the use of different solvents by the laboratories when testing the same substance, which might increase the variability of the IC_{50} results among the laboratories, the SMT assigned the solvents to be used (see **Table 5-10**). The objectives of the solubility testing were to evaluate the utility and appropriateness of the solubility protocol, and to evaluate the concordance among laboratories in selecting the solvents for each of the 72 reference substances.

5.8.1 Solubility Data

BioReliance evaluated the solubility of the reference substances, first in media, then in DMSO, and then in ETOH, at 400 and 200 mg/mL. Based on their experience, a solubility protocol was developed for the testing laboratories. This revised protocol required testing at lower concentrations, and use of the various solvents at concentrations that would be equivalent when applied to the cell cultures (see **Table 2-5**). The solubility flow chart (**Figure 2-7**) illustrates the tests for solubility in 3T3 and NHK medium, DMSO, and ETOH. **Table 5-10** provides the solubility test results.

944

Table 5-10 Solubility Test Results (mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Phase I																	
Arsenic III trioxide	0.25	0.05	<2	<2	Medium	***0.025[6]***	***0.025[6]***	***<0.2***	***<0.2***	***0.135[6]***	***0.135[6]***	***<0.2***	***<0.2***	***<0.02[6]***	***<0.02[6]***	***<0.2***	***<0.2***
Ethylene glycol	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Propranolol HCl	<2	10	200	20	DMSO	0.2	2	200	NT	20	20	200	NT	20	2	NT	NT
Phase II																	
Aminopterin	2	2	NT	NT	DMSO	2.0	<2	200	NT	<2	2	200	NT	0.2	0.2	200	NT
Cadmium II chloride	<2	<2	200	<200	DMSO	<2	<2	200	NT	<2	<2	200	NT	<0.2	<0.2	20	<20
Chloramphenicol	2	2	400	<200	DMSO	2.0	<2	200	NT	<2	<2	200	NT	0.2	0.2	20	20
Colchicine	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Lithium I carbonate	0.25	10	<2	NT	Medium	0.2	2.0	<20	<20	0.2	2	<200	<200	0.2	2	<2	<2
Potassium I chloride	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
2-Propanol	400	400	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium I fluoride	20	20	<200	<200	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium selenate	200	200	<200	<200	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Phase III																	
Acetaminophen	10	10	400	<200	DMSO	2	2	NT	NT	2	2	NT	NT	<2	<2	200	NT
Acetonitrile	400	400	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Acetylsalicylic acid	10	10	400	200	DMSO	2	2	NT	NT	<2	<2	200	NT	2	2	NT	NT
5-Aminosalicylic acid	2	2	<200	<200	Medium	2	2	NT	NT	2	2	NT	NT	2	2	NT	NT
Amitriptyline HCl	200	200	NT	NT	DMSO	<2	<2	200	NT	<2	<2	200	NT	0.2	0.2	200	NT

Table 5-10 Solubility Test Results (mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Atropine sulfate	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Boric aid	40	40	200	<200	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Busulfan	<2	<2	40	<200	DMSO	<2	<2	200	NT	<2	<2	50[6]	<200	<0.2	<0.2	20	<200
Caffeine	10	10	20	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Carbamazepine	<2	<2	40	<200	DMSO	0.2	0.2	20	20	<2	<2	200	NT	<0.2	<0.2	2	<20
Carbon tetrachloride	2	10	NT	NT	DMSO	20	20	NT	NT	<0.2	<0.2	2	NT	20	20	NT	NT
Chloral hydrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Citric acid	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Cupric sulfate pentahydrate	1	0.5	<2	2	Medium	2	0.2	<200	<200	2	2	NT	NT	0.2	0.2	<200	NT
Cycloheximide	20	20	400	<200	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Dibutyl phthalate	<2	<2	400	400	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Dichlorvos	10	10	NT	NT	DMSO	2	2	NT	NT	<2	<2	200	NT	2	2	NT	NT
Diethyl phthalate	<2	<2	400	400	DMSO	<2	<2	200	NT	0.2	<0.2	200	NT	<2	<2	200	NT
Digoxin	0.05	0.05	200	< 200	DMSO	<2	<2	200	NT	<0.2	<0.2	200	NT	<2	<2	200	NT
Dimethylformamide	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Diquat dibromide monohydrate	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Disulfoton	<2	<2	500	NT	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Endosulfan	<0.05	<0.05	40	NT	DMSO	<0.2	<0.2	20	<200	<0.2	<0.2	2	<200	<0.2	<0.2	20	<200
Epinephrine bitartrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	2	2	NT	NT
Ethanol	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Fenpropathrin	<20	<20	500	NT	DMSO	<2	<2	200	NT	<0.2	<0.2	200	NT	<2	<2	200	NT

Table 5-10 Solubility Test Results (mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Gibberellic acid	10	10	NT	NT	Medium	2	2	NT	NT	2	2	NT	NT	2	2	NT	NT
Glutethimide	<2	<2	500	NT	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Glycerol	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Haloperidol	<20	<20	40	NT	DMSO	<0.2	<0.2	20	<20	<0.2	<0.2	20	<20	<2	<2	20	<20
Hexachlorophene	0.05	<0.05	400	400	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Lactic acid	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Lindane	<0.05	<0.05	400	<200	DMSO	<2	<2	200	NT	<2	<2	200	NT	<0.2	<0.2	20	<200
Meprobamate	1	1	200	NT	DMSO	2	2	200	NT	2	2	200	NT	<0.2	<0.2	200	NT
Mercury II chloride	0.125	0.125	400	<200	DMSO	<2	<2	200	NT	<2	<2	200	NT	<0.2	<0.2	200	NT
Methanol	40	40	400	400	DMSO	20	20	NT	NT	20	20	NT	NT	<2	<2	200	NT
Nicotine	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Paraquat	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Parathion	0.05	<0.05	400	400	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Phenobarbital	2	2	200	<200	DMSO	2	2	NT	NT	<2	<2	200	NT	<2	<2	200	NT
Phenol	40	40	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Phenylthiourea	2	2	400	<200	DMSO	2	<2	200	NT	20	20	NT	NT	<2	<2	200	NT
Physostigmine	2	2	400	200	DMSO	2	2	NT	NT	<2	<2	200	NT	<2	<2	200	NT
Potassium cyanide	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Procainamide HCl	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Propylparaben	0.25	0.25	400	400	DMSO	<2	<2	200	NT	<2	<2	200	NT	<2	<2	200	NT
Sodium arsenite	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT

Table 5-10 Solubility Test Results (mg/mL)

Reference Substance	BioReliance[1]				SMT[2] Selection	ECBC[3]				FAL[3]				IIVS[3]			
	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH		3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH	3T3[4] Medium	NHK[5] Medium	DMSO	ETOH
Sodium chloride	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium dichromate dihydrate	400	400	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium hypochlorite	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Sodium oxalate	<0.05	20	0.125	<0.05	Medium	<0.2	20	0.2	<2	20	20	NT	NT	***<0.2***	***<0.2***	***<0.2***	***<0.2***
Strychnine	< 2	<2	2	2	Medium	0.2	<0.2	2	2	0.2	0.2	<200	<200	***<0.2***	***<0.2***	***<0.2***	***<0.2***
Thallium I sulfate	1	0.5	<2	<2	Medium	0.2	0.2	<200	<200	***<0.2***	***<0.2***	***<0.2***	***<0.2***	0.2	0.2	<20	<200
Trichloroacetic acid	200	200	NT	NT	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
1,1,1-Trichloroethane	10	10	400	400	Medium	20	20	NT	NT	20	20	NT	NT	20	20	NT	NT
Triethylenemelamine	<2	<2	2	<20	DMSO	0.2	0.2	<200	<200	<0.2	<0.2	2	<2	***<0.2***	***<0.2***	***<0.2***	***<0.2***
Triphenyltin hydroxide	<0.05	<0.05	10	<20	DMSO	<0.2	<0.2	2	<20	<0.2	<0.2	2	<200	<2	<2	2	<20
Valproic acid	10	2	NT	NT	DMSO	2	2	NT	NT	<2	<2	200	NT	2	<2	200	NT
Verapamil HCl	<0.05	0.25	200	NT	DMSO	<2	<2	200	NT	<2	<2	200	NT	<0.2	<0.2	20	NT
Xylene	1	1	500	NT	DMSO	<2	<2	200	NT	2	<2	200	NT	<2	<2	200	NT

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; SMT=Study Management Team; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; DMSO=Dimethyl sulfoxide; ETOH=ethanol; NT=Not tested.
Note: Table sorted by study phase and alphabetical by substance.
[1]The solubility protocol used was different from that used by the testing laboratories.
[2]Solvents selected by the SMT for cytotoxicity testing. The BioReliance results were used to determine solvents for Phases I and II. Results from all laboratories were used to determine solvents for Phase III. 3T3 and NHK media were treated as a single solvent. If a substance insoluble in one medium, and not the other, and soluble in DMSO, then DMSO was selected for use with both cell types.
[3]Used protocol in **Figure 2-7**.
[4]Dulbecco's Modification of Eagle's Medium.
[5]Keratinocyte Growth Medium (KGM® from CAMBREX Clonetics®).
[6]The results were obtained using a deviation from the standard protocol.
Laboratories agreed on solvent. Laboratories did not agree on solvent. **bold** Protocol did not provide enough guideline information to select a single solvent.

5.8.2 Solubility and Volatility Effects in the Cytotoxicity Tests

The laboratories reported solubility results for the stock solutions of reference substance for each 3T3 and NHK test. Prior to the addition of the NR dye medium, the laboratories visually observed the test cultures and documented noticeable precipitate. **Table 5-11** illustrates the existence of solubility issues (in both the 3T3 and NHK NRU test methods) as evidenced by the observation of precipitates with some reference substances. **Sections 3.3** and **5.4.2** provide additional information on ability of the laboratories to achieve sufficient toxicity for the calculation of an IC_{50} in the presence of limited solubility. **Table 5-11** also notes the presence of volatility, as indicated by the use of film plate sealers during incubation.

Table 5-11 Reference Substances with Precipitate (PPT) and Volatility Issues[1]

Reference Substances	3T3 NRU Test Method				NHK NRU Test Method			
	PPT in 2X Stock Dilutions	PPT in 1X Plate Dilutions	PPT in Stock and Plate Dilutions	Volatility	PPT in 2X Stock Dilutions	PPT in 1X Plate Dilutions	PPT in Stock and Plate Dilutions	Volatility
Acetonitrile				X				X
Aminopterin		X			X			
5-Aminosalicylic acid	X							
Arsenic III trioxide	X				X			
Cadmium II chloride		X					X	
Carbamazepine			X					
Carbon tetrachloride			X		X			
Citric acid						X		
Cupric sulfate pentahydrate						X		
Dibutyl phthalate		X					X	
Dichlorvos				X				X
Diethyl phthalate	X						X	
Digoxin			X					
Dimethylformamide						X		
Disulfoton			X				X	
Endosulfan	X			X				X
Ethanol				X				X
Fenpropathrin			X				X	
Gibberellic acid	X				X			
Glutethimide					X			
Lindane			X	X			X	
Lithium I carbonate	X				X			
Nicotine				X				X
Parathion	X						X	
Phenol				X				X
Potassium I chloride		X						
Potassium cyanide		X		X				X
2-Propanol				X				X
Sodium arsenite		X						X
Sodium chloride						X		
Sodium I fluoride		X				X		
Sodium hypochlorite				X				
Sodium oxalate			X			X		

Table 5-11 Reference Substances with Precipitate (PPT) and Volatility Issues[1]

Reference Substances	3T3 NRU Test Method				NHK NRU Test Method			
	PPT in 2X Stock Dilutions	PPT in 1X Plate Dilutions	PPT in Stock and Plate Dilutions	Volatility	PPT in 2X Stock Dilutions	PPT in 1X Plate Dilutions	PPT in Stock and Plate Dilutions	Volatility
Strychnine	X				X			
Trichloroacetic acid						X		
1,1,1-Trichloroethane	X						X	
Valproic acid	X							
Verapamil HCl					X			
Xylene	X				X			

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; PPT=Precipitate.
Note: Table sorted alphabetical by reference substance.
[1]Results are based on at least one laboratory having precipitate or volatility issues with a substance. Volatility was denoted by the use of plate sealers during testing. 2X stock dilutions are prepared for each of 8 test substance concentrations. 1X plate dilutions are the result of diluting the 2X stock solutions with medium in the 96-well plates.

5.9 Summary

- The BioReliance, ECBC, and IIVS laboratories performed the 3T3 and NHK NRU tests in compliance with GLP guidelines.
- The quality and consistency of the reference substances was maintained during the study by the central purchase and distribution of individual lots of reference substances to the testing laboratories.
- Modifications and revisions made to the protocols during Phases I and II contributed to the optimization of the final protocols used in Phase III of the study. As a general rule, the protocol changes enhanced the performance of the methods and allowed more tests to meet the acceptance criteria.
- FAL improved the quality of its NHK data prior to Phase II testing by modifying the methods used to propagate the cells. Positive control IC_{50} data in Phases II and III from FAL more closely resemble the data from the other laboratories.
- Summary test data and IC_{50} results are presented in tabular and graphic formats. Comparisons of 3T3 NRU IC_{50} values to NHK NRU IC_{50} values show that the values for 85% of the reference substances are within one order of magnitude of each other. Digoxin and aminopterin yielded differences of

up to five orders of magnitude when the IC_{50} values of the 3T3 and NHK NRU test methods were compared.

- Although each laboratory followed the same solubility protocol, they sometimes obtained different results. This may have been due to the subjective judgment of whether or not solubility was achieved. Additionally, the laboratories may have used solubility procedures that were beyond the level of detail in the solubility protocol.

6.0 ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS 6-3

6.1 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Rodent Acute Oral Toxicity 6-6
6.1.1 Linear Regression Analyses for the Prediction of Rat Acute Oral LD_{50} Values from *In Vitro* IC_{50} Values 6-6
6.1.2 Comparison of Combined-Laboratory 3T3 and NHK Regressions to the RC Millimole Regression 6-9

6.2 Analysis of Outlier Substances for the RC Millimole Regression.....6-9
6.2.1 Identification of Outlier Substances.....6-11
6.2.2 Evaluation of Outlier Substances.....6-14

6.3 Improving the Prediction of *In Vivo* Rat Oral LD_{50} Values from *In Vitro* IC_{50} Data.....6-20
6.3.1 The RC Rat-Only Millimole Regression.....6-20
6.3.2 The RC Rat-Only Weight Regression.....6-21

6.4 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting GHS Acute Oral Toxicity Category 6-23
6.4.1 Prediction of GHS Acute Oral Toxicity Category by the RC IC_{50} Values Using the RC Rat-Only Millimole Regression.....6-24
6.4.2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Millimole Regression.....6-27
6.4.3 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression.....6-30
6.4.4 Summary of the Regressions Evaluated.....6-34

6.5 Correlation of NRU Concentration-Response Slope with Rat Lethality Dose-Response Slope 6-36

6.6 Strengths and Limitations of the *In Vitro* Cytotoxicity Test Methods with the IC_{50}–LD_{50} Regressions for Predictions of Rodent Acute Oral Toxicity.....6-38
6.6.1 *In Vitro* Cytotoxicity Methods.....6-38
6.6.2 Use of Mole-Based vs. Weight-Based Regressions for the Prediction of Toxicity for Low and High Molecular Weight Substances.....6-41

6.7 Salient Issues of Data Interpretation 6-44

6.8 Comparison of NRU Test Results to Established Performance Standards.....6-44

6.9 Summary.....6-51

[This Page Intentionally Left Blank]

6.0 ACCURACY OF THE 3T3 AND NHK NRU TEST METHODS

This section discusses the accuracy of the 3T3 and NHK NRU test methods for predicting the rodent acute oral toxicity (the LD_{50}) of chemicals. Accuracy, the agreement between a test result and an accepted reference value, is a critical component of the evaluation of the validation status of a method (ICCVAM 2003). Although the 3T3 and NHK NRU test methods are not suitable as replacements for acute oral toxicity assays, the rationale for evaluating the accuracy of LD_{50} predictions from the *in vitro* IC_{50} values is that the animal savings produced by using these *in vitro* test methods to predict starting doses for acute oral toxicity assays will be greatest when the starting dose is as close as possible to the "true" LD_{50} value (see **Section 10** for the evaluation of the potential reduction of animal use).

The ability of the 3T3 and NHK NRU test methods to correctly predict rodent acute oral toxicity is based on the appropriateness of the *in vivo – in vitro* (i.e., IC_{50}-LD_{50}) regression model. The IC_{50}-LD_{50} regression establishes the relationship between the *in vitro* IC_{50} values and the LD_{50} values that will be used to set the starting doses for the computer-simulated acute oral toxicity assays in this study (see **Section 10**). The regressions generated by the three laboratories for each NRU test method were not statistically different, and the data from the 3T3 and NHK NRU test methods were combined (using a geometric mean IC_{50} of the three individual laboratory geometric mean IC_{50} values) into single regressions (see **Section 6.1**). Only rat LD_{50} data were used for these regressions to reduce the variation that would be produced by combining data from multiple species. **Table 6-1** describes the datasets used for the analyses in **Sections 6.1** through **6.4**.

To test the assumption in the *Guidance Document* that the RC millimole regression can be obtained using a basal cytotoxicity method with a single cell type and cytotoxicity endpoint (ICCVAM 2001b), the regressions for each NRU test method (3T3 and NHK) were compared with regressions for the same substances that were calculated using the RC IC_{50} and LD_{50} values (see **Section 6.1**). Because the 3T3 and NHK regressions were not statistically different from the RC regressions for the same chemicals, the RC data were used to develop a regression to predict LD_{50} values from the NRU-generated IC_{50} values because

90 this regression was based on a larger number of substances than the NICEATM/ECVAM
91 regressions (see **Section 6.3**).
92
93 **Table 6-1 Datasets Used for Accuracy Analyses[1]**
94

Use	3T3 NRU[1]	NHK NRU[1]	Characteristics of Dataset
Testing with NRU test methods	72	72	Substances tested; 58 substances were common to the RC
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 reference 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 reference rat oral LD_{50} values
RC millimole regression	NA	NA	RC IC_{50} (mM) and RC oral LD_{50} (mmol/kg) values for 347 substances (282 rat and 65 mouse LD_{50} values)
RC rat-only millimole regression	NA	NA	RC IC_{50} (mM) and RC oral LD_{50} values (mmol/kg) for 282 substances with rat oral LD_{50} data
RC rat-only weight regression	NA	NA	RC IC_{50} (µg/mL) and RC oral LD_{50} values (mg/kg) for 282 substances with rat oral LD_{50} data
Analysis of outliers for the RC millimole regression	67	68	Substances with IC_{50} values from at least one laboratory
Prediction of GHS accuracy using IC_{50} values in RC rat-only regressions	67	68	Substances with IC_{50} values from at least one laboratory

95 Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NA=Not applicable; NHK=Normal
96 human keratinocytes; NRU=Neutral red uptake.
97 [1]Number of substances.
98
99 The RC millimole regression was used to identify outlier substances (i.e., those that did not
100 fit the regression within the established acceptance limits; see **Section 6.2**) tested in the
101 validation study because:
102 • Acceptance limits for the RC millimole regression had been established
103 • The 3T3 and NHK NRU IC_{50} – rat oral LD_{50} regressions were not
104 significantly different from the RC millimole regressions calculated for the
105 same substances

106 • Use of the RC millimole regressions allow a comparison of the outlier
107 substances determined using RC data to those determined using the 3T3 and
108 NHK data
109
110 To improve upon the RC millimole regression's[1] ability to accurately predict LD_{50} values
111 from IC_{50} values, and to also make this approach relevant to the testing of mixtures and
112 substances without known molecular weights, two regressions were calculated (see **Section**
113 **6.3**). The first regression – the RC rat-only millimole regression – uses the 282 (of 347)
114 substances in the RC dataset that had reported rat LD_{50} values. The LD_{50} data for the
115 regression were limited to one species to decrease the variability in LD_{50} values that would
116 occur if the data from more than one species were combined. Rats were selected because they
117 are the preferred species for acute oral toxicity testing (EPA 2002b; OECD 2001a; OECD
118 2001d) (see **Section 6.3.1**). The RC rat-only millimole regression was transformed to one
119 based on weight units (mg/kg body weight for LD_{50} and µg/mL for IC_{50}) in order to make the
120 regression equation more generally applicable to the testing of mixtures and substances of
121 unknown molecular weights.
122
123 The ability of the 3T3 and NHK NRU IC_{50} data to correctly predict rat acute oral LD_{50} values
124 based on using the RC rat-only millimole regression and the RC rat-only weight regression,
125 was evaluated by determining the extent to which the appropriate GHS acute oral toxicity
126 category was identified for each reference substance (see **Section 6.4**). The rationale for
127 evaluating the accuracy of LD_{50} predictions is that the acute oral toxicity test methods (i.e.,
128 UDP, FDP, and ATC) call for starting doses to be placed as close as possible and just below
129 the true LD_{50}. When the starting dose is close to the true LD_{50} for a test substance, fewer
130 animals are needed. When the starting dose is below the true LD_{50}, there is reduced pain and
131 suffering because doses tend to be lower, and the test bias is more conservative. This
132 approach permits an assessment of accuracy that is specific to each GHS hazard

[1] The RC millimole regression was created using rat and mouse oral LD_{50} values from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 substances with known molecular weights (Halle 1998, 2003)

classification category. The discordant reference substances from the predictions of GHS category are presented in **Appendix L2**.

The remainder of **Section 6** discusses physical, chemical, and biological, characteristics of substances that may have an impact on the accuracy of the 3T3 and NHK methods.

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

The rat LD_{50} values provided in **Section 4.2** are used as the reference values for assessing the ability of the 3T3 and NHK test methods to accurately predict acute oral toxicity. The accuracy of the two *in vitro* cytotoxicity test methods is assessed in two ways: (1) by the goodness of fit of the *in vitro* IC_{50} data to the rat LD_{50} data in linear regression analyses, and (2) by the concordance (i.e., extent of agreement) between the GHS acute oral toxicity categories (UN 2005) assigned based on rat LD_{50} data and those predicted using *in vitro* IC_{50} values.

6.1.1 Linear Regression Analyses for the Prediction of Rat Acute Oral LD_{50} Values from *In Vitro* IC_{50} Values

As described in **Section 5.5.4.3**, linear regressions for each laboratory and *in vitro* method were calculated using log IC_{50} values (mM) versus the corresponding reference log LD_{50} values (mmol/kg) identified in **Table 4-2**. The reference substances used to calculate each of the laboratory regressions met the following criteria for each test method:

- The substance was included in the RC
- All three laboratories reported IC_{50} values
- There was an associated rat acute oral LD_{50} reference value (see **Table 4-2**).

There were 47 and 51 reference substances that fit these criteria for the 3T3 and NHK test methods, respectively. The slopes for the all of the laboratory-specific regressions were statistically significantly different from zero ($p < 0.0001$), which indicates a significant correlation between *in vitro* IC_{50} values and the corresponding rat acute oral LD_{50} values. Comparison of the individual laboratory regressions to one another using the goodness of fit F-test for regression slopes and intercepts described in **Section 5.5.4.3** indicated that the

laboratory-specific regressions for either NRU method were not significantly different from one another. For the 3T3 method, p =0.605 for the slope comparisons and p =0.947 for the intercept comparisons. For the NHK method, p =0.792 for the slope comparisons and p =0.999 for the intercept comparisons.

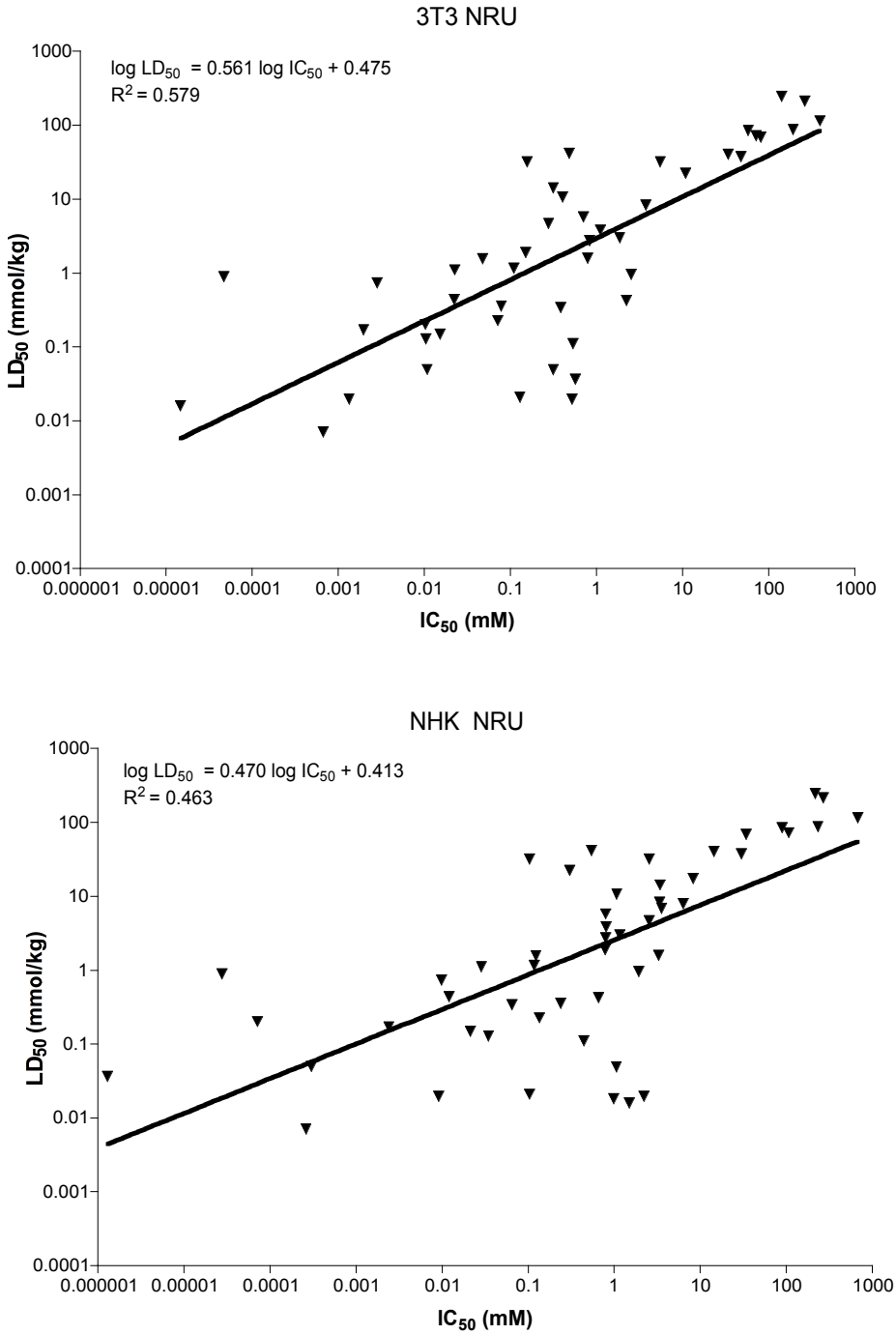
Because the individual laboratory regressions were not significantly different, the laboratory data were combined into a single regression for each method using the geometric mean of the mean IC_{50} values determined by each laboratory for each substance (see the "Combined-laboratory" regressions in **Table 6-2** and **Figure 6-1**). The combined-laboratory 3T3 regression yielded a better fit to the reference LD_{50} data (R^2=0.579) than the NHK regression (R^2=0.463).

Table 6-2 Linear Regression Analyses of the 3T3 and NHK and Rat Acute Oral LD_{50} Test Results[1]

Laboratory	N	Slope	Intercept	R^2
3T3 Method				
ECBC[2]	47	0.573	0.541	0.613
FAL[2]	47	0.539	0.373	0.519
IIVS[2]	47	0.552	0.507	0.586
Combined-laboratory[3]	47	0.561	0.475	0.579
NHK Method				
ECBC[2]	51	0.491	0.412	0.480
FAL[2]	51	0.428	0.407	0.422
IIVS[2]	51	0.483	0.416	0.478
Combined-laboratory[3]	51	0.470	0.413	0.463

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; 3T3=BALB/c 3T3 fibroblasts; N=Number of substances used to calculate the regression; NHK=Normal human keratinocytes; NRU=Neutral red uptake; R^2=Coefficient of determination.
[1]Log IC_{50} in mM; log LD_{50} in mmol/kg.
[2]Regression based on a single point per substance (i.e., the geometric mean of the within laboratory replicate IC_{50} values and the reference rat acute oral LD_{50} from **Table 4-2**).
[3]Regression based on a single point per substance (i.e., the geometric mean of the geometric mean IC_{50} values obtained for each laboratory and the reference rat acute oral LD_{50} from **Table 4-2**).

Figure 6-1 Combined-Laboratory 3T3 and NHK NRU Regressions



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; R^2=Coefficient of determination.
Points show the geometric means of the laboratory geometric mean IC_{50} values and the reference rat acute oral LD_{50} values (from **Table 4-2**) for 47 reference substances for the 3T3 and 51 reference substances for NHK test methods.
Solid lines show the combined-laboratory regressions for each method (see **Table 6-2**).

6.1.2 Comparison of the Combined-Laboratory 3T3 and NHK Regressions to the RC Millimole Regression

The validation study tested 58 RC substances using the 3T3 and NHK NRU test methods (see **Figure 3-1**). The resulting method regressions for each cell type were compared to the RC regressions for the same substances to test the assumption in the *Guidance Document* that the RC millimole regression can be obtained with a basal cytotoxicity test method using a single cell type and endpoint (ICCVAM 2001b). The 47 substances used to calculate these regressions met the following criteria:

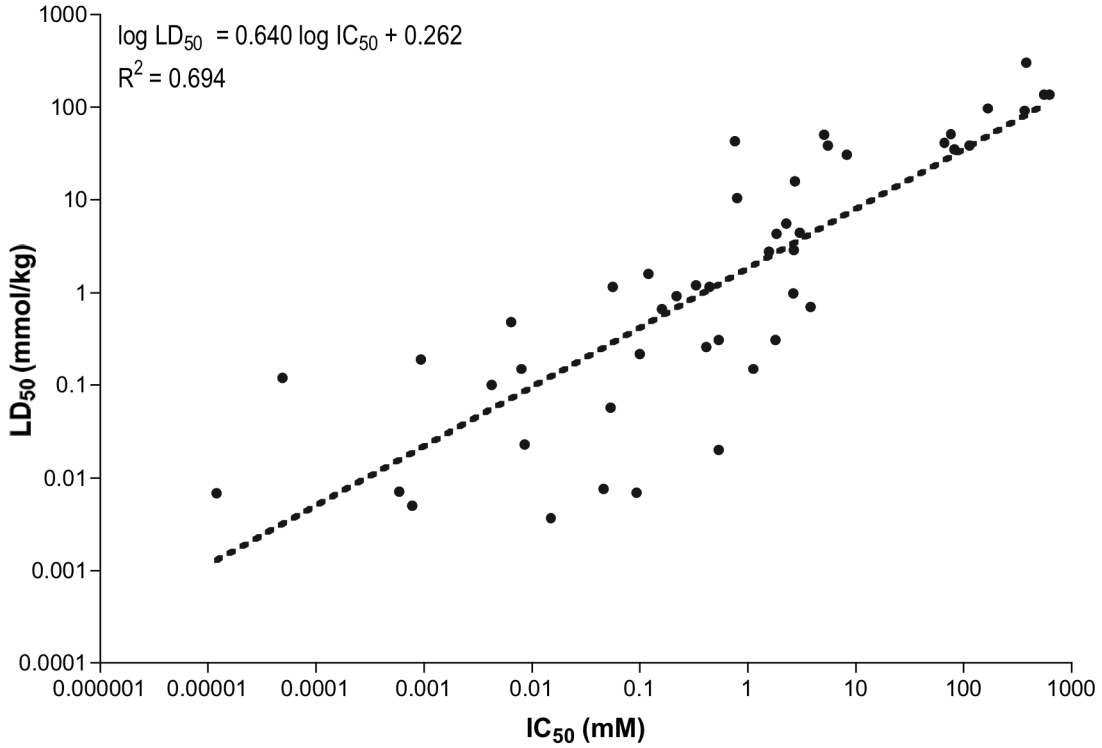
- The substance was included in the RC
- All three laboratories reported IC_{50} values for both the 3T3 and NHK NRU test methods
- There was an associated rat oral reference LD_{50} value (see **Table 4-2**)

The regression calculated for the 47 substances using the RC IC_{50} and LD_{50} data is shown in **Figure 6-2**. A graphic comparison of the RC regressions and the 3T3 and NHK combined-laboratory regressions is in **Figure 6-3**. A statistical comparison of slope and intercept (simultaneously) using an F test showed that neither the 3T3 regression (p=0.612) nor the NHK regression (p = 0.759) was significantly different from the 47 RC substance regression.

6.2 Analysis of Outlier Substances for the RC Millimole Regression

The RC millimole regression and each *in vitro* NRU test method were used to identify outliers among the reference substances tested in the validation study (i.e., those for which the rodent LD_{50} was not accurately predicted by the *in vitro* IC_{50}). The outlier substances were then evaluated to determine if they had common characteristics that could assist in identifying the types of substances that are not suited for use in the 3T3 or NHK NRU test methods for determining starting doses for acute oral toxicity assays.

Figure 6-2 Regression for 47 RC Substances Using RC Data

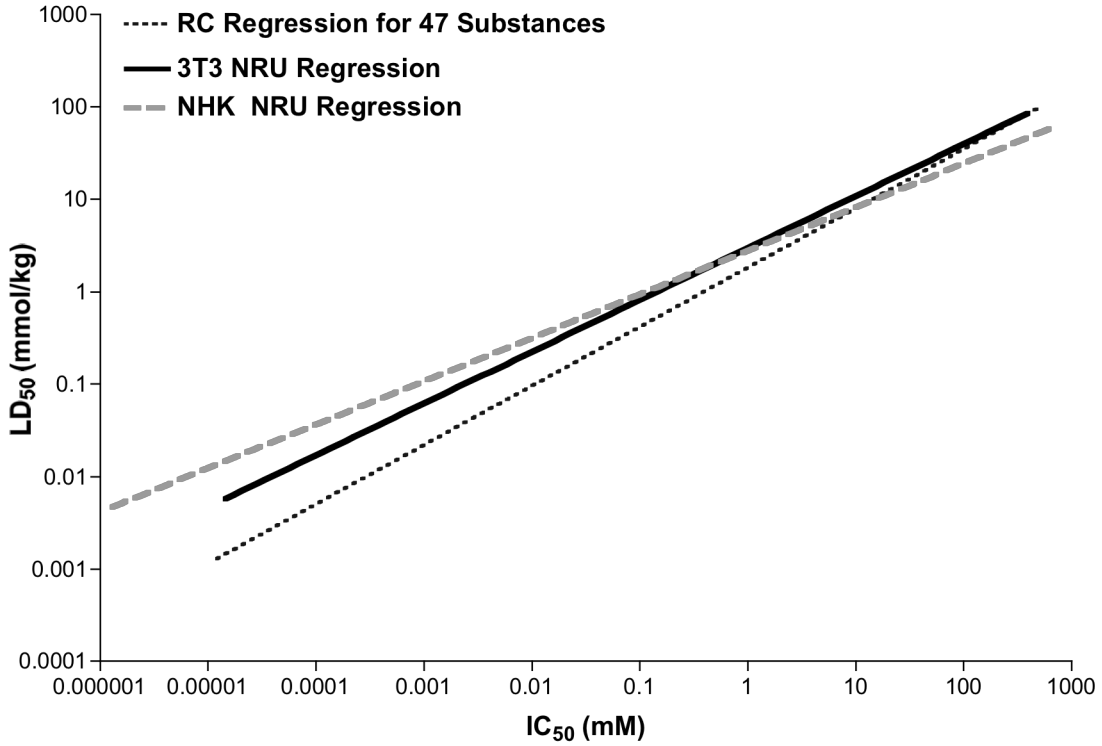


Abbreviations: RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
Points show the IC_{50} values and the reference rodent (rat and mouse) acute oral LD_{50} values from the RC for 47 reference substances. The dashed line shows the calculated regression.

The RC millimole regression was used to determine the outlier status of reference substances because:

- The RC millimole regression had associated acceptance limits (Halle 1998, 2003): a difference greater than 0.699 (or log 5) for log-observed LD_{50} (in mmol/kg) from the log-predicted LD_{50} identifies a substance as an outlier
- The 3T3 and NHK IC_{50} – rat oral LD_{50} regressions were not significantly different from the RC millimole regressions calculated for the same substances
- Use of the RC millimole regression allows a comparison of the outlier substances determined using RC IC_{50} values to those determined using the 3T3 and NHK NRU IC_{50} values.

Figure 6-3 Regression for 47 RC Substances with the 3T3 and NHK Regressions



Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=normal human keratinocytes; NRU=neutral red uptake.
The regression for 47 RC substances using RC data is log LD_{50} = 0.640 log IC_{50} + 0.262 (R^2=0.694). The combined-laboratory 3T3 regression for the same 47 substances, is log LD_{50} = 0.561 log IC_{50} + 0.475 (R^2 = 0.579) (from **Table 6-1**). The combined-laboratory NHK regression for the same 47 substances, is log LD_{50} = 0.471 log IC_{50} + 0.445 (R^2 = 0.487) (from **Table 6-2**).

6.2.1 Identification of Outlier Substances

For each *in vitro* NRU test method, the predicted LD_{50} values for the reference substances were determined using the geometric mean IC_{50} values of the three geometric mean laboratory values in the RC millimole regression. Outliers were identified using the RC method (Halle 1998): a difference greater than 0.699 (or log 5) for log-observed LD_{50} (in mmol/kg) minus the log-predicted LD_{50} identifies a substance as an outlier (see **Appendix J1** for the 3T3 NRU test method and **Appendix J2** for the NHK NRU test method for the predicted LD_{50} values). For the best comparison with the RC outlier results, the outlier evaluation for the 3T3 and NHK NRU test methods used same observed LD_{50} values as those used in the RC database for the 58 reference substances that were included in the RC

258 database (see **Table 3-2)**. For the non-RC substances, the observed values (in **Table 3-2**)
259 were obtained from other databases such as RTECS® or Hazardous Substances Database
260 (NLM 2002). The outlier analysis included all the reference substances that yielded IC_{50}
261 values from at least one laboratory in the validation study whether the *in vivo* LD_{50} values
262 were from rats or mice. Thus, 70 substances were used for the 3T3 NRU outlier analysis and
263 71 substances were used for the NHK NRU outlier analysis. **Table 6-3** lists the outlier
264 substances for the RC millimole regression when using the RC IC_{50} values and the 3T3 and
265 NHK NRU IC_{50} values.
266

Table 6-3 Outlier Substances for the RC and the 3T3 and NHK Methods When the RC Millimole Regression is Used[1]

Substances Included in the RC Identified as Outliers in:		
RC[2]	**3T3[3]**	**NHK[4]**
	Acetaminophen (+)	
	Arsenic III trioxide (–)	*Arsenic III trioxide (–)*
		Aminopterin (–)
5-Aminosalicylic acid (+)		5-Aminosalicylic acid (+)
Busulfan (–)	**Busulfan (–)**	**Busulfan (–)**
Caffeine (–)		Caffeine (–)
Cycloheximide (–)	**Cycloheximide (–)**	**Cycloheximide (–)**
Dibutyl phthalate (+)	***Dibutyl phthalate (+)***	***Dibutyl phthalate (+)***
	Diethyl phthalate (+)	***Diethyl phthalate (+)***
Digoxin (–)	***Digoxin (–)***	
Disulfoton (–)	***Disulfoton (–)***	***Disulfoton (–)***
Epinephrine bitartrate (–)	Epinephrine bitartrate (–)	Epinephrine bitartrate (–)
Ethanol (+)	**Ethanol (+)**	**Ethanol (+)**
Lindane (–)	*Lindane (–)*	
Mercury II chloride (–)	Mercury II chloride (–)	Mercury II chloride (–)
		Methanol (+)
Nicotine (–)	Nicotine (–)	Nicotine (–)
Paraquat (–)		Paraquat (–)
Parathion (–)	***Parathion* (–)**	***Parathion* (–)**
Phenobarbital (–)	Phenobarbital (–)	Phenobarbital (–)
Phenylthiourea (–)	Phenylthiourea (–)	Phenylthiourea (–)
Potassium cyanide (–)	Potassium cyanide (–)	Potassium cyanide (–)
Propylparaben (+)	Propylparaben (+)	Propylparaben (+)
		Sodium oxalate (–)

Table 6-3 Outlier Substances for the RC and the 3T3 and NHK Methods When the RC Millimole Regression is Used[1]

Substances Included in the RC Identified as Outliers in:		
RC[2]	**3T3[3]**	**NHK[4]**
Thallium I sulfate (–)	Thallium I sulfate (–)	
Triethylenemelamine (–)	Triethylenemelamine (–)	Triethylenemelamine (–)
1,1,1-Trichloroethane (+)		
Verapamil HCl (–)	**Verapamil HCl (–)**	***Verapamil HCl* (–)**
		Xylene (+)
Outliers That Were Not Included in the RC		
	Dichlorvos (–)	Dichlorvos (–)
	Endosulfan (–)	Endosulfan (–)
	Fenpropathrin (–)	*Fenpropathrin* (–)
	Physostigmine (–)	Physostigmine (–)
	Sodium hypochlorite (+)	Sodium hypochlorite (+)
	Sodium selenate (–)	Sodium selenate (–)
	Strychnine (–)	*Strychnine* (–)

Abbreviations: RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; (–)=Toxicity was underpredicted by the IC_{50} and RC millimole regression (i.e., the LD_{50} value predicted by the IC_{50} was higher than the *in vivo* LD_{50} value); (+)=Toxicity was overpredicted by the IC_{50} and RC millimole regression (i.e., the LD_{50} value predicted by the IC_{50} was lower than the *in vivo* rodent LD_{50} value).
[Note: Empty cells indicate that the substance was not an outlier for that particular IC_{50} value.]
[1]Log LD_{50} (mmol/kg) = 0.435 log IC_{50} (mM) + 0.625. Log LD_{50} (mmol/kg) values for outlier substances were >0.699 from the RC millimole regression.
[2]Using RC IC_{50} in the RC millimole regression for the 58 RC substances tested in the validation study.
[3]Using the 3T3 NRU IC_{50} in the RC millimole regression for the 70 reference substances that yielded IC_{50} values from any laboratory in the validation study.
[4]Using the NHK NRU IC_{50} in the RC millimole regression the RC for the 71 reference substances that yielded IC_{50} values from any laboratory in the validation study.
Bolded substances have active metabolites *in vivo* (see **Table 3-7**).
Substances that showed evidence of insolubility (i.e., precipitates) during testing (see **Table 5-11**) are identified by italics.

When the RC millimole regression and the RC method of identifying outlier substances were used (Halle 1998, 2003), there were 28 outliers for the 3T3 NRU test method and 31 for the NHK NRU test method. The top part of **Table 6-3** shows a comparison of the 22 RC substances that were identified by the RC as outliers (see **Table 3-2**) and the RC reference substances that were identified as outliers using either the 3T3 or NHK NRU IC_{50} values with the RC millimole regression. For the 58 RC substances that were tested in the validation study, 18 of the 22 RC outliers also responded as outliers in both NRU test methods, but some of the substances were outliers only in one of the two NRU test methods. The RC regression outliers, 5-aminosalicylic acid, caffeine, paraquat, and 1,1,1-trichloroethane were not outliers when 3T3 data were used, and the RC outliers, digoxin, lindane, thallium sulfate,

292 and 1,1,1-trichloroethane, were not outliers when the NHK NRU test method was used. In
293 contrast the 3T3 NRU test method identified three substances as outliers that were not
294 identified by the RC: acetaminophen, arsenic trioxide, and diethyl phthalate, and the NHK
295 NRU test method identified six: aminopterin, arsenic trioxide, diethyl phthalate, methanol,
296 sodium oxalate, and xylene. Seven additional substances, that were not included in the RC
297 database, were identified as outliers using the NRU IC_{50} values in the RC millimole
298 regression: dichlorvos, endosulfan, fenpropathrin, physostigmine, sodium hypochlorite,
299 sodium selenate, and strychnine.

300 6.2.2 Evaluation of Outlier Substances

301 A number of physico-chemical and toxicologic characteristics were evaluated for their
302 frequency of occurrence among the 28 and 31 outlier substances in the 3T3 and NHK NRU
303 test methods, respectively, to identify attributes that may have contributed their outlier status.
304 This section provides a summary of these analyses based on the RC millimole regression and
305 outlier criteria. The frequency of outliers versus the total number of reference substances for
306 each physico-chemical and toxicologic category examined is shown in **Appendix L1**.

307 6.2.2.1 *Physical Characteristics*

308 A number of physical characteristics were evaluated for their frequency of occurrence in the
309 set of outlier substances versus the complete set of reference substances. The characteristics
310 chosen were those that were assumed to be readily available, or relatively easy to measure,
311 for new substances that may be tested in these NRU assays. The characteristics examined
312 included chemical class, molecular weight, boiling point, IC_{50}, pH, and log K_{ow} (i.e., log
313 octanol:water partition coefficient). Unfortunately, these attributes were not available for all
314 substances. For example, log K_{ow} was available for 50 of the 70 (71%) substances evaluated
315 for the 3T3 NRU test method and for 51 of the 71 (72%) substances evaluated for the NHK
316 NRU test method. Boiling point was available for only 24 of 70 (34%) substances evaluated
317 for the 3T3 NRU test method and for 25 of the 71 (35%) substances evaluated for the NHK
318 NRU test method. For substances with log K_{ow} >3.00, 8/13 (62%) were outliers for both the
319 3T3 and NHK test methods. For molecular weights >400 g/mole, 4/7 (57%) substances were
320 outliers using the 3T3 NRU test method and 3/7 (43%) were outliers using the NHK NRU
321 test method. For substances with boiling points >200°C, 9/13 (69%) were outliers using the
322 3T3 NRU test method and 8/13 (62%) were outliers using the NHK NRU test method.

323 6.2.2.2 *Chemical Class*

324 Examination of outliers by chemical class for the RC millimole regression showed that all of
325 the chemical classes that contained at least three reference substances also contained at least
326 one outlier for one test method. Two classes contained 100% outliers for both test methods:
327 organophosphates (3/3) and organic sulfur compounds (5/5). The remaining classes with
328 higher frequencies of outliers included: 2/3 (67%) amines were outliers for both test methods,
329 7/14 (50%) heterocylics were outliers for the 3T3 NRU and 10/14 (71%) heterocyclics were
330 outliers for the NHK NRU, 2/5 (40%) chlorine compounds were outliers for both test
331 methods, 2/6 (33%) sodium compounds were outliers for both test methods, 3/10 (30%)
332 alcohols were outliers for the 3T3 NRU and 4/10 (40%) alcohols were outliers for the NHK
333 NRU, and 4/14 (29%) carboxylic acids were outliers for the 3T3 NRU and 6/14 (43%)
334 carboxylic acids were outliers for the NHK NRU.

335 6.2.2.3 *Solubility*

336 Another attribute that may cause a substance to be an outlier is the lack of solubility in the
337 test system. Because the SMT expected the toxicity of insoluble substances to be
338 underpredicted in the *in vitro* assays, substances that formed precipitates in the tests were
339 noted and compared with the outlier substances. However, insolubility was not consistently
340 associated with the outlier substances for which toxicity was underpredicted. For example,
341 eight of the 22 (36%) underpredicted substances identified by applying the 3T3 results to the
342 RC millimole regression exhibited signs of insolubility in at least one laboratory. NHK
343 results showed that seven of 23 (30%) underpredicted substances exhibited signs of
344 insolubility in at least one laboratory (see **Table 5-11** for substances that had precipitates in
345 the assays). Additionally, there was evidence of insolubility in the 3T3 and NHK NRU test
346 methods of dibutyl phthalate and diethyl phthalate, but toxicity was overpredicted for both
347 substances, rather than underpredicted. This overprediction may be a characteristic of the
348 phthalates, but more substances would have to be tested before a general rule could be
349 adopted.
350
351 There were 25 substances that showed evidence of insolubility in the 3T3 test method in at
352 least one laboratory, and 11 (44%) of these were outliers. Of the 24 substances showed
353 evidence of insolubility in at least one NHK laboratory, 11 (46%) were outliers.

354 6.2.2.4 *Metabolism*

355 It was anticipated that the toxicity of substances metabolized *in vivo* to active compounds
356 (see **Section 3.3.4.3** and **Table 3-7**) would be underpredicted *in vitro* by 3T3 and NHK cells,
357 which have little or no metabolic capability (Babich 1991; INVITTOX 1991). Of the 72
358 reference substances, 19 (26%) are known to have active metabolites *in vivo,* and 10 (45%)
359 of these were classified as outliers for 3T3. Of these 10 substances, which accounted for 36%
360 of the 28 outlier substances, the toxicity of six (60%) was underpredicted, while the toxicity
361 of four (40%) was overpredicted. Among the 31 outliers in the NHK NRU test method, nine
362 (29%) are metabolized to active metabolites. Nine of the 19 substances known to produce
363 active metabolites *in vivo* were discordant for the NHK NRU test method. NHK cells
364 underpredicted the toxicity of five (56%) of these nine substances and overpredicted the
365 other four (44%). These nine outlier substances accounted for 29% of the 31 outliers in the
366 NHK NRU test method. Thus, the fact that a substance has active metabolites that are not
367 expected to be produced in the *in vitro* tests does not necessarily indicate that its toxicity will
368 be underpredicted by *in vitro* basal cytotoxicity test methods.
369
370 Similarly, Halle (1998, 2003) noted that the RC substances that required metabolic activation
371 to produce *in vivo* toxicity were not necessarily outliers with respect to their fit to the RC
372 millimole regression. They found that eight (50%) of the 16 substances that required
373 metabolic activation to product toxicity were outliers (see **Table L3-3** in **Appendix L3**).

374 6.2.2.5 *Mechanism of Toxicity*

375 Substances whose mechanisms of toxicity would not be detected in the 3T3 or NHK cells
376 would be expected to fit the RC millimole regression poorly. In particular, toxic mechanisms
377 that include, for example, specific actions on the central nervous system (CNS) or the heart
378 are not expected to be active in the 3T3 or NHK cells. Neurotoxic mechanisms would
379 include, for example, cholinesterase inhibition, CNS nicotinic receptor blockade or
380 activation, or any activity other than membrane destabilization such as that produced by a
381 solvent, or disturbance of energy utilization such as interruption of oxidative
382 phosphorylation. Representative cardiotoxic mechanisms would include calcium channel
383 blockage and beta-adrenergic receptor activation or blockage.
384

385 The 72 reference substances used to validate the 3T3 and NHK NRU test methods included
386 16 (22%) that had specific CNS toxicity (see **Table 6-4**). Of these 16 substances, 10 (63%)
387 were outliers in both *in vitro* NRU test methods. Three of the six (50%) reference substances
388 that are cardiotoxic were outliers in the 3T3 NRU test method and two (33%) were outliers in
389 the NHK NRU test method. When all the reference substances with mechanisms that are not
390 expected to be active in the 3T3 and NHK cells (i.e., in **Table 6-4**) are summed, 13/22 (59%)
391 are outliers for the 3T3 NRU and 12/22 (55%) are outliers for the NHK NRU. These
392 substances represented 13/28 (46%) and 12/31 (39%) of the total outlier substances for the
393 3T3 and NHK NRU test methods, respectively. Halle (1998, 2003) reported similar findings
394 for the RC database (i.e., approximately half of the substances expected to be outliers based
395 on their mechanisms of toxicity were outliers) (see **Appendix L3**).
396

397

Table 6-4 Substances With Mechanisms of Toxicity Not Expected to Be Active in the 3T3 or NHK Cells in Culture

Substance	Mechanism of Toxicity[1]	3T3 Outlier[2]	NHK Outlier[2]
Neurotoxic			
Atropine sulfate	Antimuscarinic; anticholinergic action; competitive antagonism of anticholinesterase at cardiac and CNS receptor sites.	No	No
Caffeine	Inhibition of phosphodiesterase leading to AMP accumulation; translocation of intracellular Ca^{++}; adenosine receptor antagonism; neurotoxic.	No	Yes
Carbamazepine	Therapeutically decreases firing of noradrenergic neurons.	No	No
Chloral hydrate	Potentiation of $GABA_A$ receptor activity; inhibition of N-methyl-D-aspartate activity; modulation of 5-hydroxytryptamine$_3$ receptor-mediated depolarization of the vagas nerve[3].	No	No
Dichlorvos	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Disulfoton	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Endosulfan	Affects brain neurotransmitter levels[4].	Yes	Yes
Fenpropathrin	Delays closure of sodium channel causing persistent depolarization of membrane.	Yes	Yes
Glutethimide	CNS depression; anticholinergic activity.	No	No
Haloperidol	Blocks dopamine receptors.	No	No
Lindane	CNS depression through inhibition of GABA receptor linked chloride channel at the picrotoxin binding site, leading to blockade of chloride influx into neurons.	Yes	No
Nicotine	Cholinergic block causing polarization of CNS and PNS synapses.	Yes	Yes
Parathion	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Phenobarbital	CNS depression through inhibition of GABA synapses; inhibits hepatic NADH cytochrome oxidoreductase.	Yes	Yes
Physostigmine	Inhibition of acetylcholinesterase resulting in acetylcholine accumulation in CNS and effector organs.	Yes	Yes
Strychnine	Increases glutamic acid in the CNS.	Yes	Yes
Cardiotoxic			
Amitriptyline HCl	Blocks norepinephrine, 5-hydroxytryptamine, and dopamine presynaptic uptake; prevents reuptake of heart norepinephrine.	No	No

Table 6-4 Substances With Mechanisms of Toxicity Not Expected to Be Active in the 3T3 or NHK Cells in Culture

Substance	Mechanism of Toxicity[1]	3T3 Outlier[2]	NHK Outlier[2]
Digoxin	Impairs ion transport and increases sarcoplasmic calcium by binding to Na^+/K^+ ATPase, increasing automaticity of cardiac cells.	Yes	No
Epinephrine bitartrate	Adrenergic receptor stimulation.	Yes	Yes
Potassium chloride	Disturbs cardiac membrane potential and electrical activity.	No	No
Procainamide HCl	Slows impulse conduction in the heart[5].	No	No
Verapamil HCl	Inhibition of transmembrane Ca^{++} flux in excitatory tissues; alpha-adrenergic blockade.	Yes	Yes

Abbreviations: NA=Not available or information not found; CNS=Central nervous system; GABA=Gamma aminobutyric acid; PNS=Peripheral nervous system; NADH=Nicotine adenine dinucleotide (reduced).
[1]From Ekwall et al. (1998) or Hazardous Substances Data Bank (NLM 2001, 2002) unless otherwise noted.
[2]As shown in **Table 6-3**.
[3]EPA (2000b).
[4]ATSDR (2000a).
[5]Hardman et al. (1996).

6.3 Improving the Prediction of *In Vivo* Rat Oral LD_{50} Values from *In Vitro* IC_{50} Data

Because the 3T3 and NHK IC_{50} – rat oral LD_{50} regressions were not significantly different from the RC regression for the same substances, the next step was an attempt to improve the RC millimole regression for the prediction of LD_{50} values from IC_{50} values. Because the validation study provided results similar to the RC, and because the RC database has more than 3.5 times the number of substances tested in the validation study, the RC rat data (282 substances) are used to determine the relationship between IC_{50} and LD_{50}. The RC data were used to develop two new regressions, the RC rat-only millimole regression and the RC rat-only weight regression. For reference, the original RC millimole regression, log LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625 (Halle 1998, 2003), is shown in **Table 6-5** and **Figure 6-4**.

6.3.1 The RC Rat-Only Millimole Regression

The first regression used the RC data for the 282 substances with rat LD_{50} data and the original units of mM for IC_{50} and mmol/kg for LD_{50} (see **Table 6-3** and **Figure 6-4**). Only rat data were used because:

- Rats and mice are not always equally sensitive to all substances
- The majority of acute oral LD_{50} data used in the RC millimole regression were from studies using rats (282 rat data points versus 65 mouse data points) (Halle 1998, 2003)
- Most acute oral toxicity testing is performed with rats.

The RC rat-only millimole regression is applicable to substances of known molecular weight that are relatively pure.

Table 6-5 Linear Regression Analyses to Improve the Prediction of Rodent Acute Oral LD_{50} Values from *In Vitro* NRU IC_{50} Using the RC Database[1]

Data Used	Slope	Intercept	R^2
347 RC substances (282 rat and 65 mouse LD_{50} values) – millimole units[2]	0.435	0.625	0.452[3]
282 RC substances with rat LD_{50} data – millimole units[2]	0.439	0.621	0.452
282 RC substances with rat LD_{50} data – weight units[4]	0.372	2.024	0.325

Abbreviations: NRU=Neutral red uptake; RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
[1]Slopes of all regressions were significantly different ($p < 0.05$) from zero at $p < 0.0001$.
[2]IC_{50} in mM; LD_{50} in mmol/kg.
[3]Calculated from RC data (i.e., not reported by Halle [1998, 2003]).
[4]IC_{50} in μg/mL; LD_{50} in mg/kg.

Table 6-5 shows that the RC millimole regression using only rat acute oral LD_{50} data was essentially identical to the original regression that used both rat and mouse data. The slope changed from 0.435 to 0.439 and the intercept changed from 0.625 to 0.621; these changes were not statistically significantly different.

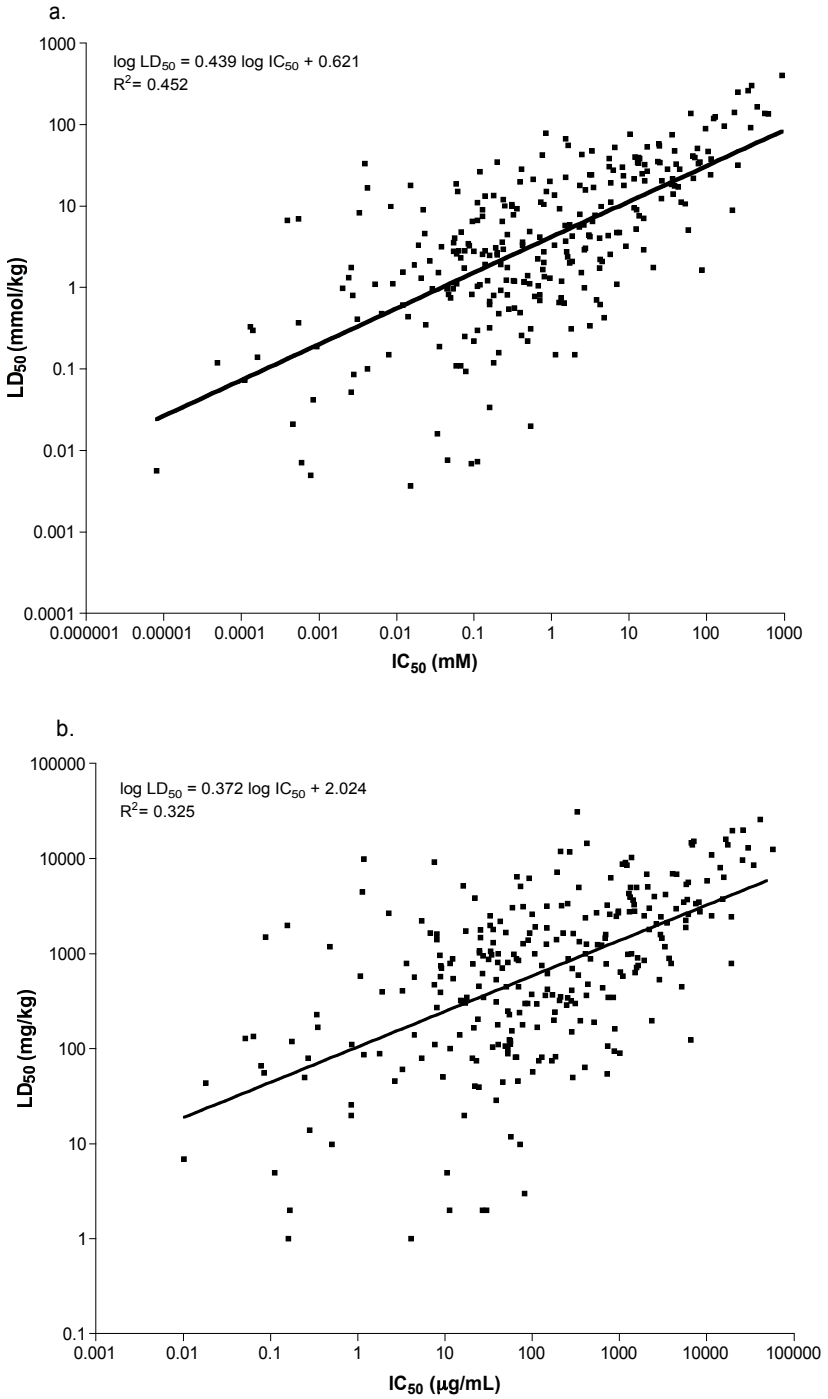
6.3.2 The RC Rat-Only Weight Regression

The second regression used the same RC rat acute oral LD_{50} data for the 282 substances but was calculated using weight units rather than millimolar units (see **Table 6-5** and **Figure 6-5**). Weight units (i.e., mg/kg for the LD_{50} and μg/mL for the IC_{50}) were selected for the units of measurement because

- Millimole units are not applicable to mixtures and substances with unknown structures or molecular weights.
- They are the most practical, i.e., hazard classification in all regulatory systems is based on LD_{50} values expressed in mg/kg (see **Table 1-2**).

The RC rat-only weight regression is applicable for use with complex mixtures, substances whose structures or molecular weights are unknown, and substances that are relatively impure (i.e., mixtures that are primarily composed of a named substance).

Figure 6-4 RC Rat-Only Millimole Regression (a) and RC Rat-Only Weight Regression (b)



Abbreviations: RC=Registry of Cytotoxicity; R^2=Coefficient of determination.
Regressions and IC_{50} and rat oral LD_{50} datapoints for 282 substances from the RC (see **Table 6-5**).

462 **6.4 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting GHS Acute**
463 **Oral Toxicity Categories**

464 Based on the correlations/regressions obtained between the 3T3 and NHK NRU IC_{50} values
465 and the rat LD_{50} values, it is clear that these *in vitro* methods are not suitable as replacements
466 for rodent acute oral toxicity tests. The use of *in vitro* methods to reduce animal use for
467 rodent acute oral toxicity assays (i.e., to assist in determining the starting doses for *in vivo*
468 assays) also depends upon their accuracy for the prediction of LD_{50} values. However, this
469 latter (adjunct) use does not require the same precision in LD_{50} prediction as complete
470 replacement would.
471
472 The NRU-predicted LD_{50} values were determined using the *in vitro* NRU IC_{50} values in the
473 regressions presented in **Table 6-5.** The predicted LD_{50} values were used to assign each
474 substance to a predicted GHS acute oral toxicity category (UN 2005). The accuracy of the
475 3T3 and NHK NRU test methods for predicting GHS acute oral toxicity categories was
476 determined by comparison with categorization based on rat acute oral LD_{50} data. The
477 rationale for evaluating the accuracy of LD_{50} predictions was that the animal savings
478 produced by using these *in vitro* NRU test methods to predict starting doses for rodent acute
479 oral toxicity assays would be greatest when the starting dose is as close as possible to the
480 LD_{50}. This approach was used because regulatory authorities use rodent acute oral toxicity
481 test results for hazard classification and labelling of products to protect handlers and
482 consumers.
483
484 The *in vitro* NRU test methods were evaluated for their ability to predict GHS acute oral
485 toxicity categories using the two regressions presented in **Section 6.3**, the RC rat-only
486 millimole regression and the RC rat-only weight regression. The same reference substances
487 were evaluated for each regression. Sixty-seven and 68 substances were evaluated using the
488 3T3 and NHK NRU test methods, respectively. Of the original 72 reference substances
489 tested, epinephrine bitartrate, colchicine, and propylparaben were excluded because they had
490 no rat acute oral LD_{50} reference data (see **Table 4-2**). Carbon tetrachloride and methanol
491 were excluded from the 3T3 evaluations because no laboratory attained sufficient toxicity in
492 any test for the calculation of an IC_{50} (see **Table 5-4**). Carbon tetrachloride was excluded

493 from the NHK evaluations because no laboratory attained sufficient toxicity in any test for
494 the calculation of an IC_{50}.(see **Table 5-5**).
495
496 For comparison with the NRU test method results and RC rat-only regressions, **Section 6.4.1**
497 and provides the accuracy analysis for the RC database used with the RC millimole
498 regression. **Sections 6.4.2** and **6.4.3** provide the accuracy information for the 3T3 and NHK
499 NRU test methods for the RC rat-only millimole regression and RC-rat only weight
500 regression, respectively. A summary of predictivity[2] is provided for each predicted toxicity
501 category, along with the percentage of substances whose toxicity was underpredicted or
502 overpredicted.

503 6.4.1 Prediction of GHS Acute Oral Toxicity Category by the RC IC_{50} Values Using the
504 RC Millimole Regression

505 **Table 6-6** shows the concordance of the observed (i.e., *in vivo*) and predicted GHS acute oral
506 toxicity categories (UN 2005) for the 347 RC IC_{50} values in the RC millimole regression, log
507 LD_{50} (mmol/kg) = 0.435 x log IC_{50} (mM) + 0.625 (Halle 1998, 2003). Accuracy is the
508 agreement of the *in vitro* category predictions with those based on the 347 rodent (282 rat
509 and 65 mouse) oral LD_{50} values used in the RC database (Halle 1998, 2003). Substances for
510 which the *in vitro* toxicity category prediction did not match the *in vivo* category were
511 considered discordant for the GHS acute oral toxicity category predictions.
512
513 The overall accuracy of the RC IC_{50} values for correctly predicting GHS acute oral toxicity
514 classification category using the RC millimole regression was 40% (140/347substances)
515 (**Table 6-6**). Rodent acute oral toxicity was overpredicted for 34% (118/347) and
516 underpredicted for 26% (89/347) of the substances. For this analysis, with respect to the
517 predictions of each GHS category:

518 • None (0%) of the 12 substances with LD_{50} <5 mg/kg (GHS Category I) was
519 correctly predicted.

[2] Proportion of correct *in vivo* category matches for all substances with *in vitro* predictions for a particular category. Predictivity is one of the measures of test accuracy (ICCVAM 2003).

- Four (15%) of 26 substances in the $5 < LD_{50} \leq 50$ mg/kg category (GHS Category II) were correctly predicted.
- Twenty (29%) of 69 substances in the $50 < LD_{50} \leq 300$ mg/kg category (GHS Category III) were correctly predicted.
- Ninety-seven (69%) of 140 substances in the $300 < LD_{50} \leq 2000$ mg/kg category (GHS Category IV) were correctly predicted. This toxicity category was also predicted for 106 other substances (52%; 106/203) that did not fall in this category. Thus, the overall predictivity for this category was 48% (97/203 substances predicted for this category matched the *in vivo* category).
- Fourteen (25%) of the 56 substances in the $2000 < LD_{50} \leq 5000$ mg/kg category (GHS Category V) were correctly predicted.
- Five (11%) of the 44 substances with $LD_{50} > 5000$ mg/kg (GHS Unclassified) were correctly predicted.

The highest accuracy, 69%, for the RC IC_{50} values in the RC millimole regression were obtained for substances in the $300 < LD_{50} \leq 2000$ mg/kg category (GHS Category IV). The lowest accuracy, 0%, was obtained for substances with $LD_{50} < 5$ mg/kg (GHS Category I). Although the 11% accuracy was low for substances with $LD_{50} > 5000$ mg/kg (GHS Unclassified), the highest predictivity, 83%, was obtained for substances in this group. The RC millimole regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-6**).

Rodent acute oral toxicity was overpredicted for 34% (118) and underpredicted for 26% (89) of the 347 RC substances. Thus, there was a total of were 207 discordant substances. GHS category was overpredicted for 57% (118/207) of the discordant substances and underpredicted for 43% (89/207) of the discordant substances.

.

548 **Table 6-6 Prediction of GHS Acute Oral Toxicity Category by the RC IC_{50} Values and the RC**
549 **Millimole Regression[1]**
550

***In Vivo* Rodent Oral LD_{50}[2] (mg/kg)**	**IC_{50}-Predicted GHS Category (mg/kg)[3]**						**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
	LD_{50} <5	**5< LD_{50} ≤50**	**50< LD_{50} ≤300**	**300< LD_{50} ≤2000**	**2000< LD_{50} ≤5000**	**LD_{50} >5000**				
LD_{50} < 5	0	5	3	4	0	0	12	0%	0%	100%
5< LD_{50} ≤50	0	4	13	9	0	0	26	15%	0%	85%
50< LD_{50} ≤300	0	9	20	38	2	0	69	29%	13%	58%
300< LD_{50} ≤2000	0	4	24	97	14	1	140	69%	20%	11%
2000< LD_{50} ≤5000	0	1	5	36	14	0	56	25%	75%	0%
LD_{50} >5000	0	0	1	19	19	5	44	11%	89%	0%
Total	0	23	66	203	49	6	347	40%	34%	26%
Predictivity	0%	17%	30%	48%	29%	83%				
Category Overpredicted	0%	22%	24%	25%	33%	17%				
Category Underpredicted	0%	61%	45%	27%	39%	0%				

551 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); RC=Registry of Cytotoxicity. Shaded cells are those
552 containing the correct predictions; RTECS®=Registry of Toxic Effects for Chemical Substances®.
553 [1]The RC millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.435 + 0.625. Numbers in table represent numbers of substances.
554 [2]Rat (282 values) and mouse (65 values) oral LD_{50} values, mostly from the 1983/84 RTECS®, that were converted to mmol/kg for used in the RC (Halle 1998,
555 2003).
556 [3]IC_{50} values from the RC are geometric mean IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity
557 endpoints (Halle 1998, 2003). GHS categories were predicted by using the IC_{50} values to calculate predicted LD_{50} values with the RC
558 millimole regression equation. Predicted LD_{50} values in mmol/kg for each substance were converted to mg/kg and used to classify the
559 substance in the appropriate predicted GHS acute oral toxicity category.
560

561

562 6.4.2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test
563 Methods Using the RC Rat-Only Millimole Regression

564 **Table 6-7** shows the concordance of the observed (i.e., *in vivo*) and predicted GHS acute oral
565 toxicity categories (UN 2005) for each *in vitro* test method using the geometric mean IC_{50}
566 values (of the three laboratories) in the RC rat-only millimole regression, log LD_{50}
567 (mmol/kg) = 0.439 x log IC_{50} (mM) + 0.621. Accuracy is the agreement of the *in vitro*
568 category predictions with those based on the rat acute oral LD_{50} reference values in **Table 4-**
569 **2**. Substances for which the *in vitro* toxicity category prediction did not match the *in vivo*
570 category were considered discordant for the GHS acute oral toxicity category predictions.

571 6.4.2.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Millimole Regression*

572 The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral
573 toxicity classification category using the RC rat-only millimole regression was 31% (21/67
574 substances) (**Table 6-7**). Rat acute oral toxicity was overpredicted for 34% (23) and
575 underpredicted for 34% (23) of the substances. For this analysis, with respect to the
576 predictions of each GHS category:

577 • None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) was
578 correctly predicted.
579 • One (9%) of 11 substances in the 5 < LD_{50} ≤50 mg/kg category (GHS
580 Category II) was correctly predicted.
581 • Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg category (GHS
582 Category III) were correctly predicted.
583 • Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg category
584 (GHS Category IV) were correctly predicted. This toxicity category was also
585 predicted for 32 other substances (71%; 32/45) that did not fall in this
586 category. Thus, the overall predictivity for this category was 29% (13/45
587 substances predicted for this category matched the *in vivo* category).
588 • None (0%) of the 10 substances in the 2000 < LD_{50} ≤5000 mg/kg category
589 (GHS Category V) were correctly predicted.
590 • Two (17%) of the 12 substances with LD_{50} >5000 mg/kg (GHS Unclassified)
591 were correctly predicted.
592

Table 6-7 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression[1]

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 -Predicted GHS Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5< LD_{50} ≤50	50< LD_{50} ≤300	300< LD_{50} ≤2000	2000< LD_{50} ≤5000	LD_{50} >5000				
LD_{50} < 5	0	2	0	4	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	1	6	3	1	0	11[4]	9%	0%	91%
50< LD_{50} ≤300	0	0	5	7	0	0	12	42%	0%	58%
300< LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000< LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	8	2	2	12[6,7]	17%	83%	0%
Total	0	4	13	45	3	2	67	31%	34%	34%
Predictivity	0%	25%	38%	29%	0%	100%				
Category Overpredicted	0%	50%	46%	31%	33%	0%				
Category Underpredicted	0%	25%	15%	40%	67%	0%				
Reference Rat Oral LD_{50}[2]	**NHK -Predicted Toxicity Category (mg/kg)**						**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
	LD_{50} <5	5< LD_{50} ≤50	50< LD_{50} ≤300	300< LD_{50} ≤2000	2000< LD_{50} ≤5000	LD_{50} >5000				
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	2	5	3	1	0	11[4]	18%	0%	82%
50< LD_{50} ≤300	0	1	6	5	0	0	12	50%	8%	42%
300< LD_{50} ≤2000	0	1	2	12	1	0	16	75%	19%	6%
2000< LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	7	6	0	13[7]	0%	100%	0%
Total	0	5	15	40	8	0	68	29%	40%	31%
Predictivity	0%	40%	40%	30%	0%	0%				
Category Overpredicted	0%	20%	47%	28%	25%	0%				
Category Underpredicted	0%	40%	13%	43%	75%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; RC=Registry of Cytotoxicity. Shaded cells are those containing the correct predictions.

[1]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.

[2]Reference rat oral LD_{50} values in mg/kg from **Table 4-2**.

[3]Epinephrine bitartrate excluded because no rat reference acute oral LD_{50} was identified (see **Table 4-2**).

[4]Colchine excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[7]Propylparaben excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

The overall accuracy of the NHK NRU test method for correctly predicting the GHS acute oral toxicity classification, when the prediction was based on the RC rat-only millimole regression, was 29% (20/68 substances) (see **Table 6-5**). Toxicity was overpredicted for 40% (27) and underpredicted for 31% (21) of the 68 substances. The pattern of concordance between *in vitro* and *in vivo* results for the NHK NRU test method with the RC rat-only millimole regression was similar to that for the 3T3 NRU test method with the exception that none of the substances with a toxicity of LD_{50} >5000 mg/kg were correctly predicted. For this analysis, with respect to the predictions of each GHS category:

- None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were correctly predicted.
- Two (18%) of 11 substances in the 5< LD_{50} ≤50 mg/kg category (GHS Category II) were correctly predicted.
- Six (50%) of 12 substances in the 50< LD_{50} ≤300 mg/kg category (GHS Category III) were correctly predicted.
- 12 (75%) of 16 substances in the 300< LD_{50} ≤2000 mg/kg category (GHS Category IV) were correctly predicted; however, this category was also predicted for 28 (70%; 28/40) substances that did not match the category. Thus, the overall predictivity for this category was 30% (12/40).
- None (0%) of the 10 substances in the 2000< LD_{50} ≤5000 mg/kg category (GHS Category V) were correctly predicted.
- None (0%) of the 13 substances with LD_{50} >5000 mg/kg (GHS Unclassified) were correctly predicted.

The RC rat-only millimole regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted toxicity for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-7**). Although substances at the very low and high ends of the toxicity range were poorly predicted, those in the middle range (i.e., 300 < LD_{50} ≤2000 mg/kg) were predicted much better, with 75 to 81% accuracy. The pattern of accuracy for the GHS categories was similar to the pattern seen with the RC IC_{50} and LD_{50} values and the RC millimole regression (see **Table 6-6**) (i.e., lowest accuracy for

636 very toxic and very nontoxic substances and highest accuracy for substances with 300 < LD_{50}
637 ≤2000 mg/kg).

638 6.4.2.2 *Discordant Substances in the Prediction of GHS Acute Oral Toxicity Category by*
639 *the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression*

640 **Appendix L2** identifies the discordant substances, that is, those for which the *in vitro*
641 predicted GHS acute oral toxicity category did not match the GHS acute oral toxicity
642 category assigned based on the reference rat acute oral LD_{50} data in **Table 4-2**. Of the total
643 number of substances used for this evaluation (67 for 3T3, 68 for NHK), the 3T3 test method
644 underpredicted the GHS category for 23 (50%) and overpredicted for 23 (50%) of the 46
645 discordant substances. The NHK test method underpredicted toxicity for 21 (44%) and
646 overpredicted for 27 (56%) of the 48 discordant substances.

647 6.4.3 <u>Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test</u>
648 <u>Methods Using the RC Rat-Only Weight Regression</u>

649 **Table 6-8** shows the concordances of the observed and predicted GHS acute oral toxicity
650 categories for each *in vitro* NRU method using the geometric mean IC_{50} values from the
651 three laboratories and the RC rat-only weight regression (**Table 6-5)**. The regression formula
652 for the RC rat-only weight regression was log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 +
653 2.024. Accuracy is the agreement of the GHS acute oral toxicity category predictions made
654 using the *in vitro* NRU data with those based on the reference rat acute oral LD_{50} values
655 (**Table 4-2)**.

656 6.4.3.1 *In Vitro – In Vivo Concordance Using the RC Rat-Only Weight Regression*

657 The overall accuracy of the 3T3 NRU test method with the RC rat-only weight regression
658 was 31% (21/67) (**Table 6-8**). The toxicity was overpredicted for 33% (24) and
659 underpredicted for 36% (22) of the substances. For this analysis, with respect to the
660 predictions of the GHS category:

661 • None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were
662 correctly predicted.
663 • One (9%) of 11 substances in the 5< LD_{50} ≤50 mg/kg category (GHS
664 Category II) was correctly predicted.

665 • Four (33%) of 12 substances in the 50< LD_{50} ≤300 mg/kg category (GHS
666 Category II) were correctly predicted; however, because 10 other substances
667 were also predicted to be in this category, the overall predictivity was 29%
668 (4/14).
669

670 **Table 6-8 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU**
671 **Test Methods and the RC Rat-Only Weight Regression[1]**
672

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 -Predicted Toxicity Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5< LD_{50} ≤50	50< LD_{50} ≤300	300< LD_{50} ≤2000	2000< LD_{50} ≤5000	LD_{50} >5000				
LD_{50} <5	0	0	2	4	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50< LD_{50} ≤300	0	0	4	8	0	0	12	33%	0%	67%
300< LD_{50} ≤2000	0	1	3	12	0	0	16	75%	25%	0%
2000< LD_{50} ≤5000	0	0	0	6	4	0	10[5]	40%	60%	0%
LD_{50} >5000	0	0	0	5	7	0	12[6,7]	0%	100%	0%
Total	0	2	14	40	11	0	67	31%	33%	36%
Predictivity	0%	50%	29%	30%	36%	0%				
Category Overpredicted	0%	0%	50%	43%	0%	0%				
Category Underpredicted	0%	50%	21%	28%	64%	0%				
Reference Rat Oral LD_{50}[2] (mg/kg)	**NHK -Predicted Toxicity Category (mg/kg)**									
	LD_{50} <5	**5< LD_{50} ≤50**	**50< LD_{50} ≤300**	**300< LD_{50} ≤2000**	**2000< LD_{50} ≤5000**	**LD_{50} >5000**	**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50< LD_{50} ≤300	0	1	5	6	0	0	12	42%	8%	50%
300< LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000< LD_{50} ≤5000	0	0	0	9	1	0	10[5]	10%	90%	0%
LD_{50} >5000	0	0	0	6	6	1	13[7]	8%	92%	0%
Total	0	4	14	42	7	1	68	31%	37%	32%
Predictivity	0%	25%	36%	31%	14%	100%				
Category Overpredicted	0%	25%	50%	33%	0%	0%				
Category Underpredicted	0%	50%	14%	36%	86%	0%				

673 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes;
674 NRU=Neutral red uptake; RC=Registry of Cytotoxicity.
675 [1]The RC rat-only weight regression is log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 + 2.024.
676 [2]Reference rat oral LD_{50} values in mg/kg from **Table 4-2**.
677 [3]Epinephrine bitartrate excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).
678 [4]Colchine excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).
679 [5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
680 [6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.
681 [7]Propylparaben excluded because no rat acute oral LD_{50} was identified (see **Table 4-2**).

682 • Twelve (75%) of 16 substances in the 300< LD_{50} ≤2000 mg/kg category
683 (GHS Category IV) were predicted correctly. Because a total of 40 substances
684 were predicted to be in this category, the overall predictivity was 30% (12/40).
685 • Four (40%) of 10 substances in the 2000< LD_{50} ≤5000 mg/kg category (GHS
686 Category V) were correctly predicted; however, because a total of 11
687 substances were predicted to be in this category, the overall predictivity was
688 36% (4/11).
689 • None (0%) of the 12 substances with LD_{50} >5000 mg/kg (GHS Unclassified)
690 were correctly predicted.
691
692 The overall accuracy of the NHK predictions using the RC rat-only weight regression was
693 31% (21/68) (see **Table 6-8**). The *in vivo* GHS toxicity categories were overpredicted for
694 37% (22) and underpredicted for 32% (25) of the substances. For this analysis, with respect
695 to the predictions of the GHS category:
696 • None (0%) of the six substances with LD_{50} <5 mg/kg (GHS Category I) were
697 correctly predicted.
698 • One (9%) of 11 substances in the 5 < LD_{50} ≤ 50 mg/kg category (GHS
699 Category II) was correctly predicted.
700 • Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg category (GHS
701 Category III) were correctly predicted; however, because six other substances
702 were also predicted to be in this category, the overall predictivity was 33%
703 (3/9).
704 • Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg category
705 (GHS Category IV) were predicted correctly; however, because 29 other
706 substances were also predicted to be in this category, the overall predictivity
707 was 31% (13/42).
708 • One (10%) of 10 substances in the 2000 < LD_{50} ≤5000 mg/kg category (GHS
709 Category V) was correctly predicted.
710 • One (8%) of 13 substances with LD_{50} >5000 mg/kg (GHS Unclassified) was
711 correctly predicted.

712 The RC rat-only weight regression generally underpredicted toxicity for substances in the
713 highest toxicity (i.e., lowest LD_{50}) categories and overpredicted toxicity for substances in the
714 lowest toxicity (i.e., highest LD_{50}) categories (see **Table 6-8**). Although substances at the
715 very low and high ends of the toxicity range were poorly predicted, those in the middle range
716 (i.e., $300 < LD_{50} \leq 2000$ mg/kg) were predicted much better, with 75 to 81% accuracy. The
717 pattern of accuracy for the GHS categories was similar to the pattern seen with the RC IC_{50}
718 and LD_{50} values and the RC millimole regression (see **Table 6-6**) and with the NRU IC_{50} and
719 rat oral LD_{50} values and the RC rat-only millimole regression (see **Table 6-7**) (i.e., lowest
720 accuracy for very toxic and very nontoxic substances and highest accuracy for substances
721 with $300 < LD_{50} \leq 2000$ mg/kg).

722 6.4.3.2 *Discordant Substances in the Prediction of GHS Acute Oral Toxicity Category by*
723 *the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression*

724 **Appendix L2** shows the substances for which the *in vitro* predicted GHS acute oral toxicity
725 category using the RC rat-only weight regression did not match those that were based on the
726 rat acute oral LD_{50} reference data. The two *in vitro NRU test* methods over- and under-
727 predicted the GHS acute oral toxicity category for similar numbers of substances, compared
728 with the GHS acute oral toxicity categories for the rat acute oral LD_{50} reference values in
729 **Table 4-2**. The 3T3 NRU test method overpredicted the GHS acute oral toxicity category for
730 22 (48%) of 46 discordant substances, and underpredicted of 24 (52%) substances. The NHK
731 NRU test method overpredicted the GHS acute oral toxicity category for 25 (53%) of 47
732 discordant substances, and underpredicted 22 (47%) substances.

733 6.4.4 Summary of the Regressions Evaluated

734 **Table 6-9** summarizes the regressions evaluated in **Section 6.4** for accuracy in predicting the
735 GHS acute oral toxicity categories (UN 2005), and the proportion of over- or under-
736 predictions. Prediction accuracy using the RC IC_{50} and LD_{50} values and the RC millimole
737 regression was higher that that for the NRU test methods with the RC rat-only regressions
738 (i.e., 40% for the RC vs. 29% to 31% for the NRU test methods). Prediction accuracy was
739 slightly higher for the 3T3 NRU test method compared with the NHK NRU (i.e., 31% for
740 3T3 vs. 29% for NHK) using the RC rat-only millimole regression, and the same as the NHK
741 NRU test method (i.e., 31%) using the RC rat-only weight regression. The proportion of
742 discordant substances using the RC IC_{50} values and the RC millimole regression (60%) was

743 lower than that using the *in vitro* NRU test methods and the RC rat-only regressions (69% to
744 71%). The proportion of discordant substances from the 3T3 test method, 69%, was the same
745 whether it was determined with the RC rat-only millimole regression or the RC rat-only
746 weight regression. The proportion of discordant substances for the NHK test method was
747 slightly lower with RC rat-only weight regression than with the RC rat-only millimole
748 regression (69% vs. 71%). The RC IC_{50} values and the RC millmole regression were
749 expected to perform better than the *in vitro* NRU methods and the RC rat-only regressions
750 since the IC_{50} and LD_{50} values used to evaluate the performance of the RC millimole
751 regression were exactly the same as those used to calculate the linear regression formula. The
752 NRU IC_{50} values and the reference oral LD_{50} values used to evaluate the RC rat-only
753 regressions were different from those used to calculate the RC rat-only regressions.
754
755 **Table 6-9 Comparison of Regressions and *In Vitro* NRU Test Methods for Their**
756 **Performance in Predicting GHS Acute Oral Toxicity Categories**
757

Regression	N[1]	R^2 Statistic	Accuracy	Discordant Substances[2]
RC millimole[3]	347	0.452	RC IC_{50} – 40%	RC IC_{50} – 207/347 (60%)
RC rat-only millimole[3]	282	0.452	3T3– 31% NHK– 29%	3T3– 46/67(69%) NHK– 48/68 (71%)
RC rat-only weight[3]	282	0.325	3T3– 31% NHK– 31%	3T3– 46/67 (69%) NHK– 47/68 (69%)

758 Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005);
759 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; RC=Registry of
760 Cytotoxicity; R^2=Coefficient of determination.
761 [1]Number of substances used in regression.
762 [2]Proportion of discordant substances.
763 [3]From **Table 6-5**.
764

765 The accuracy of the GHS category predictions using the *in vitro* NRU test methods with the
766 RC rat-only regressions obtained for the reference substances from this validations study may
767 or may not be applicable to other substances. A number of reasons may explain the low
768 accuracy for the reference substances. One is the skewness of the substances selected for
769 testing with respect to fit to the RC millimole regression (see **Figure 3-1**). **Table 3-4** shows
770 that 22 (38%) of the 58 RC substances selected for testing were known to poorly fit the RC
771 millimole regression (i.e., the predicted LD_{50} was outside the RC acceptance interval).

Toxicity[3] was underpredicted for 17 (77%) of these outlier substances and overpredicted (i.e., predicted LD_{50} was lower than measured *in vivo* LD_{50}) for the remaining five (23%). **Table 6-3** shows that 43% (30/70 for 3T3) and 44% (31/71 for NHK) of the reference substances that yielded IC_{50} values were outliers. Other reasons for the low accuracy for GHS acute oral toxicity prediction, such as those discussed in **Section 1.2.3**, include the major differences between cell cultures and whole animals regarding the absorption, distribution (including binding to serum proteins), availability, metabolism, and excretion of reference substances.

6.5 Correlation of NRU Concentration-Response Slope with Rat Lethality Dose-Response Slope

Because the slope calculations available for the NRU concentration-response curve analyses were based on the Hill function, the SMT determined whether the Hill Slope correlated with the rodent dose-mortality slope. If the two were correlated, the Hill Slope from the NRU test methods could be used to estimate the dose-mortality slope, which could, in turn, be used to estimate the most appropriate dose progression for UDP testing in rodents. A more immediate use for the validation study results, however, would be for the computer simulation modeling of animal testing for the UDP and ATC acute oral toxicity methods (described in **Sections 10.2** and **10.3**).

Dose-mortality slope information was available for 22 of the 72 reference substances, as shown in **Table 6-10**. Hill function slopes were available for 20 and 21 of the 22 substances for the 3T3 and the NHK NRU test methods, respectively. The Hill function slopes were transformed to absolute values because geometric means cannot be calculated for negative numbers, and geometric mean Hill function slopes were calculated for the acceptable NRU tests for each reference substance. When there was more than one dose-mortality slope available for a substance, a geometric mean was calculated from the available values. The absolute values of the geometric mean Hill function slopes are plotted against the geometric mean dose-mortality slopes in **Figure 6-4**. To determine whether there was a relationship between the absolute value of the Hill Slope and the dose-mortality slope, Spearman

[3] Toxicity is inversely proportional to LD_{50}. High LD_{50} values reflect low toxicity and low LD_{50} values reflect high toxicity

correlation analyses and least squares linear regression analyses were performed for each method. Both analyses showed that the absolute value of the *in vitro* Hill function slope was not related to the dose-mortality slope. The Spearman correlation analysis yielded nonsignificant correlations for both *in vitro* NRU test methods (3T3 r_s=-0.051 with p=0.831, and NHK r_s=-0.142 with p=0.541). Linear regression analyses for the prediction of dose-mortality slope by the absolute value of the Hill function slope also showed that the slopes of the regressions were not significantly different from zero (3T3 p=0.774, and NHK p=0.994). Because there was no relationship between Hill function slope and dose-mortality slope, the Hill function slope was not used to predict the dose-mortality slope for the simulation modeling of animal testing for the UDP and ATC acute oral toxicity methods in **Sections 10.2** and **10.3**.

Table 6-10 Reference Substances with Dose-Mortality and NRU Hill Slopes

Reference Substance	Dose-Mortality Slope[1]	3T3 Hill Slope[2]	NHK Hill Slope[2]
Acetylsalicylic acid	1.45	1.658	1.906
Boric acid	7.70	1.511	1.083
Caffeine	6.27	1.069	1.215
Carbon tetrachloride	2.06	NA	NA
Dichlorvos	1.24	2.240	1.383
Dimethylformamide	1.11	1.875	3.157
Diquat dibromide	16.57	4.273	1.289
Ethanol	4.57	1.725	2.049
Ethylene glycol	38.38	2.016	2.904
Glycerol	8.90	1.941	2.398
Hexachorophene	12.84	1.466	2.470
Lactic acid	4.04	4.541	2.934
Methanol	8.53	NA	1.173
Nicotine	3.00	11.019	0.682
Parathion	1.31	1.551	1.467
Potassium cyanide	14.50	1.931	1.207
Sodium arsenite	7.60	2.317	1.717
Sodium I fluoride	1.26	3.952	2.569
Trichloroacetic acid	20.97	1.883	1.369
Triethylene melamine	2.10	0.963	1.355
Valproic acid	1.20	2.467	1.440
Xylene	9.60	1.871	2.452
Carbon tetrachloride	2.06	NA	NA

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; NA=Not available.
[1]Geometric mean if there was more than one value for each substance (from **Appendix H2**).
[2]Geometric mean of absolute values from acceptable *in vitro* NRU tests.

820 **6.6 Strengths and Limitations of the Use of *In Vitro* Cytotoxicity Test Methods**
821 **with the IC_{50}-LD_{50} Regressions for Prediction of Rodent Acute Oral Toxicity**

822 6.6.1 *In Vitro* Cytotoxicity Methods

823 The NRU basal cytotoxicity methods tended to underpredict the toxicity of the most toxic
824 substances and to overpredict the toxicity of the least toxic substances for each regression
825 evaluated. The 3T3 and NHK NRU test methods were best at predicting the toxicity of
826 substances with 300 < LD_{50} ≤2000 mg/kg. The accuracy of the *in vitro* prediction of this
827 GHS category using the RC rat-only millimole regression and the RC rat-only weight
828 regression was 75-81%. GHS toxicity categories of substances with higher or lower LD_{50}
829 values were correctly predicted with less than 50% accuracy. The worst accuracy, 0%, was
830 observed for:

831 - Substances with LD_{50} ≤5 mg/kg in both *in vitro* test methods and regressions
832 - Substances with 2000< LD_{50} ≤5000 mg/kg using 3T3 with the RC rat-only
833 millimole regression
834 - Substances with 2000< LD_{50} ≤5000 mg/kg or LD_{50} >5000 mg/kg using NHK
835 with RC rat-only millimole regression
836 - Substances with LD_{50} >5000 mg/kg using 3T3 with RC rat-only weight
837 regression

838

839 Some substances with low toxicity and low solubility could not be tested in the *in vitro* NRU
840 test methods because the concentration of dissolved substance was inadequate to obtain an
841 IC_{50} value. None of the laboratories obtained adequate toxicity in any of the 3T3 tests of
842 carbon tetrachloride or methanol, and at least one laboratory failed to achieve adequate
843 toxicity with gibberellic acid or xylene. No laboratory achieved adequate toxicity in any of
844 the NHK experiments with carbon tetrachloride, and at least one laboratory could not achieve
845 adequate toxicity with methanol, 1,1,1-trichloroethane, or xylene. Another limitation of use
846 of the *in vitro* test methods is in the testing of substances that come out of solution by
847 forming a film on the medium surface or plastic well wall (i.e., "film out"), and for
848 substances that etch the laboratory ware plastics (ICCVAM 2006). Substances that etch
849 plastics can be detected by looking for the presence of etched rings in the 96-well plates after
850 exposure. Some substances that produce films in medium also etch plastic.

851

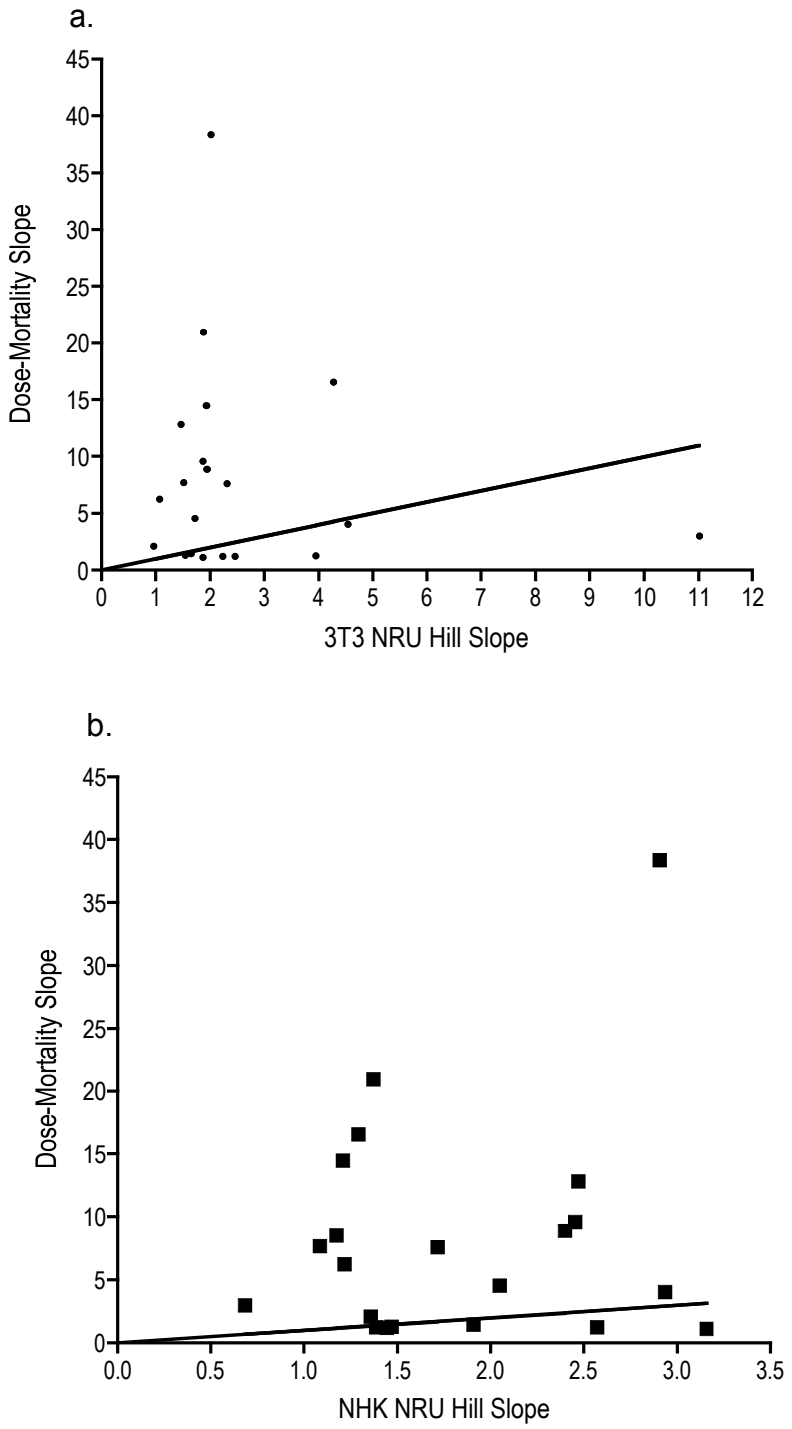
852 The prediction of rodent acute oral toxicity (and the starting doses for acute oral toxicity
853 tests) by the *in vitro* NRU methods is expected to be poor for substances with mechanisms of
854 toxicity that are not effective in the 3T3 and NHK cells. Such toxic mechanisms include
855 specific, receptor-mediated actions on the CNS or the heart.

856

857 The evaluation of the 3T3 and NHK NRU test methods for predicting starting doses for
858 rodent acute oral toxicity testing with its potential to reduce and refine animal use is provided
859 in **Section 10**.

860

Figure 6-5 Correlation of Dose-Mortality Slope to Hill Function Slope



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake.
Hill function slopes and dose-mortality slopes for the reference substances shown in **Table 6-8** for (a) the 3T3 data and (b) the NHK data. The solid line indicates the theoretical, one-to-one correspondence of Hill function slope with dose-mortality slope. Spearman's correlation coefficients were r_s=-0.051 (p=0.831) for the 3T3 and r_s=-0.142 (p=0.541) for the NHK data.

867 6.6.2 Use of Mole-Based vs. Weight-Based Regressions for the Prediction of Toxicity
868 for Low and High Molecular Weight Substances

869 The ICCVAM ATWG expressed concern that the RC rat-only weight regression may less
870 accurately predict the toxicity of low and high molecular weight substances than the RC rat-
871 only millimole regression. Using the RC IC_{50} and LD_{50} values for the 282 RC substances
872 with rat oral LD_{50} data, analyses were performed to:

873 • Determine the difference in the over and under-prediction of rodent acute oral
874 toxicity (i.e., LD_{50}) from IC_{50} values between low molecular weight
875 substances (i.e., ≤100 g/mole) and substances with molecular weights >100
876 g/mole
877 • Determine the difference in the over and under-prediction of rodent acute oral
878 toxicity from IC_{50} values between high molecular weight substances (i.e.,
879 ≥400 g/mole) vs. substances with molecular weights <400 g/mole.
880 • Compare the RC rat-only millimole regression with the RC rat-only weight
881 regression with respect to the over- and under-prediction of the toxicity of low
882 and high molecular weight substances
883

884 This analysis used the RC data rather than the validation studies data because the RC
885 contains data for many more substances. The analysis assumes that the regressions either
886 underpredicted or overpredicted the toxicity of all of the substances evaluated. In other
887 words, there was a difference between the LD_{50} predicted by the regression and the *in vivo*
888 LD_{50} used to calculate the regression even if it was a tiny fraction (i.e., no substances fit the
889 regression exactly). The complete analysis and discussion are presented in **Appendix J7**.
890

891 Of the 282 RC substances with rat acute oral LD_{50} values, there were 51 with molecular
892 weights ≤100 g/mole and 231 with molecular weights >100 g/mole. For the 51 substances
893 with molecular weight ≤100 g/mole, the RC rat-only millimole regression underestimated the
894 toxicity of 20/51 (39%) substances and overestimated the toxicity of 31/51 (61%) substances
895 (i.e., all were either under- or over-estimated). The RC rat-only weight regression
896 underestimated the toxicity of 24/51 (47%) substances and overestimated the toxicity of
897 27/51 (53%) substances (i.e., all were either under- or over-estimated). Fisher's exact test

898 indicated that there was no difference between the millimole and weight regressions with
899 respect to the under or over-prediction of toxicity for the low molecular weight substances
900 (two-tailed p=0.549) (see **Table 6-11**).
901

902 **Table 6-11 Over- and Under- Prediction of Toxicity for Low and High Molecular**
903 **Weight Substances Using RC Rat-Only Weight and Millimole**
904 **Regressions**
905

Comparison	For	Fisher's Exact Test[1]
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 51 substances with molecular weight ≤100 g/mole	0.549
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 231 substances with molecular weight >100 g/mole	0.575
51 Low molecular weight (≤100 g/mole) substances vs. 231 other substances (>100 g/mole)	RC rat-only millimole regression	0.355
51 Low molecular weight (≤100 g/mole) substances vs. 231 other substances (>100 g/mole)	RC rat-only weight regression	0.756
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 20 substances with molecular weight ≥400 g/mole	0.480
RC rat-only millimole vs. RC rat-only weight regression	Under- and over-prediction of toxicity for 262 substances with molecular weight <400 g/mole	NT
20 High molecular weight substances (≥400 g/mole) vs. 262 other substances (<400 g/mole)	RC rat-only millimole regression	0.362
20 High molecular weight substances (≥400 g/mole) vs. 262 other substances (<400 g/mole)	RC rat-only weight regression	0.033

906 Abbreviations: RC=Registry of Cytotoxicity; NT=Not tested because the proportions were the same. Toxicity was
907 underpredicted for 121/262 (46%) substances and overpredicted for 141/262 (54%) substances.
908 [1]P-values.
909

910 For the 231 substances with molecular weights >100 g/mole, the RC rat-only millimole
911 regression underestimated the toxicity of 108/231 (47%) substances and overestimated the
912 toxicity of 123/231 (53%). The RC rat-only weight regression underestimated the toxicity of
913 101/231 (44%) substances and overestimated the toxicity of 130/231 (57%). Fisher's exact
914 test indicated that there were no significant differences between the millimole and weight
915 regressions for the under- and over-prediction of toxicity for the 231 substances with
916 molecular weight >100 g/mole (two-tailed p =0.575). Fisher's exact test also showed that
917 there were no significant differences in the under- and over-prediction of the toxicity of the

51 substances with molecular weight ≤100 g/mole compared to the under- and over-prediction of the toxicity of the 231 with molecular weight >100 g/mole (two-tailed p=0.756 for the RC rat-only weight regression, and two-tailed p=0.355 for the RC rat-only millimole regression).

Of the 282 RC substances with rat acute oral LD_{50} values, there were 20 with molecular weights ≥400 g/mole and 262 with molecular weights <400 g/mole. The RC rat-only millimole regression underestimated the toxicity of 7/20 (35%) of the ≥400 g/mole substances and overestimated 13/20 (65%). The RC rat-only weight regression underestimated the toxicity of 4/20 (20%) of the substances and overestimated 16/20 (80%). Fisher's exact test indicated that there were no differences between the millimole and weight regressions for the under- and over-prediction of toxicity for the 20 high molecular weight substances (two-tailed p=0.4801).

For the remaining 262 substances with molecular weights <400 g/mole, both the RC rat-only millimole and the RC rat-only weight regressions underestimated the toxicity of 121/262 (46%) substances and overestimated 141/262 (54%). Thus, there were no statistical differences in the under- and over-esimation of toxicity for the 262 substances with molecular weights <400 g/mole regardless of which regression was used. Fisher's exact test also showed that there was no statistical difference in the under- and over-prediction of the toxicity of substances with high molecular weight (≥400 g/mole) compared with the under- and over-prediction the lower molecular weight substances using the RC rat-only millimole regression (two-tailed p=0.362). In contrast the use of the RC rat-only weight regression, resulted in a small but statistically significant difference in the under- and over-prediction of the toxicity of substances with high molecular weight (>400 g/mole) compared with the under- and over-prediction of the toxicity of substances with lower molecular weight (two-tailed p =0.033). The weight-based regression significantly overestimated the toxicity of the high molecular weight substances (compared with substances with lower molecular weight) while the millimole regression did not.

949 **6.7 Salient Issues of Data Interpretation**

950 One of the most important considerations for the 3T3 and NHK NRU test methods, as for
951 any test method, is the ability to generate good concentration-response results. In addition to
952 technical difficulties with these test methods, such as occasional poor cell growth and the
953 formation of NRU crystals, this validation study yielded non-monotonic concentration-
954 response curves for certain substances.
955
956 A number of substances produced non-monotonic concentration-response curves in the 3T3
957 and/or the NHK NRU range finding or definitive tests. Because the *in vitro* NRU test
958 methods, and the calculation of IC_{50} values from the resulting concentration curves, presume
959 that the toxic response is linear, the data from non-linear responses (e.g., biphasic curves), as
960 seen with aminopterin, do not always permit an IC_{50} determination by the standard Hill
961 function analysis. In such cases, the lowest concentration that killed approximately 50% of
962 the cells in the range finding test was used to set the concentration range for the definitive
963 test. The definitive test used more closely spaced concentrations in an attempt to obtain a
964 monotonic concentration-response curve. However, 100% toxicity (or 0%) viability was
965 often unattainable in such definitive tests that exhibited a plateau of toxicity well over 0%
966 viability (e.g., 20%). Care must be used in the calculation of the IC_{50} for curves for which
967 toxicity plateaus to assure that the value reflects the concentration at 50% inhibition of the
968 VC value rather than simply the midpoint of the highest and lowest response.
969
970 Because of low toxicity and/or low solubility, some substances did not produce sufficient
971 toxicity for the calculation of an IC_{50} value. Carbon tetrachloride, methanol, xylene,
972 gibberellic acid, lithium carbonate, and 1,1,1-trichloroethane failed to yield acceptable IC_{50}
973 results in at least one laboratory because of insufficient toxicity. All of these substances, with
974 the exception of methanol, produced precipitate in the cell culture medium.

975 **6.8 Comparison of NRU Test Results to Established Performance Standards**

976 The *Guidance Document* method of evaluating *in vitro* basal cytotoxicity assays for
977 predicting starting doses for rodent acute oral toxicity assays provides the existing
978 performance standard (ICCVAM 2001b) for the 3T3 and NHK NRU test methods. The

979 *Guidance Document* recommends testing 10 to 20 reference substances from the RC in an *in*
980 *vitro* basal cytotoxicity assay for predicting starting doses for rodent acute oral toxicity
981 testing (ICCVAM 2001b). These substances should cover a wide range of toxicity and fit the
982 RC millimole regression as closely as possible. The *Guidance Document* recommends using
983 the IC_{50} results for the selected reference substances from the candidate method to calculate a
984 new regression line with the LD_{50} values used by the RC. If the resulting regression is
985 parallel to the RC millimole regression and within the ± log 5 (i.e., ± 0.699) prediction
986 interval for the RC, candidate assay may be considered effective for predicting starting doses
987 for substances in rodent acute oral toxicity assays.
988
989 One goal of the testing in Phases Ib and II of this study was to establish whether the results
990 from the 3T3 and NHK NRU test methods were consistent with the RC millimole regression.
991 As discussed in **Section 3.3.5**, two of the major criteria for selecting the 12 coded substances
992 tested from the 72 reference substances were:
993 (a) Two substances must be included from each of the unclassified and classified
994 GHS acute oral toxicity categories, and
995 (b) The substances must fit as closely to the RC millimole regression as possible.
996
997 Unfortunately, the SMT could not identify 12 substances that fit both criteria because there
998 was only one substance, aminopterin, in the LD_{50} <5 mg/kg category that fit the RC
999 millimole regression. The other substance chosen from that toxicity category was sodium
1000 selenate. Because sodium selenate was not included in the RC, there was no indication of
1001 how closely it would fit the RC millimole regression, and it was therefore not included in the
1002 Phases Ib and II regression analyses. The other 10 substances selected for testing in phases Ib
1003 and II were colchicine, arsenic trioxide, cadmium chloride, sodium fluoride, propranolol,
1004 lithium carbonate, potassium chloride, chloramphenicol, 2-propanol, and ethylene glycol.
1005
1006 The geometric mean log IC_{50} (mM) values from the 3T3 and NHK NRU test methods from
1007 each laboratory were used with the oral log rodent LD_{50} (mmol/kg) values from the RC (see
1008 **Appendices J1** and **J2**) for the least squares linear regression analyses (see **Section 5.5.3.3**)
1009 for the substances tested in Phases Ib and II. The slopes for all regressions were significantly

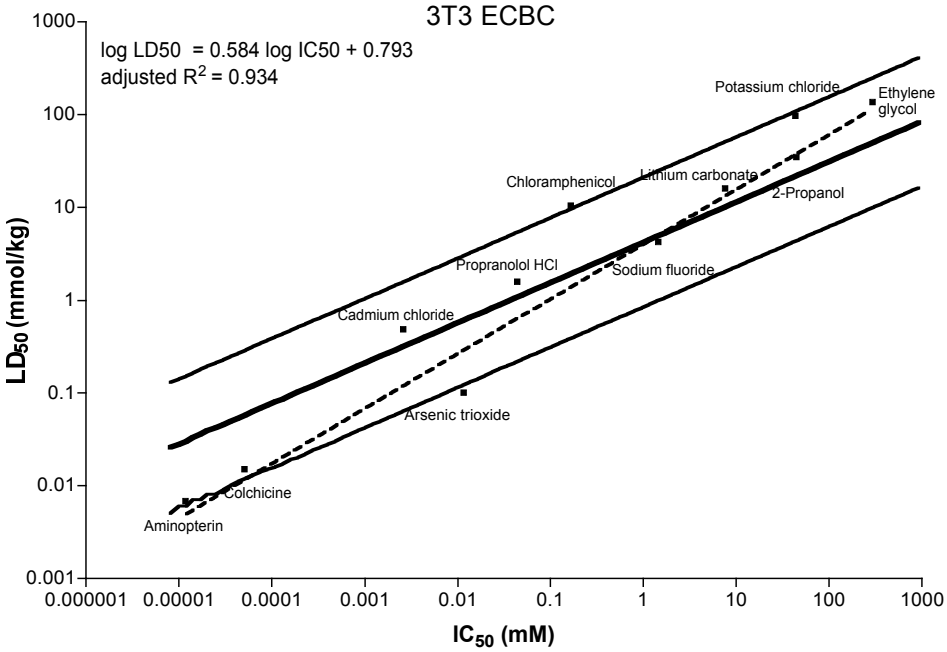
1010 different from zero at $p < 0.0001$, which indicated that there was a significant relationship
1011 between IC_{50} and LD_{50}. The R^2 values for the regressions from each laboratory, shown in
1012 **Table 6-12**, show that the 3T3 NRU test method produced better-fitting regressions than the
1013 corresponding NHK NRU test method ($R^2 = 0.940$ to 0.953 vs. 0.577 to 0.621). The
1014 relatively low R^2 values for the NHK NRU test method were attributed to the much lower
1015 toxicity of aminopterin in those cells (see **Figures 6-6** to **6-8** and **Tables 5-3** and **5-4**). All
1016 test method and laboratory-specific regressions were consistent with the RC millimole
1017 regression. **Table 6-12** shows that all joint comparisons of slopes and intercepts with the RC
1018 millimole regression were not significant (i.e., $p > 0.01$). The RC millimole regression slope
1019 and intercept were used as constants for this comparison.
1020
1021 A graphic comparison of the IC_{50} regressions with the RC millimole regression as suggested
1022 by the *Guidance Document* (ICCVAM 2001b) demonstrated that they were generally within
1023 the RC millimole regression acceptance limits (see **Figures 6-6, 6-7,** and **6-8**). According to
1024 the *Guidance Document* (ICCVAM 2001b), *in vitro* basal cytotoxicity assays providing such
1025 consistency with the RC millimole regression are acceptable for predicting starting doses for
1026 rodent acute oral toxicity assays.
1027
1028 As an additional analysis, a regression for the 11 substances tested in Phases Ib and II (the
1029 *RC-11 millimole regression),* was calculated using the log RC IC_{50} (mM) and log LD_{50}
1030 (mmol/kg) values (see **Table 6-12**). Each of the laboratory regressions for each test method
1031 was then compared to the RC-11 millimole regression using an F test for a joint comparison
1032 of slope and intercept. None of the regressions were significantly different from the RC-11
1033 millimole regression (p values ranged from 0.755 to 0.933).
1034
1035
1036

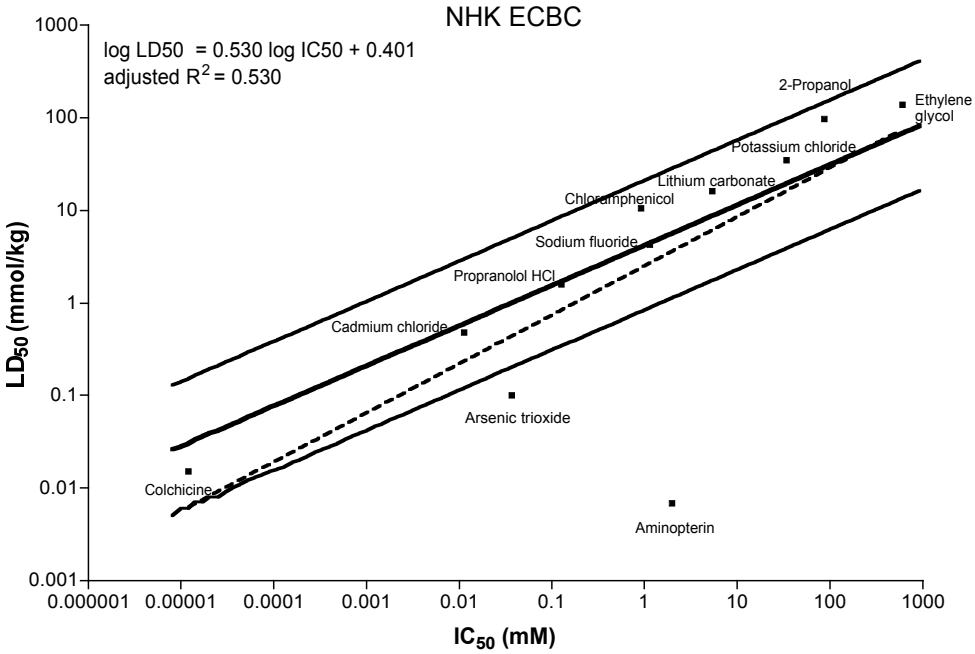
1036 **Table 6-12 Linear Regressions for 11 Substances Tested in Phases Ib and II**
1037

	3T3 Regression[1]				
Laboratory	**Intercept**	**Slope**	**R^2 Statistic**	**Test Against RC Millimole Regression[2]**	**Test Against RC-11 Millimole Regression[3]**
ECBC	0.793	0.584	0.940	0.040	0.829
FAL	0.709	0.598	0.953	0.024	0.909
IIVS	0.710	0.584	0.949	0.041	0.933
	NHK Regression[1]				
Laboratory	**Intercept**	**Slope**	**R^2 Statistic**	**Test Against RC Millimole Regression[2]**	**Test Against RC-11 Millimole Regression[3]**
ECBC	0.401	0.530	0.577	0.620	0.805
FAL	0.429	0.548	0.621	0.569	0.853
IIVS	0.373	0.549	0.590	0.538	0.755

1038 Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical
1039 Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; RC=Registry of Cytotoxicity; 3T3=BALB/c 3T3
1040 fibroblasts; NHK=Normal human keratinocytes; R^2=Coefficient of determination.
1041 [1]Laboratory and test method regressions were calculated after log transforming the NRU IC_{50} in mM and the RC LD_{50} in
1042 mmol/kg for the 11 RC substances tested in study phases Ib and II (shown in **Figures 6-6** through **6-8)**.
1043 [2]Simultaneous comparison of slope and intercept with RC millimole regression: log LD_{50} (mmol/kg) = 0.425 x log IC_{50}
1044 (mM) + 0.625; R^2=0.452; the reported values are p values of the statistic.
1045 [3]Simultaneous comparison of slope and intercept with RC-11 millimole regression (defined as a regression on the 11
1046 substances): log LD_{50} (mmol/kg) = 0.552 x log IC_{50} (mM) + 0.602; R^2=0.971; the reported values are p values of the
1047 statistic.
1048
1049
1050

Figure 6-6 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for ECBC

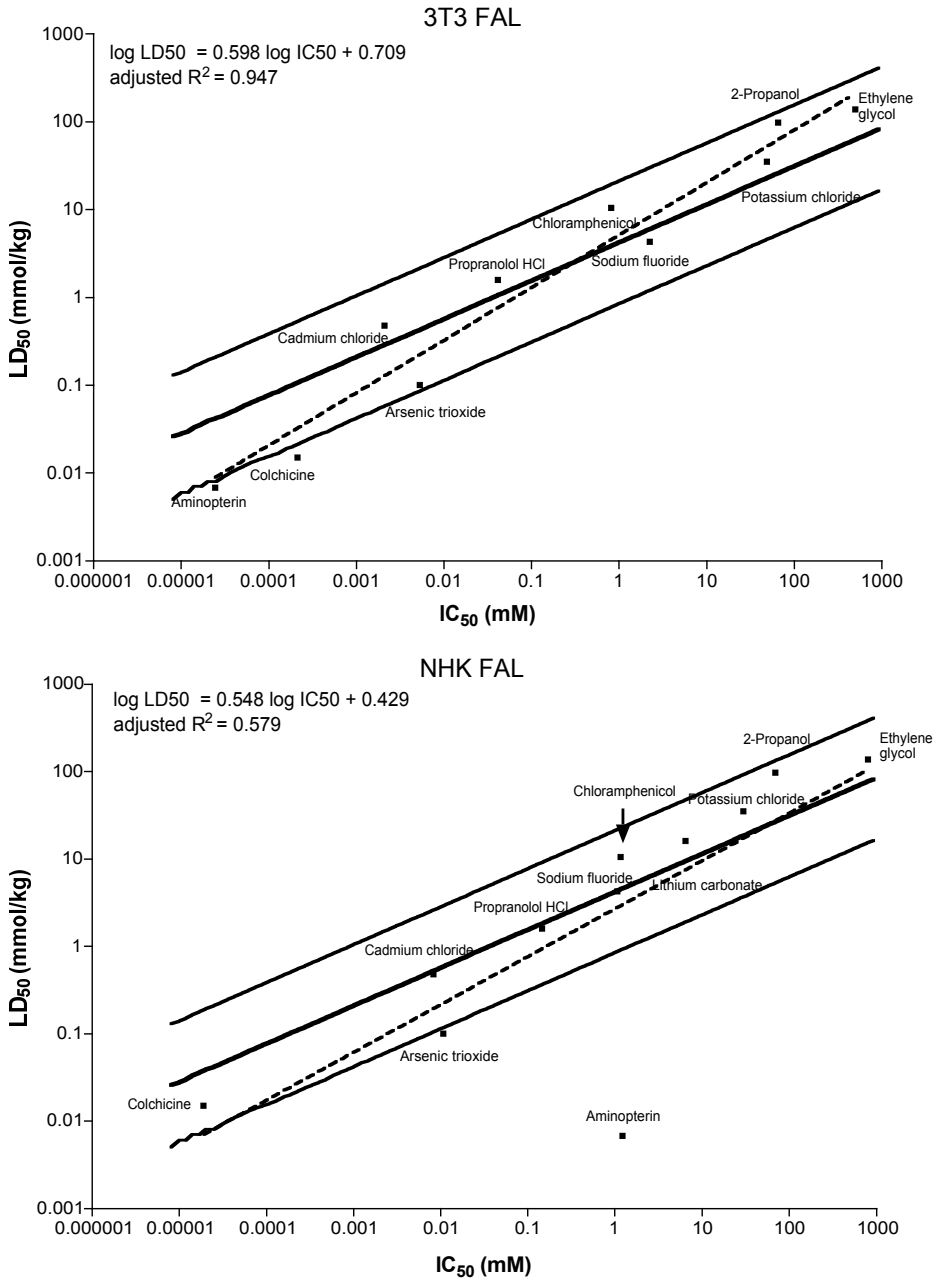




Abbreviations: ECBC=Edgewood Chemical Biological Center; RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK= Neutral red uptake using normal human keratinocytes; R^2=Coefficient of determination.

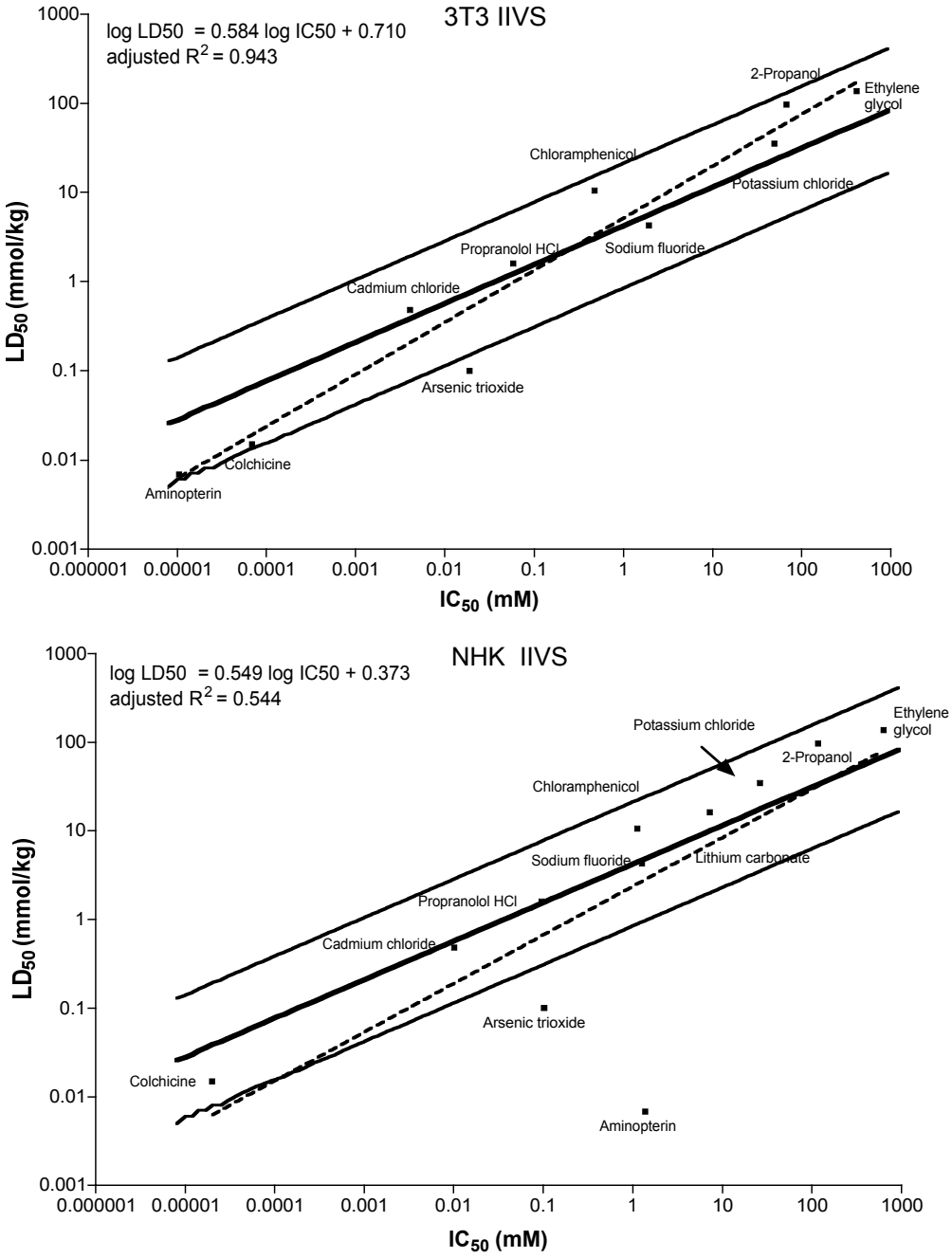
[1]Regressions of substances tested in study phases Ib and II do not include sodium selenate because it was not included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values. The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the ECBC regressions.

1059
1060 **Figure 6-7 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for FAL**



1061
1062 Abbreviations: FAL: Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory
1063 RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK=Neutral red uptake
1064 using normal human keratinocytes; R^2=Coefficient of determination.
1065 [1]Regressions of substances tested in study phases Ib and II do not include sodium selenate because it was not
1066 included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values.
1067 The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the FAL
1068 regressions.
1069

Figure 6-8 *In Vitro – In Vivo* Regressions[1] for Phases Ib and II for IIVS



Abbreviations: IIVS: Institute for *In Vitro* Sciences; RC=Registry of Cytotoxicity; 3T3=Neutral red uptake using BALB/c 3T3 fibroblasts; NHK=Neutral red uptake using normal human keratinocytes; R^2=Coefficient of determination.

[1]Regressions of substances tested in study phases Ib and II do not include sodium selenate because it was not included in the RC. Regressions were calculated using the NRU IC_{50} values and the RC LD_{50} values. The solid lines show RC millimole regression (bold) and acceptance limits (lighter). The dashed shows the IIVS regressions.

1078 **6.9 Summary**

1079 The millimole regressions developed using the validation study IC_{50} and LD_{50} values were
1080 not significantly different from the regressions for the same 47 RC substances using the RC
1081 data (F test; p =0.612 for the 3T3 regression and p =0.759 for the NHK regression). Because
1082 this validation study provided results similar to the RC, which has more than 3.5 times the
1083 number of substances, the 282 RC substances with rat LD_{50} values were used to determine
1084 the relationship between the IC_{50} and LD_{50} data. One linear regression was developed using
1085 millimole units for the measurement of substances, the RC rat-only millimole regression, and
1086 one was developed using weight units (which are more practical in a routine testing
1087 situation), the RC rat-only weight regression. The RC rat-only millimole regression is
1088 applicable to substances of known molecular weight while the RC rat-only weight regression
1089 is applicable for use with complex mixtures, substances whose molecular weight is unknown.
1090

1091 Characteristics that seemed promising for characterizing the RC millimole regression outliers
1092 were chemical class, boiling point, molecular weight, and log K_{ow}. Different chemical classes
1093 behaved differently with respect to being outliers; ranging from 5/5 (100%) for the organic
1094 sulfur compounds for both test methods to 4/14 (29%) for alcohols for the 3T3 NRU. Of the
1095 reference substances with boiling points >200°C, 9/13 (69%) were outliers for the 3T3 NRU
1096 and 8/13 (62%) were outliers for the NHK NRU. With respect to molecular weights, 4/7
1097 (57%) substances with molecular weight >400 g/mole were outliers using the 3T3 data, and
1098 3/7 (43%) were outliers using the NHK data. When log K_{ow} was used, 8/13 (62%) substances
1099 with a log K_{ow} >3 were outliers for both test methods.
1100

1101 The lack of fit of individual substances to the RC millimole regression was not consistently
1102 related to insolubility or to the fact that the test method systems had little to no metabolic
1103 capability. Of the substances that exhibited precipitation, 11/25 (44%) were outliers in the
1104 3T3 NRU assays and 11/24 (46%) were outliers in the NHK NRU assays. However, although
1105 the 3T3 and NHK cells have little to no metabolic capability, the toxicity of substances
1106 known to produce active metabolites *in vivo* was not underpredicted by these assays. Of the
1107 19 substances known to produce active metabolites *in vivo*, 10 (53%) were outliers in the 3T3
1108 NRU test method; the toxicity of six (60%) was underpredicted while the toxicity of four

1109 (40%) overpredicted. These 10 substances accounted for 36% of the 28 outliers identified by
1110 the 3T3 NRU test method. Similarly, nine (47%) of the 19 substances known to produce
1111 active metabolites *in vivo* were outliers in the NHK NRU test method. Of these nine, the
1112 NHK NRU test method underpredicted the toxicity of five (56%) and overpredicted four
1113 (44%). These nine outliers accounted for 29% of the 31 outliers identified by the NHK NRU
1114 test method.
1115
1116 The examination of outliers based on mechanisms of toxicity showed that 10/16 (63%)
1117 substances with specific neurotoxic mechanisms were outliers in both the 3T3 and NHK
1118 NRU test methods. Three of the six (50%) cardiotoxic substances were outliers in the 3T3
1119 NRU test method and two (33%) were outliers in the NHK NRU test method. When all the
1120 reference substances with mechanisms of toxicity that are not expected to be active in the
1121 3T3 and NHK systems (i.e., in **Table 6-3**) were summed, 13/22 (59%) were outliers for the
1122 3T3 NRU and 12/22 (55%) were outliers for the NHK NRU.
1123
1124 The accuracy of the 3T3 and NHK NRU test methods for predicting the GHS acute oral
1125 toxicity categories was 31% (21/67) and 29% (20/68), respectively, when used with the RC
1126 rat-only millimole regression. The corresponding accuracy with the RC rat-only weight
1127 regression was 31% for both methods (21/67 for 3T3, and 21/68 for NHK). Accuracy was
1128 highest for substances in the $300 < LD_{50} \leq 2000$ mg/kg range. The accuracies of the
1129 regressions, with respect to the GHS categories, were similar for both regressions (millimole
1130 and weight) and all three laboratories.
1131 - 0% for substances with $LD_{50} \leq 5$ mg/kg (GHS Category I)
1132 - 9% to 18% for substances with $5 < LD_{50} \leq 50$ mg/kg (GHS Category II)
1133 - 33% to 50% for substances with $50 < LD_{50} \leq 300$ mg/kg (GHS Category III)
1134 - 75% to 81% for substances with $300 < LD_{50} \leq 2000$ mg/kg (GHS Category IV)
1135 - 0% to 40% for substances with $2000 < LD_{50} \leq 5000$ mg/kg (GHS Category V)
1136 - 0% to 17% for substances with $LD_{50} > 5000$ mg/kg (GHS Unclassified)
1137
1138 The overall accuracy for prediction of GHS category prediction using the RC IC_{50} and LD_{50}
1139 values and the RC millimole regression was higher that that for the NRU test methods with

1140 the RC rat-only regressions (i.e., 40% for the RC vs. 29% to 31% for the NRU test methods
1141 and RC rat-only regressions). However, the pattern of accuracy for the GHS categories was
1142 similar. For all the accuracy analyses, the lowest accuracy was obtained for very toxic and
1143 very nontoxic substances and highest accuracy was obtained for substances with $300 < LD_{50}$
1144 ≤ 2000 mg/kg.
1145
1146 The accuracy of GHS acute oral toxicity category predictions using the *in vitro* NRU test
1147 methods with the RC rat-only regressions obtained for the reference substances may or may
1148 not be broadly applicable to substances that might require acute oral toxicity testing. The
1149 reasons for the low accuracy obtained in this validation study include: the differences
1150 between cell cultures and whole animals regarding the absorption, distribution, availability,
1151 metabolism, and excretion of reference substances, and the presence or absence of toxicity
1152 targets; the skewness of the selection of substances for testing (with respect to fit to the
1153 regression); and the structure of the GHS acute oral toxicity categories.
1154

[This Page Intentionally Left Blank]

7.0 RELIABILITY OF THE 3T3 AND NHK NRU TEST METHODS.....7-3

7.1 Reference Substances Used to Determine the Reliability of the 3T3 and NHK NRU Test Methods.....7-3

7.2 Reproducibility Analyses for the 3T3 and NHK NRU Test Methods.....7-5
7.2.1 Comparison of Laboratory-Specific IC_{50}-LD_{50} Linear Regression Analyses to the Mean Laboratory Regression.....7-6
7.2.2 ANOVA Results for the 3T3 and NHK NRU Test Methods 7-6
7.2.3 CV Results for the 3T3 and NHK NRU Test Methods.....7-22
7.2.4 Comparison of Maximum to Minimum IC_{50} Values Using Laboratory Means 7-41
7.2.5 Comparison of Maximum:Minimum IC_{50} Ratios with the Maximum:Minimum LD_{50} Ratios 7-43
7.2.6 Normalization of Reference Substance IC_{50} Values Using SLS Values 7-44

7.3 Historical Positive Control Data 7-48
7.3.1 ANOVA and Linear Regression Results for the 3T3 NRU Test Method....7-50
7.3.2 ANOVA and Linear Regression Results for the NHK NRU Test Method..7-52

7.4 Laboratory Concordance for Solvent Selection 7-54

7.5 Summary.....7-57

[This Page Intentionally Left Blank]

7.0 RELIABILITY OF THE 3T3 AND NHK NRU TEST METHODS

The reliability of the 3T3 and NHK NRU test methods was assessed by determining intra- and inter-laboratory reproducibility. Intralaboratory reproducibility is the agreement of results produced when people in the same laboratory perform the method using the same test protocol at different times (ICCVAM 2003). Interlaboratory reproducibility is the agreement of results among different laboratories using the same protocol and reference substances. Interlaboratory reproducibility indicates the extent to which a method can be successfully transferred among laboratories. Repeatability, usually applied to results within a laboratory, is the closeness of agreement between test results obtained when the procedure is performed on the same substance under identical conditions within a given time. This study was not designed to assess intralaboratory repeatability.

The interlaboratory reproducibility of the test results was assessed by comparing the laboratory-specific IC_{50}-LD_{50} regressions for the 3T3 and NHK NRU test methods to the mean (i.e., across-laboratory mean) laboratory regressions (see **Section 7.2.1**). This comparison is relevant because the 3T3 and NHK NRU test methods are intended for use with IC_{50}-LD_{50} regressions to determine starting doses for acute oral toxicity tests. Interlaboratory reproducibility of the 3T3 and NHK NRU test methods was also determined using ANOVA, CV analysis, and comparison of maximum:minimum IC_{50} ratios calculated using laboratory mean values **(**see **Sections 7.2.2**, **7.2.3**, and **7.2.4**, respectively), as discussed in **Section 5.5.2.2**. Inter- and intra-laboratory reproducibility of the PC (SLS) was determined using ANOVA, CV analysis, and/or linear regression over time (see **Section 7.3**). The extent of laboratory concordance in selecting the solvent to be used for each test substance (described in **Section 2.10**) is provided in **Section 7.4**.

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

The validation study was designed for the purpose of using the IC_{50} results of 72 reference substances (see **Table 3-2)** to determine the reliability of the IC_{50} values from the 3T3 and NHK NRU test methods. The number of reference substances used for the reproducibility analysis was not the same as the number of reference substances used for the accuracy

analyses in **Section 6.4**. In the former case, only reference substances for which all three laboratories reported replicate IC_{50} values were used, while in the latter case, substances with rat acute oral LD_{50} data only and at least one laboratory reporting replicate IC_{50} values were used. **Table 7-1** lists the reference substances that failed to yield sufficient toxicity for the calculation of an IC_{50} in each laboratory, and the number of remaining reference substances with replicate IC_{50} values. The laboratories obtained acceptable IC_{50} values for 66 to 68 reference substances using the 3T3 NRU test method, and for 69 to 70 substances using the NHK NRU test method. When only reference substances with IC_{50} values from all three laboratories are considered, 64 and 68 substances were available to evaluate the reliability of the 3T3 and NHK NRU test methods, respectively. The substances that were excluded from the 3T3 reliability analysis were carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic acid, and xylene. The substances that were excluded from the NHK reliability analysis were carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene.

Table 7-1 Reference Substances Excluded from Reproducibility Analyses Because of Insufficient Cytotoxicity

Laboratory	3T3 NRU Test Method		NHK NRU Test Method	
	Reference Substances Lacking IC_{50} Results	N[1]	Reference Substances Lacking IC_{50} Results	N[1]
ECBC	Carbon tetrachloride Methanol 1,1,1-Trichloroethane Xylene	68	Carbon tetrachloride Methanol Xylene	69
FAL	Carbon tetrachloride Disulfoton Gibberellic acid Lithium carbonate Methanol Xylene	66	1,1,1-Trichloroethane Carbon tetrachloride Xylene	69
IIVS	Carbon tetrachloride Lithium carbonate Methanol Valproic acid	68	Carbon tetrachloride 1,1,1-Trichloroethane	70

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; N=Number of substances.
[2]Number of substances with replicate IC_{50} values.

Despite the fact that IC_{50} values were not obtained by all the laboratories for all reference substances, **Table 7-2** shows that the complete range of LD_{50} responses, as defined by the GHS classification for acute oral toxicity in **Table 3-1**, was covered by the reference substances for which replicate IC_{50} values were obtained. The 3T3 NRU IC_{50} values ranged from 0.005 to 38,878 µg/mL, while the NHK values covered a larger range, from 0.00005 to 49,800 µg/mL (see **Tables 5-4** and **5-5**).

Table 7-2 Number of Reference Substances Tested vs Number of Reference Substances Yielding IC_{50} Values from Each Laboratory, by GHS Acute Oral Toxicity Category

GHS Category[1] (mg/kg)	Reference Oral LD_{50}[2]	3T3 NRU Test Method[3]	NHK NRU Test Method[3]
$LD_{50} \leq 5$	7	6	7
$5 < LD_{50} \leq 50$	12	12	12
$50 < LD_{50} \leq 300$	12	12	12
$300 < LD_{50} \leq 2000$	16	14	16
$2000 < LD_{50} \leq 5000$	11	9	9
$LD_{50} > 5000$	14	11	12

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).
[1]GHS category for acute oral toxicity.
[2]Number of reference substances tested in each category. Reference acute oral LD_{50} values from rats and mice were generated after evaluating LD_{50} values located through literature searches and references from toxicity databases such as RTECS® (from **Table 4-2**).
[3]Number of reference substances with IC_{50} values from all three laboratories.

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

The interlaboratory reproducibility of the 3T3 and NHK NRU IC_{50} values was assessed by comparing the laboratory-specific IC_{50}-LD_{50} linear regressions for each method to a regression calculated using the mean IC_{50} values of the laboratories. The interlaboratory reproducibility of the 3T3 and NHK NRU test methods was also assessed using ANOVA, CV analysis, and analysis of the laboratory mean maximum:minimum IC_{50} ratios, as described in **Section 5.5.2.2.** Intralaboratory reproducibility was assessed using a CV analysis.

121 7.2.1 Comparison of Laboratory-Specific IC_{50}-LD_{50} Linear Regression Analyses to the
122 Mean Laboratory Regression

123 The comparisons of laboratory-specific IC_{50}-LD_{50} linear regressions to the mean laboratory
124 regression for each method were made because the 3T3 and NHK NRU test methods are
125 intended for use with IC_{50}-LD_{50} regressions to determine starting doses for acute oral toxicity
126 tests. Laboratory-specific IC_{50}-LD_{50} linear regressions were generated and displayed
127 graphically for each method using the 64 and 68 reference substances for the 3T3 and NHK
128 NRU test methods, respectively, as indicated in **Section 7.1**. The regressions used the
129 geometric mean IC_{50} values for each substance with the rodent acute oral LD_{50} reference
130 value (**Table 4-2**). To determine whether the laboratory-specific regressions were
131 significantly different from one another, they were compared against the mean laboratory
132 regression for each NRU test method that was calculated using the geometric mean of the
133 laboratory mean IC_{50} values and the rodent acute oral LD_{50} reference values. The mean
134 laboratory regression for each NRU test method is in **Figure 7-1** with 95% confidence limits,
135 and shows that the laboratory-specific regressions were all within the 95% confidence limits
136 of the mean laboratory regression.

137 7.2.2 ANOVA Results for the 3T3 and NHK NRU Test Methods

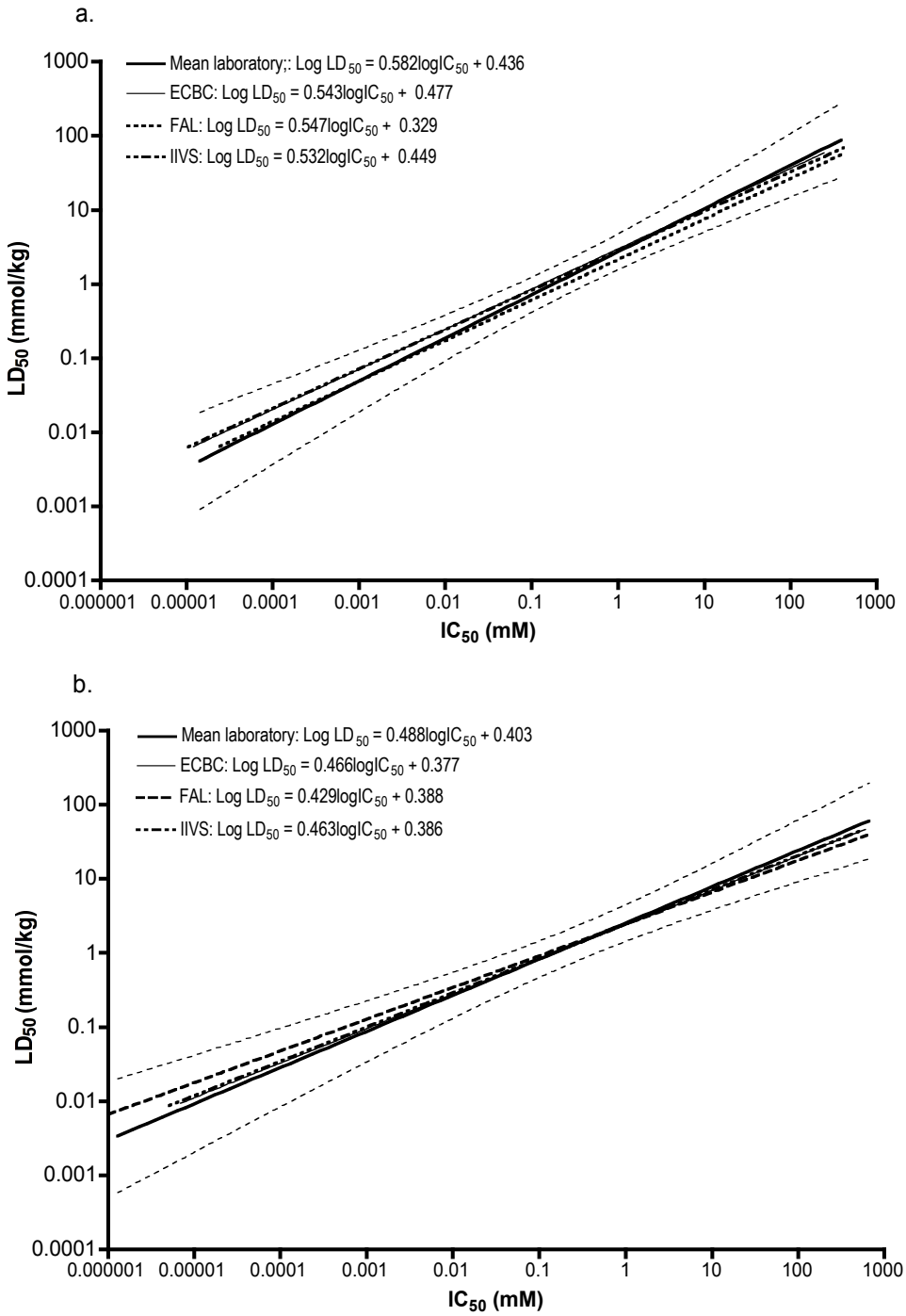
138 The ANOVA was performed as discussed in Section **5.5.2.2**. Because the sample sizes from
139 this study were small, usually three observations per laboratory, there may be differences that
140 were statistically significant only because there were too few observations within the
141 laboratories to adequately characterize variability or because the within-laboratory variability
142 was small.

143 7.2.2.1 *Differences Among the IC_{50} Values in Laboratories Using the 3T3 NRU Test*
144 *Method*

145 The ANOVA results in **Table 7-3** show that there were statistically significant ($p < 0.01$)
146 differences among the laboratories for 23 of the 64 (36%) reference substances evaluated.
147 The p values from the contrast analyses, post-hoc tests to determine which laboratory was
148 significantly different from the others at $p < 0.01$ (see **Section 5.5.2.2)**, are also provided in
149 **Table 7-3**. The substances for which statistically significant ANOVA and contrast results
150 were obtained are listed in **Table 7-4** along with columns showing the laboratory with

151 significantly differing values from the other two laboratories. Because significant laboratory
152 differences may have resulted from the insolubility or volatility of the test substance, **Table**
153 **7-4** also indicates whether any laboratory reported insolubility or volatility during conduct of
154 the test. Insolubility was suggested by the presence of precipitates in either the stock
155 solutions or in cell culture. Volatility was identified by the need for plate sealers to contain
156 volatile contamination of lower concentration wells by higher concentrations. Insolubility
157 and volatility were reported for only six of the 23 chemicals showing significant
158 interlaboratory variability. In contrast, 22 of the 41 substances that were classified as
159 generating interlaboratory reproducible data exhibited precipitates and/or volatility.
160
161 For the 23 substances that yielded significantly different results among laboratories, contrast
162 analyses indicated that the IC_{50} values produced by ECBC and FAL were frequently different
163 from the other laboratories. ECBC tended to report the lowest IC_{50} values (i.e., highest
164 toxicity) among the laboratories while FAL tended to report the highest values of the three
165 laboratories. ECBC reported significantly different results from the other two laboratories for
166 15 of the 23 substances; for 13 of the 15, ECBC's mean value IC_{50} was the lowest among the
167 laboratories. FAL reported significantly different results from the other two laboratories for
168 20 of the 23 substances; for 18 of the 20, FAL's IC_{50} value was the highest among the
169 laboratories. IIVS reported significantly different values for 11 of the 26 substances, with no
170 tendency toward highest or lowest IC_{50} values.
171
172

Figure 7-1 Mean Laboratory and Laboratory-Specific 3T3 and NHK NRU Regressions



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; NHK=Normal human keratinocytes.
Solid lines show the mean laboratory linear regressions for the 3T3 NRU (a) and the NHK NRU (b) test methods with dashed curved lines to show the 95% confidence limits of the regression. The regressions were calculated using 64 and 68 reference substances for the 3T3 and NHK NRU test methods, respectively, as described in **Section 7.1**. Regressions used geometric mean IC_{50} values and reference acute oral LD_{50} values from **Table 4-2**.

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Acetaminophen	50.1	1.6		28	1.7	0.171	
ECBC	40.8		22		1.61		NA
FAL	66.2		35		1.82		NA
IIVS	43.4		26		1.64		NA
Acetonitrile	8484	1.5		21	3.93	0.553	
ECBC	6433		2		3.81		NA
FAL	9690		58		3.99		NA
IIVS	9330		13		3.97		NA
Acetylsalicylic acid	760	3.1		56	2.88	<0.001	
ECBC	646		10		2.81		0.581
FAL	1234		24		3.09		<0.001
IIVS	401		16		2.6		<0.001
5-Aminosalicylic acid	1698	1.4		19	3.23	0.054	
ECBC	1467		14		3.17		NA
FAL	2070		16		3.32		NA
IIVS	1557		12		3.19		NA
Aminopterin	0.007	2.4		54	-2.14	0.036	
ECBC	0.005		20		-2.28		NA
FAL	0.012		46		-1.93		NA
IIVS	0.005		23		-2.33		NA
Amitriptyline HCl	7.23	1.3		14	0.86	0.348	
ECBC	6.03		23		0.78		0.163
FAL	7.86		28		0.9		0.469
IIVS	7.81		18		0.89		0.445
Arsenic trioxide	2.51	3.9		61	0.4	0.004	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	2.41		33		0.38		0.527
FAL	1.04		7		0.02		0.002
IIVS	4.09		52		0.61		0.006
Atropine sulfate	85.6	2.5		49	1.93	0.049	
ECBC	54.1		55		1.73		NA
FAL	133		31		2.12		NA
IIVS	70		8		1.85		NA
Boric acid	2228	3.3		69	3.35	0.01	
ECBC	1497		32		3.18		NA
FAL	3987		17		3.6		NA
IIVS	1202		48		3.08		NA
Busulfan	135	8.0		119	2.13	0.002	
ECBC	40		48		1.6		0.012
FAL	321		56		2.51		<0.001
IIVS	43.7		4		1.64		0.033
Cadmium chloride	0.565	1.4		39	-0.25	0.124	
ECBC	0.48		14		-0.32		NA
FAL	0.4		32		-0.4		NA
IIVS	0.817		53		-0.09		NA
Caffeine	161	1.4		18	2.21	0.481	
ECBC	133		10		2.12		NA
FAL	157		52		2.2		NA
IIVS	191		7.5		2.28		NA
Carbamazepine	109	1.8		35	2.04	0.049	
ECBC	83		14		1.92		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	152		37		2.18		NA
IIVS	91.8		12		1.96		NA
Carbon tetrachloride	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Chloral hydrate	187	1.6		25	2.27	0.004	
ECBC	151		10		2.18		0.008
FAL	241		10		2.38		0.002
IIVS	170		12		2.23		0.181
Chloramphenicol	161	4.9		67	2.21	<0.001	
ECBC	55.3		22		1.74		<0.001
FAL	273		30		2.44		0.001
IIVS	156		18		2.19		0.165
Citric acid	829	2.4		41	2.92	0.002	
ECBC	473		29		2.68		0.001
FAL	1148		13		3.06		0.003
IIVS	865		19		2.94		0.298
Colchicine	0.047	4.7		85	-1.33	0.001	
ECBC	0.02		11		-1.70		0.0028
FAL	0.093		45		-1.03		0.0005
IIVS	0.028		1		-1.55		0.0914
Cupric sulfate pentahydrate	70.6	21.6		85	1.85	<0.001	
ECBC	82.7		4		1.92		0.001
FAL	123		44		2.09		<0.001

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	5.7		31		0.76		<0.001
Cycloheximide	0.293	5.9		104	-0.53	0.021	
ECBC	0.125		45		-0.9		NA
FAL	0.647		70		-0.19		NA
IIVS	0.109		23		-0.96		NA
Dibutyl phthalate	78.3	9.2		124	1.89	<0.001	
ECBC	23.5		17		1.37		0.012
FAL	191		50		2.28		<0.001
IIVS	20.7		7		1.32		0.005
Dichlorvos	20.3	3.3		57	1.31	0.002	
ECBC	9.8		35		0.99		0.001
FAL	32.8		6		1.52		0.002
IIVS	18.3		11		1.26		0.823
Diethyl phthalate	113	1.7		28	2.05	0.127	
ECBC	85.5		34		1.93		0.092
FAL	147		26		2.17		0.07
IIVS	106		24		2.03		0.846
Digoxin	520	2.8		62	2.72	0.043	
ECBC	351		39		2.54		NA
FAL	892		36		2.95		NA
IIVS	317		21		2.5		NA
Dimethylformamide	5242	1.1		6	3.72	0.296	
ECBC	5343		10		3.73		NA
FAL	5483		9		3.74		NA
IIVS	4900		4		3.69		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Diquat dibromide monohydrate	15.1	9.3		120	1.18	0.017	
ECBC	3.9		23		0.59		NA
FAL	36.1		98		1.56		NA
IIVS	5.4		25		0.73		NA
Disulfoton	98.6	2.3		55	1.99	0.003	
ECBC	137		55		2.14		NA
FAL	NA		NA		NA		NA
IIVS	60.4		87		1.78		NA
Endosulfan	8.02	4.2		78	0.9	0.046	
ECBC	5.3		57		0.72		NA
FAL	15.2		78		1.18		NA
IIVS	3.6		42		0.56		NA
Epinephrine bitartrate	59.4	1.2		12	1.77	0.048	
ECBC	51.5		12		1.71		NA
FAL	63.4		11		1.8		NA
IIVS	63.4		3		1.8		NA
Ethanol	6731	1.6		23	3.83	0.075	
ECBC	5360		33		3.73		NA
FAL	8420		14		3.93		NA
IIVS	6413		5		3.81		NA
Ethylene glycol	25292	1.7		26	4.4	0.007	
ECBC	18325		9		4.26		0.004
FAL	31650		24		4.50		0.01
IIVS	25900		12		4.41		0.505
Fenpropathrin	27.2	2.5		49	1.43	0.301	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	22.6		11		1.35		NA
FAL	42.4		63		1.63		NA
IIVS	16.7		12		1.22		NA
Gibberellic Acid	7842	1.0		3	3.89	0.621	
ECBC	8027		11		3.9		NA
FAL	NA		NA		NA		NA
IIVS	7657		10		3.88		NA
Glutethimide	192	2.3		43	2.28	<0.001	
ECBC	167		4		2.22		0.029
FAL	284.3		7		2.45		<0.001
IIVS	125.3		7		2.1		<0.001
Glycerol	28904	1.9		33	4.46	0.846	
ECBC	20000		15		4.3		NA
FAL	38878		73		4.59		NA
IIVS	27833		39		4.44		NA
Haloperidol	6.26	1.5		24	0.8	0.006	
ECBC	5.3		12		0.72		0.03
FAL	8		8		0.9		0.002
IIVS	5.5		12		0.74		0.061
Hexachlorophene	4.48	1.7		27	0.65	0.174	
ECBC	5		48		0.7		NA
FAL	5.3		33		0.72		NA
IIVS	3.1		9		0.49		NA
Lactic acid	3073	1.2		12	3.49	0.16	
ECBC	2943		11		3.47		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	3487		16		3.54		NA
IIVS	2790		9		3.45		NA
Lindane	161	2.9		58	2.21	0.066	
ECBC	125		95		2.1		NA
FAL	266		36		2.43		NA
IIVS	90.4		122		1.96		NA
Lithium carbonate	NA	NA		NA	NA	NA	NA
ECBC	564		12		2.75		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Meprobamate	539	2.5		54	2.73	<0.001	
ECBC	353		14		2.55		NA
FAL	877		15		2.94		NA
IIVS	386		2		2.59		NA
Mercury chloride	4.32	1.7		33	0.64	0.021	
ECBC	3.5		5		0.54		NA
FAL	6		31		0.78		NA
IIVS	3.5		3		0.54		NA
Methanol	NA	NA		NA	NA	NA	NA
ECBC	NA		NA		NA		NA
FAL	NA				NA		NA
IIVS	NA				NA		NA
Nicotine	378	1.7		25	2.58	0.128	
ECBC	272		24		2.43		NA
FAL	412		33		2.61		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	450		12		2.65		NA
Paraquat	23.3	1.2		8	1.37	1	
ECBC	21.3		34		1.33		NA
FAL	24.9		67		1.4		NA
IIVS	23.7		64		1.37		NA
Parathion	61.8	6.4		111	1.79	0.014	
ECBC	22.7		53		1.36		NA
FAL	141		70		2.15		NA
IIVS	22		22		1.34		NA
Phenobarbital	612	1.5		21	2.79	0.232	
ECBC	634		21		2.8		NA
FAL	726		35		2.86		NA
IIVS	476		23		2.68		NA
Phenol	70.9	2.1		41		0.011	
ECBC	50.2		22		1.7		NA
FAL	104		24		2.02		NA
IIVS	58.1		12		1.76		NA
Phenylthiourea	119	7.9		90	2.08	0.007	
ECBC	30.1		66		1.48		0.004
FAL	239		28		2.38		0.006
IIVS	89		25		1.95		0.718
Physostigmine	28.8	1.9		30	1.46	0.149	
ECBC	28.2		53		1.45		NA
FAL	37.8		5		1.58		NA
IIVS	20.4		33		1.31		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Potassium chloride	3635	1.1		7	3.56	0.846	
ECBC	3352		14		3.53		NA
FAL	3842		31		3.58		NA
IIVS	3710		11		3.57		NA
Potassium cyanide	64.3	10.4		127	1.81	<0.001	
ECBC	15.3		25		1.18		0.001
FAL	159		52		2.2		<0.001
IIVS	18.9		5		1.28		0.006
Procainamide HCl	443	1.2		11	2.65	0.007	
ECBC	400		4		2.6		0.008
FAL	431		1		2.63		0.396
IIVS	497		8		2.7		0.003
2-Propanol	3563	1.6		23	3.55	0.001	
ECBC	2610		9		3.42		<0.001
FAL	3970		4		3.6		0.004
IIVS	4110		4		3.61		0.002
Propranolol HCl	14.9	1.3		16	1.17	0.488	
ECBC	13.6		32		1.13		NA
FAL	13.5		51		1.13		NA
IIVS	17.6		21		1.25		NA
Propylparaben	29.9	3.0		64	1.48	0.001	
ECBC	20.9		16		1.32		0.045
FAL	51.8		29		1.71		<0.001
IIVS	17.1		12		1.23		0.003
Sodium arsenite	0.873	2.8		55	-0.06	0.028	

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	0.5		6		-0.3		NA
FAL	1.4		57		0.15		NA
IIVS	0.7		17		-0.15		NA
Sodium chloride	4764	1.1		3	3.68	0.759	
ECBC	4790		5		3.68		NA
FAL	4625		13		3.67		NA
IIVS	4877		9		3.69		NA
Sodium dichromate dihydrate	0.602	1.2		9	-0.22	0.822	
ECBC	0.603		14		-0.22		NA
FAL	0.657		37		-0.18		NA
IIVS	0.547		17		-0.26		NA
Sodium fluoride	79.8	1.6		22	1.9	0.016	
ECBC	61.3		9		1.79		NA
FAL	96.1		18		1.98		NA
IIVS	82		7		1.91		NA
Sodium hypochlorite	1211	2.5		57	3.08	0.04	
ECBC	823		13		2.92		NA
FAL	805		46		2.91		NA
IIVS	2005		44		3.3		NA
Sodium oxalate	40.8	1.6		23	1.61	0.643	
ECBC	42		41		1.62		NA
FAL	31		28		1.49		NA
IIVS	49.5		53		1.69		NA
Sodium selenate	34.5	4.3		60	1.54	<0.001	
ECBC	12.7		13		1.1		<0.001

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	54.2		19		1.73		<0.001
IIVS	36.5		14		1.56		0.026
Strychnine	199	4.7		83	2.3	<0.001	
ECBC	389		21		2.59		<0.001
FAL	124		16		2.09		0.018
IIVS	83.5		6		1.92		<0.001
Thallium Sulfate	7.5	4.9		72	0.88	0.165	
ECBC	2.8		24		0.45		NA
FAL	13.4		78		1.13		NA
IIVS	6.3		28		0.8		NA
Trichloroacetic acid	928	1.6		27	2.97	0.005	
ECBC	762		13		2.88		0.022
FAL	1220		6		3.09		0.002
IIVS	801		14		2.9		0.069
1,1,1-Trichloroethane	15538	2.2		52	4.19	<0.001	
ECBC	NA		NA		NA		NA
FAL	21250		11		4.33		NA
IIVS	9827		2		3.99		NA
Triethylenemelamine	0.568	16.9		135	-0.25	<0.001	
ECBC	0.086		11		-1.07		<0.001
FAL	1.45		18		0.16		<0.001
IIVS	0.169		29		-0.77		0.002
Triphenyltin hydroxide	0.022	1.7		29	-1.66	0.688	
ECBC	0.026		17		-1.59		NA
FAL	0.026		81		-1.59		NA

Table 7-3 Interlaboratory Reproducibility of the IC_{50} Values from the 3T3 NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	0.015		55		-1.83		NA
Valproic acid	1177	3.3		76	3.07	<0.001	
ECBC	547		12		2.74		NA
FAL	1807		10		3.26		NA
IIVS	NA		NA		NA		NA
Verapamil HCl	35.2	1.2		10	1.55	0.23	
ECBC	32		18		1.51		NA
FAL	34.6		5		1.54		NA
IIVS	38.9		11		1.59		NA
Xylene	NA	NA		NA	NA	NA	NA
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	724		12		2.86		NA

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.
[1]Results reported on the same row with chemical names are the means of all the laboratories. Results reported on the same row as laboratories are the laboratory means.
[2]Maximum laboratory mean IC_{50} divided by minimum laboratory mean IC_{50}.
[3]$p < 0.01$ indicated statistical significance.
[4]Contrasts were performed if ANOVA was significant ($p < 0.01$) to determine which laboratory was different from the other two laboratories. Significant contrasts were denoted by $p < 0.01$. No contrast tests were performed if only two laboratories reported IC_{50} values.

Table 7-4 Reference Substances with Significant (ANOVA) Differences Among Laboratories Using the 3T3 NRU Test Method

Reference Substance	Significant Contrast Results[1]			Insoluble/ Volatile[2]
	ECBC	FAL	IIVS	
Acetylsalicylic acid		H	L	
Arsenic trioxide		L	H	Precipitate
Busulfan		H		
Chloral hydrate	L	H		
Chloramphenicol	L	H		
Citric acid	L	H		
Colchicine	L	H		
Cupric sulfate pentahydrate	M	H	L	
Dibutyl phthalate		H	L	Precipitate
Dichlorvos	L	H		Precipitate
Ethylene glycol	L			
Glutethimide		H	L	
Haloperidol		H		
Meprobamate	L	H	M	
Phenylthiourea	L	H		
Potassium cyanide	L	H	M	Precipitate /Volatile
Procainamide HCl	L		H	
2-Propanol	L	M	H	Volatile
Propylparaben		H	L	
Sodium selenate	L	H		
Strychnine	H		L	Precipitate
Trichloroacetic acid		H		
Triethylenemelamine	L	H		

Abbreviations: ANOVA=Analysis of variance; 3T3=BALB/c 3T3 fibroblasts; H=Laboratory reported the highest mean IC_{50}; L=Laboratory reported the lowest mean IC_{50}; M=Laboratory reported a mean IC_{50} between the values of the other two laboratories; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Laboratories significantly different from the other two at $p < 0.01$.
[2]From **Table 5-11**. Precipitate reported by at least one laboratory is indicated by "Precipitate". Use of plate sealers by at least one laboratory to prevent volatile contamination of control wells indicated by "Volatility".

205 7.2.2.2 *Differences Among the IC_{50} Values in Laboratories Using the NHK NRU Test*
206 *Method*

207 The ANOVA results in **Table 7-5** indicate that there were statistically significant ($p < 0.01$)
208 laboratory differences for six of the 68 (9%) reference substances evaluated. These
209 substances are listed in **Table 7-6** along with columns showing which laboratory's IC_{50}
210 values were statistically significantly different from the other two (as indicated by the
211 contrast results), and indications of insolubility or volatility during conduct of the assay.
212 Insolubility was reported for three of the six substances, but none of the six substances were
213 volatile.
214
215 For the six substances that yielded significantly different IC_{50} values among the laboratories,
216 ECBC reported the highest IC_{50} value for four substances and the lowest for one, FAL
217 reported the lowest values for three substances and the highest for two, and IIVS reported the
218 highest IC_{50} value for one substance and the lowest for two.

219 7.2.3 <u>CV Results for the 3T3 and NHK NRU Test Methods</u>

220 CV values were calculated as described in Section **5.5.2.2**. **Tables 7-3** and **7-5** provide the
221 intra- and inter-laboratory CV values for the individual reference substances. **Table 7-7**
222 summarizes the CV values for each method and shows that median and mean values were
223 often similar. Median CV values were frequently lower than the corresponding means, which
224 indicated that large individual CV values skewed the CV distributions.

225 7.2.3.1 Reproducibility of *Intralaboratory CV Values*

226 **Table 7-7** shows that the intralaboratory CV values and mean intralaboratory CV values
227 were the same, 26%, for both NRU test methods. The median intralaboratory CV values were
228 also similar: 23% and 24% for the 3T3 and the NHK NRU test method, respectively. Of the
229 three laboratories, FAL had the highest mean and median CV values and IIVS had the lowest
230 for both methods.
231

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Acetaminophen	526	1.3		13	2.72	0.181	
ECBC	558		15		2.75		NA
FAL	447		19		2.65		NA
IIVS	571		14		2.76		NA
Acetonitrile	10104	1.2		8	4	0.964	
ECBC	10868		72		4.04		NA
FAL	10153		19		4.01		NA
IIVS	9290		4		3.97		NA
Acetylsalicylic acid	613	1.4		15	2.79	0.060	
ECBC	631		3		2.8		NA
FAL	694		14		2.84		NA
IIVS	514		15		2.71		NA
5-Aminosalicylic acid	52.3	2.6		47	1.72	0.044	
ECBC	29.9		22		1.48		NA
FAL	78.2		54		1.89		NA
IIVS	48.8		16		1.69		NA
Aminopterin	682	1.6		27	2.83	0.025	
ECBC	889		20		2.95		NA
FAL	545		8		2.74		NA
IIVS	611		12		2.79		NA
Amitriptyline HCl	9.76	1.4		19	0.99	0.365	
ECBC	10.8		31		1.03		NA
FAL	7.57		72		0.88		NA
IIVS	10.9		10		1.04		NA
Arsenic trioxide	10.4	8.2		91	1.02	<0.001	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	7.77		33		0.89		0.694
FAL	2.55		75		0.41		<0.001
IIVS	20.9		31		1.32		0.0006
Atropine sulfate	91.9	1.3		13	1.96	0.988	
ECBC	85.4		12		1.93		0.8903
FAL	104		85		2.02		0.9069
IIVS	83.2		25		1.92		0.9832
Boric acid	473	1.2		8	2.67	0.931	
ECBC	440		31		2.64		0.9692
FAL	517		73		2.71		0.7391
IIVS	464		2		2.67		0.768
Busulfan	278	1.2		11	2.44	0.659	
ECBC	253		27		2.4		NA
FAL	268		72		2.43		NA
IIVS	313		12		2.5		NA
Cadmium chloride	1.98	1.2		10	0.3	0.733	
ECBC	2.2		37		0.34		NA
FAL	1.88		65		0.27		NA
IIVS	1.86		8		0.27		NA
Caffeine	661	1.4		21	2.82	0.296	
ECBC	817		31		2.91		NA
FAL	591		32		2.77		NA
IIVS	574		1		2.76		NA
Carbamazepine	128	4.0		85	2.11	0.432	
ECBC	66.1		13		1.82		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	253		129		2.4		NA
IIVS	63.9		8		1.81		NA
Carbon tetrachloride	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Chloral hydrate	137	1.4		17	2.14	0.302	
ECBC	140		24		2.15		NA
FAL	159		32		2.2		NA
IIVS	112		2		2.05		NA
Chloramphenicol	366	1.3		13	2.56	0.750	
ECBC	318		45		2.5		NA
FAL	414		44		2.62		NA
IIVS	367		22		2.56		NA
Citric acid	424	1.7		25	2.63	0.006	
ECBC	526		16		2.72		0.009
FAL	312		17		2.49		0.002
IIVS	433		5		2.64		0.483
Colchicine	0.007	1.6		22	-2.16	0.174	
ECBC	0.005		46		-2.28		NA
FAL	0.008		10		-2.12		NA
IIVS	0.008		21		-2.09		NA
Cupric sulfate pentahydrate	197	1.1		4	2.29	0.374	
ECBC	190		10		2.28		NA
FAL	195		6		2.29		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	207		3		2.32		NA
Cycloheximide	0.082	2.3		43	-1.09	0.302	
ECBC	0.053		22		-1.28		NA
FAL	0.12		78		-0.92		NA
IIVS	0.071		19		-1.15		NA
Dibutyl phthalate	32.6	2.2		41	1.51	0.408	
ECBC	28.3		27		1.45		NA
FAL	47.4		73		1.68		NA
IIVS	22		6		1.34		NA
Dichlorvos	11.1	1.4		20	1.05	0.181	
ECBC	8.56		27		0.93		NA
FAL	12.4		30		1.09		NA
IIVS	12.2		3		1.09		NA
Diethyl phthalate	145	2.6		44	2.16	0.049	
ECBC	174		8		2.24		NA
FAL	71.5		94		1.85		NA
IIVS	189		18		2.28		NA
Digoxin	0.00314	107.6		88	-2.5	<0.001	
ECBC	0.00538		13		-2.27		<0.001
FAL	0.00005		36		-4.29		<0.001
IIVS	0.00398		7		-2.4		<0.001
Dimethylformamide	7856	1.5		19	3.9	<0.001	
ECBC	9353		2		3.97		<0.001
FAL	7817		1		3.89		0.508
IIVS	6397		3		3.81		<0.001

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Diquat dibromide monohydrate	4.73	1.9		37	0.67	0.217	
ECBC	3.59		23		0.56		NA
FAL	6.77		55		0.83		NA
IIVS	3.84		8		0.58		NA
Disulfoton	378	5.8		99	2.58	<0.001	
ECBC	140		19		2.15		0.002
FAL	808		26		2.91		<0.001
IIVS	186		32		2.27		0.018
Endosulfan	2.35	2.4		43	0.37	0.029	
ECBC	3.44		17		0.54		NA
FAL	1.42		50		0.15		NA
IIVS	2.19		20		0.34		NA
Epinephrine bitartrate	90.6	1.5		24	1.96	0.119	
ECBC	115		9		2.06		NA
FAL	81.7		35		1.91		NA
IIVS	75		16		1.88		NA
Ethanol	10184	1.4		18	4.01	0.035	
ECBC	8290		5		3.92		NA
FAL	12013		19		4.08		NA
IIVS	10250		9		4.01		NA
Ethylene glycol	42600	1.3		15	4.63	0.063	
ECBC	38000		12		4.58		NA
FAL	49800		9		4.7		NA
IIVS	40000		13		4.6		NA
Fenpropathrin	2.6	2.0		39	0.41	0.031	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	3.73		27		0.57		NA
FAL	2.23		28		0.35		NA
IIVS	1.82		17		0.26		NA
Gibberellic Acid	2866	1.0		2	3.46	0.862	
ECBC	2850		14		3.45		NA
FAL	2940		9		3.47		NA
IIVS	2807		4		3.45		NA
Glutethimide	177	1.1		5	2.25	0.968	
ECBC	187		34		2.27		NA
FAL	170		14		2.23		NA
IIVS	176		16		2.24		NA
Glycerol	27108	1.9		31	4.43	0.200	
ECBC	34267		45		4.53		NA
FAL	18023		46		4.26		NA
IIVS	29033		16		4.46		NA
Haloperidol	3.57	1.1		7	0.55	0.935	
ECBC	3.69		27		0.57		NA
FAL	3.72		49		0.57		NA
IIVS	3.29		35		0.52		NA
Hexachlorophene	0.031	2.2		41	-1.5	0.097	
ECBC	0.027		16		-1.57		NA
FAL	0.046		44		-1.34		NA
IIVS	0.021		11		-1.67		NA
Lactic acid	1308	1.0		1	3.12	0.904	
ECBC	1290		4		3.11		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	1320		5		3.12		NA
IIVS	1313		11		3.12		NA
Lindane	19.3	1.5		20	1.29	0.203	
ECBC	19.1		17		1.28		NA
FAL	23.2		31		1.37		NA
IIVS	15.6		15		1.19		NA
Lithium carbonate	477	1.3		13	2.68	0.295	
ECBC	411		29		2.61		NA
FAL	486		20		2.69		NA
IIVS	535		6		2.73		NA
Meprobamate	516	4.7		61	2.71	0.027	
ECBC	761		15		2.88		NA
FAL	163		116		2.21		NA
IIVS	624		14		2.8		NA
Mercury chloride	5.87	1.3		15	0.77	0.120	
ECBC	6.87		15		0.84		NA
FAL	5.4		19		0.73		NA
IIVS	5.35		2		0.73		NA
Methanol	1616	1.9		42	3.21	0.007	
ECBC	NA		NA		NA		NA
FAL	1133		19		3.05		NA
IIVS	2100		11		3.32		NA
Nicotine	113	1.4		17	2.05	0.700	
ECBC	94.3		26		1.97		NA
FAL	134		59		2.13		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	112		25		2.05		NA
Paraquat	66.1	2.0		40	1.82	0.047	
ECBC	48.3		13		1.68		NA
FAL	96.6		39		1.98		NA
IIVS	53.4		10		1.73		NA
Parathion	31.4	1.2		8	1.5	0.845	
ECBC	34		30		1.53		NA
FAL	31.2		38		1.49		NA
IIVS	29		29		1.46		NA
Phenobarbital	478	1.9		39	2.68	0.027	
ECBC	693		26		2.84		NA
FAL	360		27		2.56		NA
IIVS	381		18		2.58		NA
Phenol	77.7	1.6		22	1.89	0.094	
ECBC	59.1		36		1.77		NA
FAL	93.2		6		1.97		NA
IIVS	80.8		6		1.91		NA
Phenylthiourea	346	1.5		19	2.54	0.133	
ECBC	363		16		2.56		NA
FAL	401		21		2.6		NA
IIVS	272		26		2.44		NA
Physostigmine	172	1.5		22	2.24	0.623	
ECBC	164		3		2.21		NA
FAL	213		112		2.33		NA
IIVS	139		6		2.14		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
Potassium chloride	2279	1.3		13	3.36	0.396	
ECBC	2560		17		3.41		NA
FAL	2287		28		3.36		NA
IIVS	1990		8		3.3		NA
Potassium cyanide	45.1	5.3		86	1.65	0.340	
ECBC	29.3		24		1.47		NA
FAL	89		112		1.95		NA
IIVS	16.9		13		1.23		NA
Procainamide HCl	1764	1.4		16	3.25	0.053	
ECBC	1480		14		3.17		NA
FAL	1787		12		3.25		NA
IIVS	2027		11		3.31		NA
2-Propanol	5541	1.7		26	3.74	0.033	
ECBC	5263		11		3.72		NA
FAL	4273		27		3.63		NA
IIVS	7087		7		3.85		NA
Propranolol HCl	36.9	1.5		21	1.57	0.003	
ECBC	38.27		12		1.58		0.325
FAL	43.8		6		1.64		0.006
IIVS	28.6		11		1.46		0.001
Propylparaben	16.8	1.3		16	1.23	0.066	
ECBC	18.1		13		1.26		NA
FAL	18.6		15		1.27		NA
IIVS	13.8		9		1.14		NA
Sodium arsenite	0.532	2.4		44	-0.27	0.061	

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
ECBC	0.79		32		-0.1		NA
FAL	0.336		56		-0.47		NA
IIVS	0.47		14		-0.33		NA
Sodium chloride	2724	3.2		51	3.44	0.045	
ECBC	3583		7		3.55		NA
FAL	1118		124		3.05		NA
IIVS	3470		9		3.54		NA
Sodium dichromate dihydrate	0.737	1.5		19	-0.13	0.258	
ECBC	0.784		14		-0.11		NA
FAL	0.851		36		-0.07		NA
IIVS	0.576		17		-0.24		NA
Sodium fluoride	47.4	1.4		15	1.68	0.313	
ECBC	48.7		14		1.69		NA
FAL	39.7		24		1.6		NA
IIVS	53.7		13		1.73		NA
Sodium hypochlorite	1580	1.5		20	3.2	0.313	
ECBC	1863		31		3.27		NA
FAL	1243		46		3.09		NA
IIVS	1633		11		3.21		NA
Sodium oxalate	355	1.0		1	2.55	0.926	
ECBC	355		15		2.55		NA
FAL	350		42		2.54		NA
IIVS	360		26		2.56		NA
Sodium selenate	11.2	2.2		40	1.05	0.134	
ECBC	7.47		12		0.87		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
FAL	16.1		59		1.21		NA
IIVS	10		13		1		NA
Strychnine	69.3	1.9		39	1.84	0.364	
ECBC	100		76		2		NA
FAL	52.5		53		1.72		NA
IIVS	55.1		6		1.74		NA
Thallium Sulfate	0.16	1.6		23	-0.8	0.405	
ECBC	0.198		51		-0.7		NA
FAL	0.153		20		-0.82		NA
IIVS	0.127		16		-0.9		NA
Trichloroacetic acid	427	1.6		24	2.63	0.134	
ECBC	348		18		2.54		NA
FAL	541		28		2.73		NA
IIVS	394		13		2.6		NA
1,1,1-Trichloroethane	NA	NA		NA	NA	NA	
ECBC	8137		7		3.91		NA
FAL	NA		NA		NA		NA
IIVS	NA		NA		NA		NA
Triethylenemelamine	1.95	1.3		12	0.29	0.562	
ECBC	1.69		57		0.23		NA
FAL	2.03		23		0.31		NA
IIVS	2.13		23		0.33		NA
Triphenyltin hydroxide	0.013	3.0		55	-1.89	0.088	
ECBC	0.021		32		-1.68		NA
FAL	0.007		106		-2.15		NA

Table 7-5 Reproducibility of the IC_{50} Values from the NHK NRU Test Method

Reference Substance/Laboratory	Arithmetic Mean IC_{50} (mg/mL)[1]	Maximum: Minimum Mean IC_{50} Ratio[2]	Arithmetic IntraLab %CV	Arithmetic InterLab %CV	Log Arithmetic Mean IC_{50} (mg/mL)[1]	ANOVA P[3]	Contrast P[4]
IIVS	0.011		32		-1.96		NA
Valproic acid	533	1.6		28	2.73	0.081	
ECBC	468		25		2.67		0.331
FAL	702		23		2.85		0.032
IIVS	430		17		2.63		0.135
Verapamil HCl	68.7	1.3		14	1.84	0.624	
ECBC	60.5		22		1.78		NA
FAL	79.4		42		1.9		NA
IIVS	66.2		8		1.82		NA
Xylene	NA	NA		NA	NA	NA	
ECBC	NA		NA		NA		NA
FAL	NA		NA		NA		NA
IIVS	486		38		2.69		NA

Abbreviations: NHK=Normal human keratinocytes; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.

[1]Results reported on the same row with chemical names are the means of all the laboratories. Results reported on the same row as laboratories are the laboratory means.

[2]Maximum laboratory mean IC_{50} divided by minimum laboratory mean IC_{50}.

[3]$p < 0.01$ indicated statistical significance.

[4]Contrasts were performed if ANOVA was significant ($p < 0.01$) to determine which laboratory was different from the other two laboratories. Significant contrasts were denoted by $p < 0.01$. No contrast tests were performed if only two laboratories reported IC_{50} values.

Table 7-6 Reference Substances with Significant ANOVA Differences Among Laboratories for NHK NRU Test Method

Reference Substance	Significant Contrast Results[1]			Solubility/ Volatility[2]
	ECBC	FAL	IIVS	
Arsenic trioxide		L	H	Precipitate
Citric acid	H	L		Precipitate
Digoxin	H	L		
Dimethylformamide	H		L	
Disulfoton	L	H		Precipitate
Propranolol HCl		H	L	

Abbreviations: ANOVA=Analysis of variance; NHK=Normal human keratinocytes; H=Laboratory reported the highest mean IC_{50}; L=Laboratory reported the lowest mean IC_{50}; X=Laboratory reported a mean IC_{50} between the values of the other two laboratories; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
[1]Laboratories significantly different from the other two at $p < 0.01$
[2]From **Table 5-11**. Precipitate reported by at least one laboratory.

7.2.3.2 *Reproducibility of Interlaboratory CV*

The mean and median interlaboratory CV for the reference substances were lower in the NHK NRU test method (mean=28%; median=21% vs. mean=47%; median=37% for 3T3 (see **Table 7-7**).

Table 7-7 Summary of CV Results for the 3T3 and NHK NRU Test Methods

CV	3T3 NRU Test Method				NHK NRU Test Method			
	N	Mean	Median	Range	N	Mean	Median	Range
Intralaboratory CV	198	26%	23%	1-122%	204	26%	24%	1-129%
ECBC	66	23%	17%	2-95%	68	23%	20%	2-76%
FAL	66	33%	31%	1-98%	68	43%	34%	1-129%
IIVS	66	21%	14%	1-122%	68	13%	13%	1-35%
Interlaboratory CV	66	47%	37%	3-135%	68	28%	21%	1-91%

Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; N=number of values; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
Note: For the 3T3 method, the following substances were excluded because all laboratories did not obtain sufficient IC_{50} data: carbon tetrachloride; disulfoton; gibberellic acid; lithium carbonate; methanol; 1,1,1-trichloroethane; valproic acid; and xylene. For the NHK method, the following substances were excluded because all laboratories did not obtain sufficient IC_{50} data: carbon tetrachloride; methanol; 1,1,1-trichloroethane; and xylene.

7.2.3.3 *Variation of CV with Chemical Property*

To identify chemical characteristics that may be associated with high or low CV values, their associations were assessed for chemical class along with the following chemical attributes: physical state (i.e., solid or liquid), solubility, volatility, molecular weight, log K_{ow}, IC_{50}, and boiling point. The CVs were also examined with respect to their association with the GHS acute oral toxicity class (UN 2005). For categorical characteristics such as physical form, solubility (i.e., precipitate/no precipitate), volatile/not volatile, and chemical class, the mean CV values and ranges for the groups were compared to one another and to the overall mean CV and CV range for each method. No statistical analyses were performed for these comparisons. Spearman correlation analyses were performed for chemical characteristics measured by continuous variables, such as molecular weight, log K_{ow}, and IC_{50}, and boiling point.

7.2.3.4 *Results of Intralaboratory CV Analysis*

The intralaboratory CV analysis (see **Table 7-8**) uses one mean intralaboratory CV for each reference substance that was calculated from the intralaboratory CV values from each laboratory. There seemed to be little difference in CV values among the categorical physical/chemical/toxicological attributes. The mean intralaboratory CV values for solids and liquids were similar (26 vs. 23% for 3T3; 27 vs. 24% for NHK). The mean intralaboratory CV values for reference substances for which precipitates were observed were similar to values for substances with no precipitates were observed (32 vs. 23% for 3T3; 24 vs. 27% for NHK). The mean intralaboratory CV values for substances that exhibited volatility were relatively similar to those that did not (31 vs. 25% for 3T3; 27 vs. 26% for NHK). Similarly, the substances grouped by GHS acute oral toxicity category (UN 2005) had mean intralaboratory CV values that were similar (20-33% for 3T3; 19-31% for NHK) to the overall mean CV values (26% for both test methods). However, the mean intralaboratory CV values for both NRU test methods tended to increase with decreasing LD_{50}.

Mean intralaboratory CV values were calculated for the chemical classes that contained at least three of the reference substances included in the reproducibility analyses (i.e., 64 substances for 3T3 and 68 substances for NHK). Reference substances in the amide chemical class had unusually low mean intralaboratory CV values for both the 3T3 (13%) and the

NHK (10%) NRU test method compared with the overall mean CV (26% for both test methods), but there were only three substances in this chemical class (acetaminophen, dimethylformamide, procainamide HCl). Organic sulfur compounds had a high mean intralaboratory CV for the 3T3 test method (46%), but not for the NHK NRU test method (29%) compared with the overall mean intralaboratory CV for both test methods (26%). The intralaboratory CV values for the remaining chemical classes were unremarkable compared with the overall mean intralaboratory CV values.

For the characteristics amenable to correlation analysis, none of the Spearman correlation coefficients were large (absolute value of r_s <0.6), but several were statistically significantly different from zero ($p < 0.05$). Molecular weight ($p=0.016$), IC_{50} ($p=0.002$), and boiling point ($p=0.009$) exhibited statistically significant correlations to intralaboratory CV for the 3T3 test NRU method. The higher molecular weight substances had higher intralaboratory CV values and the substances with lower IC_{50} values had higher intralaboratory CV values. The finding that substances with higher boiling points had higher CV values was consistent with the categorical analysis of volatility. The substances that exhibited volatile characteristics (i.e., cross contamination of VC wells) in the 3T3 NRU test method had slightly higher mean intralaboratory CV values (31%) than the substances that did not exhibit volatile characteristics (25%).

Among the IC_{50} values obtained using the NHK NRU test method, two of the characteristics amenable to correlation analysis were statistically significantly different from zero, although the correlation coefficients did not have large magnitudes (absolute value of r_s <0.4). The log K_{ow} ($p=0.032$) and IC_{50} ($p=0.004$) exhibited statistically significant correlations ($p < 0.05$) to the intralaboratory CV. Log K_{ow} was positively correlated (i.e., higher log K_{ow} values were associated with higher mean intralaboratory CV), but the IC_{50} was negatively correlated (i.e., higher log IC_{50} values were associated with lower mean intralaboratory CV) to mean intralaboratory CV.

Table 7-8 Intralaboratory CV by Chemical Characteristics for the 3T3 and NHK NRU Test Methods

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N[1]	Range	Mean	N[1]	Range	Mean
All chemicals	64	1-122%	26%	68	1-129%	26%
Chemical form						
Solid	51	4-84	26	53	6-57	27
Liquid	13	6-48	23	15	2-40	24
Solubility						
Precipitate[2]	18	11-84	32	19	2-47	24
No precipitate	46	4-55	23	49	7-57	27
Volatility[3]						
Volatile	10	6-84	31	9	11-50	27
Nonvolatile	54	4-55	25	59[2]	2-57	26
Chemical Class						
Alcohol	9	6-42	22	9	10-37	22
Amide	3	4-28	13	3	2-16	10
Amine	3	9-35	18	3	10-24	18
Carboxylic acid	13	4-41	18	14	2-48	23
Heterocyclic	14	6-59	31	14	13-50	32
Organophosphorous	2	NA	NA	3	20-32	26
Organic sulfur	4	36-59	46	5	21-27	29
Phenol	5	14-30	20	5	11-31	19
Polycyclic	4	19-35	27	5	9-38	20
Inorganic	14	9-43	25	15	6-50	29
Inorganic chlorine	5	9-33	19	5	12-50	32
Inorganic sodium	6	9-34	20	6	17-47	30
GHS Acute Oral Toxicity Class						
$LD_{50} \leq 5$ mg/kg	6	9-46	27	7	20-40	30
$5 < LD_{50} \leq 50$	12	13-59	32	12	12-50	31
$50 < LD_{50} \leq 300$	12	11-84	33	12	17-37	25
$300 < LD_{50} \leq 2000$	14	4-51	22	16	6-57	25
$2000 < LD_{50} \leq 5000$	9	9-32	20	9	7-50	30
$LD_{50} > 5000$	11	6-42	20	12	2-40	19
Correlations	**N**	**r_s**	**P value**	**N**	**r_s**	**P value**
Molecular weight	64	0.301	0.016	68	0.181	0.140
Log K_{ow}	45[4]	0.121	0.430	48[4]	0.310	0.032
IC_{50}	64	-0.382	0.002	68	-0.346	0.004
Boiling point	24[5]	0.520	0.009	24[5]	0.226	0.289

Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NA=Not applicable because class had less than three observations; NHK=Normal human keratinocytes; N=Number of values; r_s=Spearman correlation coefficient; K_{ow}=Octanol:water partition coefficient.

[1]One intralaboratory CV for each chemical was calculated by averaging the CV values for each reference substance.

[2]Identified by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see **Table 5-11**).

[3]Identified by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-11**).

[4]Number of reference substances with CV values and log K_{ow} data.

[5]Number of reference substances with CV values and boiling point data.

343 7.2.3.5 *Results of the Interlaboratory CV Analysis*

344 **Table 7-9** shows the analysis of the interlaboratory CV values. There seemed to be little
345 difference in interlaboratory CV values for most of the categorical physical/chemical
346 characteristics. The mean interlaboratory CV values for solids and liquids were similar (48%
347 for solids vs. 42% for liquids for 3T3, and 28% for solids vs. 21% for liquids for NHK), as
348 were the values for substances for which precipitates were observed versus no precipitates
349 (58% vs. 43% for 3T3, and 24% vs. 28% for NHK), and the values for substances that
350 exhibited volatile characteristics (51% for volatile substances vs. 46% for nonvolatile
351 substances for 3T3, and 32% for volatile substances vs. 26% for nonvolatile substances for
352 NHK).
353
354 Mean interlaboratory CV values were calculated for the chemical classes that contained at
355 least three of the reference substances included in the reproducibility analyses (i.e., 64
356 substances for 3T3 and 68 substances for NHK). Reference substances in the amide chemical
357 class had low mean interlaboratory CV values for both the 3T3 (15%) and the NHK (16%)
358 NRU test methods compared with the overall mean interlaboratory CV (47% and 28%,
359 respectively). Substances in the amine class also had low mean interlaboratory CV values for
360 the 3T3 NRU (13%), but not for the NHK NRU (20%). Organic sulfur compounds had
361 unusually high mean interlaboratory CV values for the 3T3 test method (100%), but not for
362 the NHK NRU (36%) compared with the overall mean interlaboratory CV (47% and 28%,
363 respectively). Because of the low number of reference substances in these classes, these
364 results were deemed to not be significant.
365

365 **Table 7-9 Interlaboratory 3T3 and NHK NRU Test Method CV Values Sorted by**
366 **Chemical Characteristics**
367

Class/Attribute	3T3 NRU Test Method			NHK NRU Test Method		
	N	Range	Mean	N	Range	Mean
All chemicals	64[1]	3-135%	47%	68[1]	1-91%	28%
Chemical Form						
Solids	51	3-135	48	53	1-91	28
Liquids	13	6-124	42	15	1-44	21
Solubility						
Precipitate[2]	18	7-127	58	19	1-91	24
No precipitate	46	3-135	43	49	1-88	28
Volatility						
Volatile[3]	10	21-127	51	9	8-86	32
Nonvolatile	54	3-135	46	59	1-91	26
Chemical Class						
Alcohol	9	12-119	38	9	11-31	20
Amide	3	6-28	15	3	13-19	16
Amine	3	10-16	13	3	14-24	20
Carboxylic acid	13	6-124	38	14	1-61	26
Heterocyclic	14	8-135	57	14	5-85	32
Organic sulfur	4	78-119	100	5	8-99	36
Organophosphorous	2	NA	NA	3	8-99	42
Phenol	5	19-64	41	5	15-47	28
Polycyclic	4	14-85	44	5	2-88	30
Inorganic	14	3-127	50	15	4-91	30
Inorganic chlorine	5	3-127	45	5	10-86	35
Inorganic sodium	6	3-60	34	6	15-51	32
GHS Acute Oral Toxicity Class						
$LD_{50} \leq 5$ mg/kg	6	12-135	72	7	12-99	37
$5 < LD_{50} \leq 50$	12	33-127	78	12	8-91	41
$50 < LD_{50} \leq 300$	12	8-120	37	12	10-41	26
$300 < LD_{50} \leq 2000$	14	11-85	35	16	1-61	20
$2000 < LD_{50} \leq 5000$	9	3-69	29	9	1-85	27
$LD_{50} > 5000$	11	6-124	41	12	2-44	23
Correlations	**N**	**r_s**	**P value**	**N**	**r_s**	**P value**
Molecular weight	64	0.245	0.051	68	0.169	0.168
Log K_{ow}	45[4]	0.151	0.324	48[4]	0.210	0.151
IC_{50}	64	-0.304	0.015	68	-0.297	0.014
Boiling point	22[5]	0.563	0.007	25[5]	-0.051	0.809

368 Abbreviations: CV=Coefficient of variation; 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of
369 Classification and Labelling of Chemicals (UN 2005); NA=Not applicable because class had less than three observations;
370 NHK=Normal human keratinocytes; NRU=Neutral red uptake; N=Number of values; r_s=Spearman correlation coefficient;
371 K_{ow}=Octanol:water partition coefficient.
372 [1]One intralaboratory CV for each chemical was calculated by averaging the CV values for each reference substance.
373 [2]Identified by laboratory reports of precipitate in the stock reference substance solutions or in cell culture (see **Table 5-11**).
374 [3]Identified by laboratory reports of using plate sealers to avoid contamination of the VC wells (see **Table 5-11**).
375 [4]Number of reference substances with CV values and log K_{ow} data.
376 [5]Number of reference substances with CV values and boiling point data.
377

378

Mean interlaboratory CV values tended to be large for chemicals in the most toxic GHS acute oral toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for reference substances in the LD_{50} ≤5 mg/kg (72%) and 5 < LD_{50} ≤50 mg/kg (78%) classes were larger than the mean overall interlaboratory CV (47%,). For the NHK NRU test method, the mean interlaboratory CV for chemicals in the 5 < LD_{50} ≤5 mg/kg (37%) and 5 < LD_{50} ≤50 mg/kg (41%) classes were larger than the mean overall interlaboratory CV (28%).

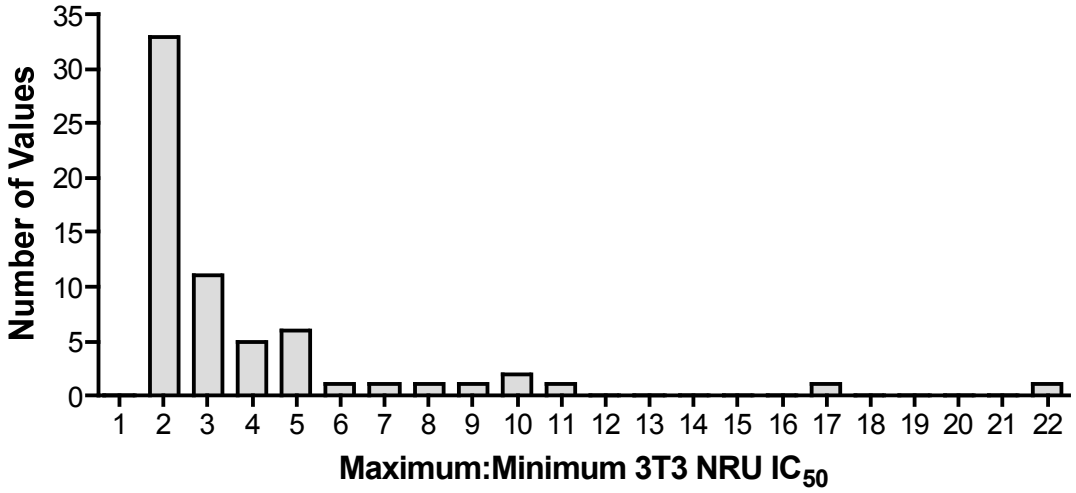
For the characteristics amenable to correlation analysis, none of the correlation coefficients were large (absolute value of r_s <0.6), but IC_{50} (p=0.015) and boiling point (p=0.007) exhibited statistically significant correlations (p <0.05) to interlaboratory CV in the 3T3 test NRU method. There was a negative correlation between interlaboratory CV and IC_{50}, but the correlation between boiling point and interlaboratory CV was positive. The positive correlation of CV with boiling point was largely consistent with the categorical analysis of volatility. The substances that exhibited volatile characteristics in the 3T3 NRU test method had slightly higher mean CV values than substances that did not exhibit volatile characteristics (51% vs. 46%). Only the IC_{50} was significantly correlated (p=0.014) to interlaboratory CV with a negative correlation (r_s=-0.271) when the NHK NRU test method was used.

7.2.4 Comparison of Maximum to Minimum IC_{50} Values Using Laboratory Means

Interlaboratory reproducibility was also compared by calculating maximum to minimum mean IC_{50} values using the laboratory means from each method, so that the reproducibility of the IC_{50} values could be compared with the reproducibility of the reference LD_{50} values derived in **Section 4.2**. The **Figure 7-2** frequency histogram for the 3T3 NRU test method maximum:minimum mean IC_{50} values shows that approximately half of the reference substances produced ratios within 1.5 - 2.5-fold of each other, and only nine chemicals had ratios greater than 5.5 fold, including one substance (cupric sulfate pentahydrate) that had a ratio of 22.

408 **Figure 7-2 Frequency of Maximum:Minimum 3T3 NRU IC_{50} Ratios**

409



410 Bars show the number of substances with maximum:minimum 3T3 NRU IC_{50} ratios within 0.5 units of the bar
411 label (e.g., the first bar indicates that there were 33 reference substances for which the laboratory mean
412 maximum:minimum 3T3 NRU IC_{50} ratios were 1.5 to2.5). The analysis includes 64 reference substances.
413 Carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic
414 acid, and xylene were excluded because not all laboratories obtained IC_{50} values.

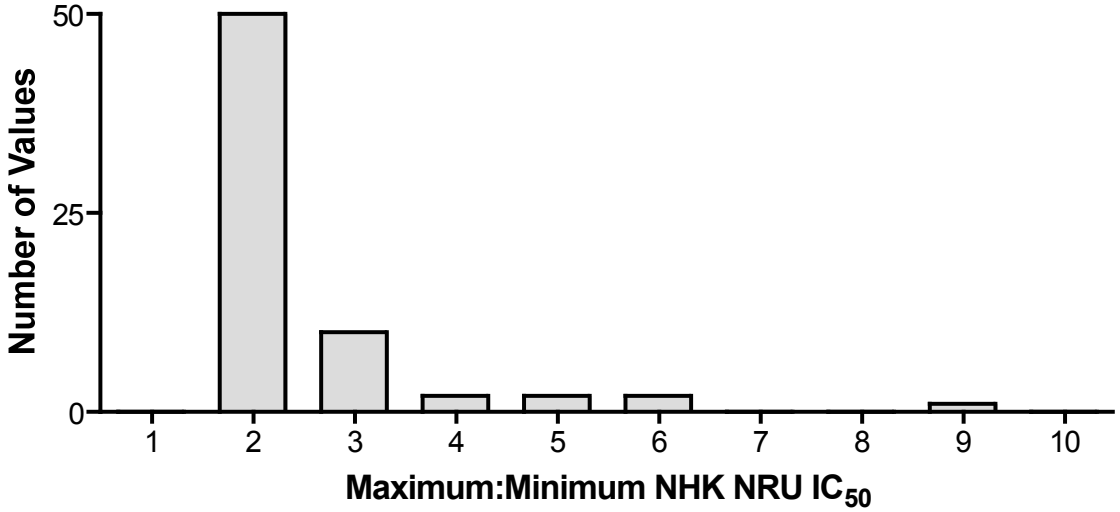
415

416 The **Figure 7-3** frequency histogram for the maximum:minimum mean IC_{50} values for the
417 NHK NRU test method shows that ratios of 60 of the 68 chemicals were within two- to three-
418 fold of one another. The highest ratio of 108 for digoxin is not shown in the figure.
419 Comparison of **Figures 7-2** and **7-3** shows that the interlaboratory reproducibility of the
420 NHK NRU test method was better than that for the 3T3 NRU test method based on the
421 distribution of the low maximum:minimum IC_{50} ratios.

422

423

Figure 7-3 Frequency of Maximum:Minimum NHK NRU IC_{50} Ratios



Bars show the number of substances with maximum:minimum NHK NRU IC_{50} ratios with 0.5 units of the bar label (e.g., the first bar indicates that there were 50 reference substances for which the laboratory mean maximim:minimum NHK NRU IC_{50} ratios were 1.5 to2.5). The analysis includes 68 reference substances. Carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene were excluded because not all laboratories obtained IC_{50} values. The maximum:minimum IC_{50} for digoxin of 108 was excluded from this figure.

7.2.5 Comparison of the Maximum:Minimum IC_{50} Ratios with the Maximum:Minimum LD_{50} Ratios

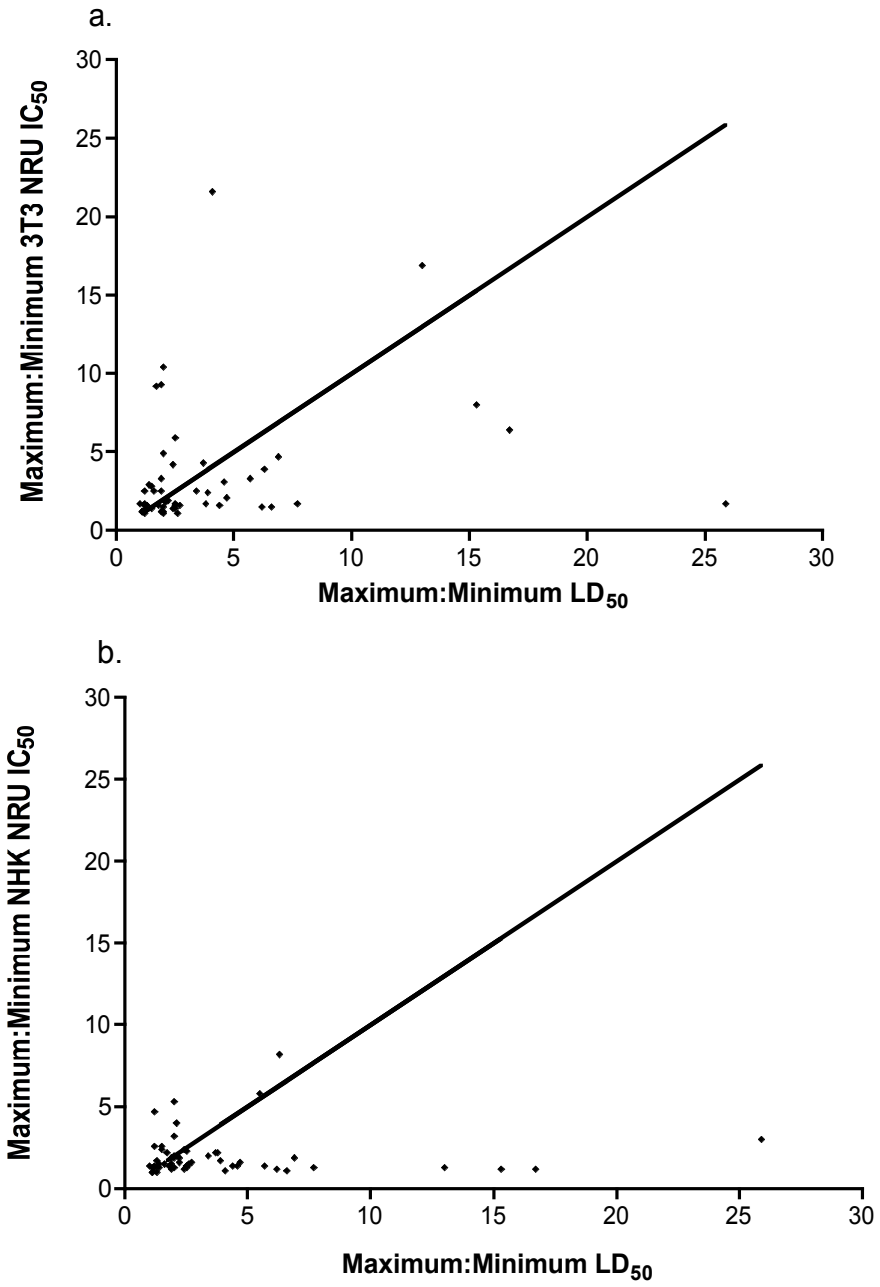
To compare the reproducibility of the NRU IC_{50} values with that of the LD_{50} values, the maximum:minimum IC_{50} ratios for each method (shown in **Tables 7-3** and **7-5**) were compared with the maximum:minimum LD_{50} ratios reported in **Table 4-2**. This analysis excluded reference substances for which fewer than three laboratories reported IC_{50} values, and reference substances for which fewer than two acceptable acute oral LD_{50} values were identified. As a result, there were 53 substances analysed for the 3T3 NRU test method and 57 for the NHK NRU test method. The following substances were excluded from both analyses because fewer than two acceptable LD_{50} values could be identified: aminopterin; colchicine; digoxin; epinephrine bitartrate; gluthethimide; phenylthiourea; physostigmine; procainamide HCl, propranolol HCl; propylparaben; and thallium sulfate. Carbon tetrachloride, disulfoton, gibberellic acid, lithium carbonate, methanol, 1,1,1-trichloroethane, valproic acid, and xylene, were excluded from the 3T3 analysis, and carbon tetrachloride, methanol, 1,1,1-trichloroethane, and xylene, were excluded from the NHK analysis, because fewer than three laboratories reported IC_{50} values.

446 **Figure 7-4** shows that the maximum:minimum LD_{50} ratios tend to be larger than either the
447 3T3 NRU IC_{50} or NHK NRU IC_{50} ratios because there are more points below the theoretical
448 one-to-one correspondence line than above the line. The difference between the LD_{50}
449 maximum:minimum values and the NRU IC_{50} maximum:minimum values is more striking
450 for the NHK since there are fewer points above the line for the NHK graph (**Figure 7-4b**)
451 than for the 3T3 graph (**Figure 7-4a**).

452 7.2.6 Normalization of Reference Substance IC_{50} Values Using SLS IC_{50} Values

453 As an alternate analysis for reproducibility, IC_{50} values for reference substances were
454 normalized to those of the corresponding SLS IC_{50}values. This approach was tested using
455 five reference substances for each test method to determine whether such normalization
456 would reduce the variability, measured using CV values, of the results. The reference
457 substances selected for this evaluation were those for which the ANOVA indicated
458 statistically significant differences among the laboratories. Because there were a number of
459 reference substances that met this criterion for the 3T3 NRU test method, the substances
460 were selected so as to cover a wide range of rodent acute oral toxicity. One reference
461 substance was selected from each GHS category with the exception of the $50 \leq LD_{50} < 300$
462 mg/kg category. There was no substance represented by this category because there were six
463 acute oral toxicity categories and only five substances were used for this assessment. The
464 reference substances, shown in **Table 7-10**, were busulfan, chloramphenicol, meprobamate,
465 propylparaben, and triethylenemelamine. Because there were only six reference substances
466 with significant ANOVAs in the NHK NRU test method, the last five reference substances in
467 **Table 7-5** (citric acid, digoxin, dimethylformamide, disulfoton, and propranolol HCl) were
468 selected for this analysis.
469

Figure 7-4 Comparison of Maximum:Minimum NRU IC_{50} Ratios to Maximum:Minimum LD_{50} Ratios



Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NRU=Neutral red uptake; NHK=Normal human keratinocytes. Comparison of maximum:minimum ratios of IC_{50} and LD_{50} for 53 reference substances for the 3T3 NRU test method (a) and 57 reference substances for the NHK NRU test method (b). Solid lines show the theoretical one to one correspondence of maximum:minimum IC_{50} to maximum:minimum LD_{50}.

477

Table 7-10 CV Values for 3T3 and NHK NRU Test Method IC_{50} Values and Normalized IC_{50} Values

Reference Substance	IC_{50} (mM)[1]	IntraLab CV[2] (%)	InterLab CV[3] (%)	SLS IC_{50}: Substance IC_{50}[4]	IntraLab CV SLS IC_{50}: Substance IC_{50}[5] (%)	InterLab CV SLS IC_{50}: Substance IC_{50}[6] (%)
3T3 NRU Test Method						
Busulfan	0.548		119	0.677		74
ECBC	0.163	48		1.05	70	
FAL	1.30	56		0.109	53	
IIVS	0.177	4		0.877	9	
Chloramphenicol	0.498		67	0.725		29
ECBC	0.171	22		0.847	30	
FAL	0.845	30		0.844	22	
IIVS	0.483	18		0.483	21	
Meprobamate	2.47		54	0.071		39
ECBC	1.62	14		0.085	23	
FAL	4.02	15		0.039	29	
IIVS	1.77	2		0.088	3	
Propylparaben	0.166		64	1.16		49
ECBC	0.116	16		1.29	20	
FAL	0.287	29		0.535	22	
IIVS	0.0949	12		1.65	9	
Triethylene-melamine	0.00278		135	191		87
ECBC	0.000421	11		354	11	
FAL	0.00710	18		21.4	24	
IIVS	0.000827	29		197	23	
Mean		**22**	**88**		**25**	**65**
NHK NRU Test Method						
Citric Acid	2.21		25	0.00587		26
ECBC	2.74	16		0.0053	14	
FAL	1.62	17		0.0076	28	
IIVS	2.25	5		0.0047	16	
Digoxin	4.02E-06		88	62378		168
ECBC	6.89E-06	13		1264	10	
FAL	6.53E-08	36		183479	44	
IIVS	5.10E-06	7		2389	26	
Dimethylform-amide	107		19	0.00011		31
ECBC	128	2		0.00007	7	
FAL	107	1		0.00013	1	
IIVS	87.5	3		0.00013	19	
Disulfoton	1.38		99	0.0140		61
ECBC	0.509	19		0.022	6	
FAL	2.94	26		0.005	5	
IIVS	0.679	32		0.015	20	
Propranolol HCl	0.125		21	0.0947		20
ECBC	0.129	12		0.081	15	

Table 7-10 CV Values for 3T3 and NHK NRU Test Method IC_{50} Values and Normalized IC_{50} Values

Reference Substance	IC_{50} (mM)[1]	IntraLab CV[2] (%)	InterLab CV[3] (%)	SLS IC_{50}: Substance IC_{50}[4]	IntraLab CV SLS IC_{50}: Substance IC_{50}[5] (%)	InterLab CV SLS IC_{50}: Substance IC_{50}[6] (%)
FAL	0.148	6		0.087	25	
IIVS	0.0967	11		0.116	9	
Mean		**14**	**50**		**16**	**61**

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NA=No acceptable IC_{50} results reported or calculation was not performed (e.g., for contrast results); CV=Coefficient of variation.
[1]Results reported on the same row with reference substance names are the arithmetic means of all the laboratories. Results reported on the same row as laboratories are the arithmetic laboratory means.
[2]CV for IC_{50} values from the acceptable tests within each laboratory.
[3]CV calculated using the arithmetic mean IC_{50} values from each laboratory.
[4]Concurrent SLS IC_{50} in mM divided by the reference substance IC_{50}. Results reported on the same row with reference substance names are the arithmetic means of all the laboratories. Results reported on the same row as laboratories are the arithmetic laboratory means.
[5]CV for SLS IC_{50}:reference substance IC_{50} values within each laboratory.
[3]CV calculated using the mean SLS IC_{50}:reference substance IC_{50} values from each laboratory.

Millimolar units were used for the IC_{50} values in this analysis since the mole is the most appropriate unit for measuring and comparing biological activity. The IC_{50} value (in mM) for each reference substance was normalized to the corresponding SLS IC_{50} value (in mM) by dividing the SLS IC_{50} by the reference substance IC_{50}. Intra- and inter-laboratory CV values were calculated for both the IC_{50} values and for the SLS IC_{50}:reference substance IC_{50} ratios to determine whether this type of normalization would reduce the interlaboratory CV values.

Table 7-10 shows that the mean intralaboratory CV of the IC_{50} values for the five substances used in the 3T3 evaluation was 22% and the interlaboratory CV was 88%. Normalizing the reference substance IC_{50} values to the SLS IC_{50} yielded a slightly higher intralaboratory CV of 25% and a lower interlaboratory CV of 65%. The mean intralaboratory CV of the IC_{50} values for the five substances used in the NHK evaluation was 14% and the interlaboratory CV was 50%. Normalizing the reference substance IC_{50} values to the SLS IC_{50} yielded a slightly higher intralaboratory CV of 16% and a higher interlaboratory CV of 61%. When the normalization ratios are examined for each chemical-by-laboratory combination (**Table 7-10**), nine CVs increased, five decreased, and one remained the same for the 3T3 NRU test method, and eight increased, six decreased, and one remained the same for the NHK NRU

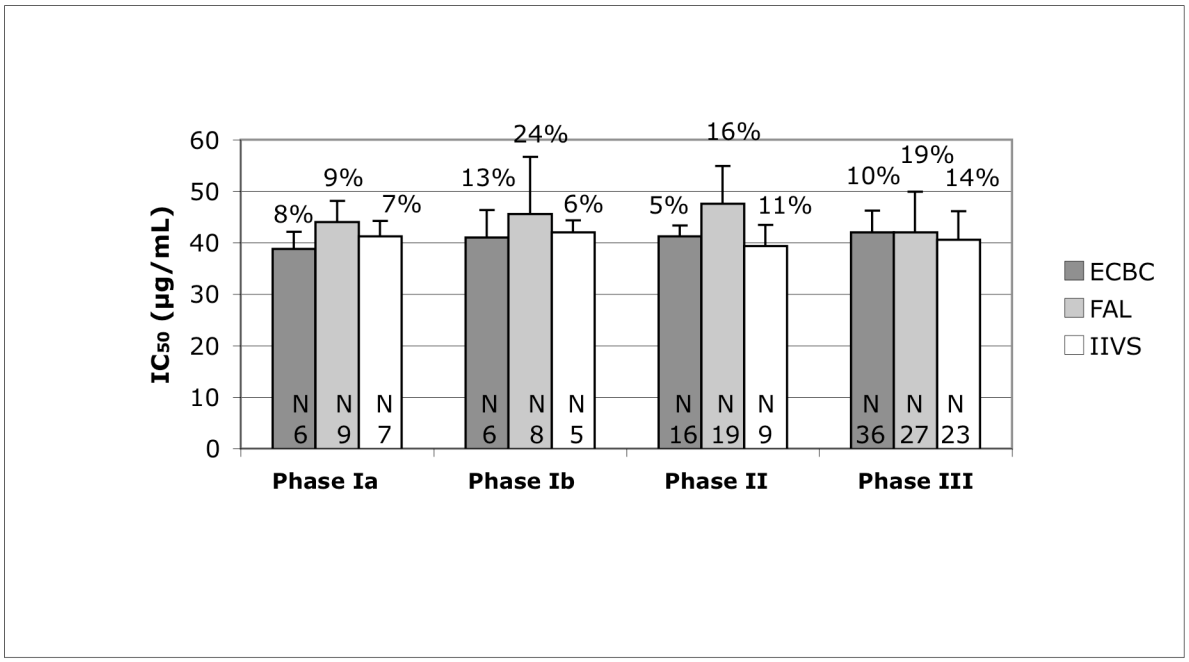
509 test method. Thus, for the reference substances used in this analysis, normalizing the
510 reference substance IC_{50} to the concurrent SLS IC_{50} did not reduce the overall variability of
511 the measurements, as measured by the CV values.

512 **7.3 Historical Positive Control Data**

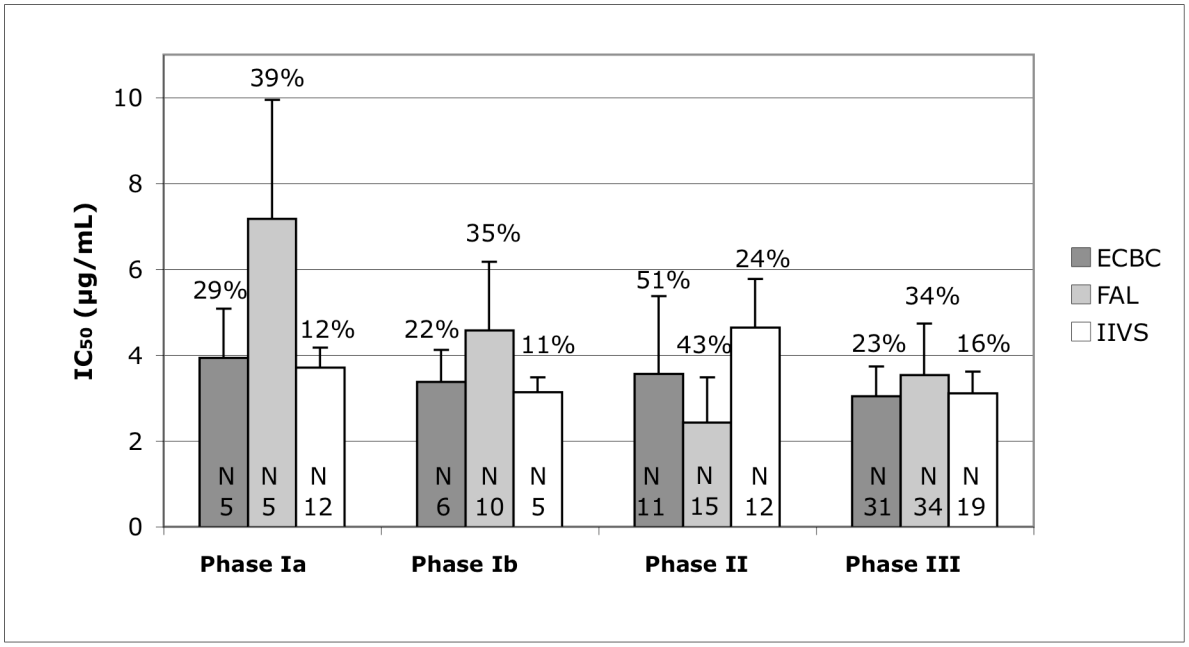
513 The reproducibility of the positive control (SLS) data was assessed by CV analysis,
514 ANOVA, and linear regression over time, as described in **Section 5.5.4.2**. To obtain an
515 assessment of the true variation of SLS IC_{50} values, the reproducibility analyses also included
516 IC_{50} values from SLS tests that failed the test acceptance criterion for the IC_{50} acceptance
517 limits determined for each study phase. Therefore, the values used for this analysis included
518 some that were not included in **Table 5-3**. These additional SLS tests, however, passed all
519 other test acceptance criteria. If more than one SLS test was performed in a single day (for
520 each method and laboratory), the IC_{50} values were averaged to determine a single IC_{50} for the
521 day so that the multiple results from that day would not overly influence the average.
522
523 **Figure 7-5** shows the average SLS IC_{50} values for each method, laboratory, and study phase.
524 The SLS IC_{50} for the 3T3 test method (**Figure 7-5a**) was relatively consistent over the entire
525 period of the study (approximately 2.5 years). The intralaboratory CV values for the
526 individual study phases ranged from 5% to 24% **(Figure 7-5a)**. With the exception of the
527 Phase Ib CV at FAL, the CV values for each laboratory and phase were less than 20%. The
528 interlaboratory CV values were even smaller, 6% in Phases Ia and Ib, 10% in Phase II, and
529 2% in Phase III.
530
531 **Figure 7-5b** shows that the SLS IC_{50} for the NHK NRU test method tended to vary with
532 time, but, with the exception of the values from FAL, there appeared to be no consistent
533 trend. The IC_{50} values from FAL, which changed their cell culture methods after Phase Ib
534 (see **Section 5.3.3.1**), tended to decrease over time. Although the change in cell culture
535 methods reduced the magnitude of the IC_{50}, the variability (as evidenced by the
536 intralaboratory CV values shown in **Figure 7-5b**) remained relatively high (CV ≥34% for all
537 FAL study phases).
538
539

Figure 7-5 SLS IC_{50} for Each Laboratory and Study Phase

a 3T3 Test Method



b NHK Test Method



Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; N=Number of values.
Note: Bars show mean SLS IC_{50} values. Error bars show standard deviation. Percent values above error bars are intralaboratory CVs.

The CV values for all the laboratories and study phases show that the SLS IC_{50} values in the NHK NRU test method are more variable within laboratories than the corresponding 3T3 SLS IC_{50} values. The CV values for the SLS IC_{50} for the NHK NRU test method ranged from 11 to 51%, with nine of the 12 values greater than 20%. The interlaboratory CV values, which were also greater than those for the 3T3 NRU test method, were 39% in Phase Ia, 21% in Phase Ib, 31% in Phase II, and 8% in Phase III.

7.3.1 ANOVA and Linear Regression Results for the 3T3 NRU Test Method

7.3.1.1 *Variation of SLS IC_{50} Values with Time*

Table 7-11 shows the SLS ANOVA results from the 3T3 test method. When the IC_{50} values in each laboratory were compared, there were no statistically significant differences ($p < 0.01$) among study phases for any laboratory. **Table 7-12** shows that the slopes of the linear regressions of the IC_{50} values over time (expressed as index values) were significantly different from zero for ECBC and FAL ($p=0.001$ and 0.012, respectively), but, because the slopes were so small (0.000204 and -0.000324), and in different directions, these differences were considered to be unimportant, regardless of the statistical conclusions. The slope of the IIVS regression of SLS IC_{50} over time was not significantly different from zero ($p=0.651$; **Table 7-12**), which was consistent with the ANOVA analysis (**Table 7-11**), and showed that SLS IC_{50} from IIVS did not vary with study phase ($p=0.854$). The ANOVA analysis, with study phase as the factor (with laboratories combined), showed that the 3T3 NRU IC_{50} values from all the laboratories were consistent over time ($p=0.304$).

7.3.1.2 *Comparison of SLS IC_{50} Values Among the Laboratories*

When all study phases from each laboratory were combined, ANOVA, with laboratory as the factor, showed that the SLS IC_{50} values in the 3T3 NRU test method differed significantly among the laboratories ($p < 0.006$) **(Table 7-11)**. However, as can be seen in **Figure 7-5**, the individual laboratory CVs are all in the same range and overlap.

578 **Table 7-11 ANOVA Results for the SLS IC_{50} Values in the 3T3 NRU Test Method**
579

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-0.876	0.042	6	0.031	-0.811	0.046	9	0.015	-0.850	0.034	7	0.854
Phase Ib	-0.864	0.066	6		-0.846	0.065	8		-0.838	0.025	5	
Phase II	-0.848	0.027	16		-0.796	0.057	19		-0.854	0.025	8	
Phase III	-0.842	0.036	36		-0.851	0.066	27		-0.844	0.041	23	
Test for differences between laboratories (phases combined)												
All Phases	-0.849	0.039	64	0.006	-0.826	0.062	63		-0.847	0.035	44	
Test for differences between phases (laboratories combined)												
Phase Ia	-0.839	0.049	22	0.304								
Phase Ib	-0.850	0.056	19									
Phase II	-0.831	0.047	34									
Phase III	0.845	0.045	86									

580 Abbreviations: ANOVA=Analysis of variance; SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; N=Number of
581 values; SD=Standard deviation; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals
582 in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
583 [1]Statistically significant at $p < 0.01$.
584
585

586 **Table 7-12 Linear Regression Analysis of SLS IC_{50} Values Over Time[1]**
587

Laboratory	Slope	P-value (Slope)[2]	Intercept
3T3 NRU Test Method			
ECBC	0.000204	0.001	-0.874
FAL	-0.000324	0.012	-0.796
IIVS	0.0000304	0.651	-0.850
NHK NRU Test Method			
ECBC	-0.000559	0.002	-1.901
FAL	-0.00112	<0.001	-1.737
IIVS	-0.000445	0.002	-1.885

588 Abbreviations: SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; N=Number
589 of values; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical
590 Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
591 [1]Time was expressed as index values. The index value of each test reflected the order of testing without respect to the time
592 lapsing between tests.
593 [2]Statistically significant from zero at $p < 0.05$.
594

595 7.3.2 ANOVA and Linear Regression Results for the NHK NRU Test Method

596 7.3.2.1 *Variation of SLS IC_{50} Values with Time*

597 **Table 7-13** shows the ANOVA results for the NHK NRU test method. When the IC_{50} values
598 within each laboratory were compared by study phase, the values were statistically different
599 ($p < 0.01$) at each laboratory. The IC_{50} values from the various study phases were also
600 significantly different from one another when the laboratory data were combined ($p < 0.001$).
601 The change in cell culture methods at FAL after Phase Ib (see **Section 5.3.3.1**) was attributed
602 to this difference. **Table 7-13** shows that FAL had clearly the lowest log mean SLS IC_{50} for
603 Phases Ia and Ib. Linear regression analyses showed that the IC_{50} slopes over time (expressed
604 as an index values) were statistically significantly less than zero for each laboratory (see
605 **Table 7-12**). Because the slopes were so small (-0.000559, -0.00112, and -0.000445), and
606 negative, their statistical significance was considered to be irrelevant.

607 7.3.2.2 *Comparison of SLS IC_{50} Among the Laboratories*

608 The ANOVA results, with laboratory as a factor (**Table 7-13)** showed that the SLS IC_{50} was
609 statistically significantly different among the laboratories when the data from the study
610 phases were pooled ($p < 0.001$). **Figure 7-5** shows that the SLS data from ECBC and IIVS
611 were rather similar for Phases Ia, Ib, and III. The SLS IC_{50} data from FAL are different from
612 the other two laboratories for Phases Ia, Ib, and II, but the SDs for Phase III show that the
613 data from all laboratories produced similar values.

614 **Table 7-13 ANOVA Results for the SLS IC_{50} Values in the NHK NRU Test Method**

Study Phase/ Laboratory	ECBC				FAL				IIVS			
	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]	Log Mean IC_{50} (mM)	SD	N	P[1]
Test for differences between phases within each laboratory												
Phase Ia	-1.867	0.135	5	0.001	-1.656	0.125	5	<0.001	-1.904	0.060	12	<0.001
Phase Ib	-1.936	0.092	6		-1.829	0.141	10		-1.965	0.046	5	
Phase II	-2.007	0.109	11		-1.982	0.173	15		-1.863	0.058	12	
Phase III	-1.990	0.098	31		-1.941	0.113	34		-1.972	0.070	19	
Test for differences between laboratories (phases combined)												
All Phases	-1.971	0.113	53	<0.001	-1.879	0.175	64		-1.924	0.073	48	
Test for differences between phases (laboratories combined)												
Phase Ia	-1.833	0.143	22	<0.001								
Phase Ib	-1.891	0.125	21									
Phase II	-1.964	0.139	38									
Phase III	-1.971	0.100	84									

615 Abbreviations: ANOVA=Analysis of variance; SLS=Sodium lauryl sulfate; 3T3=BALB/c 3T3 fibroblasts; N=Number of
616 values; SD=Standard deviation; ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals
617 in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.
618 [1]Statistically significant at $p < 0.01$.
619

620 **7.4 Laboratory Concordance for Solvent Selection**

621 The solvents used for the reference substances are shown in **Table 7-14**. For Phases Ib and II,
622 the SMT based their selection of solvents on the results provided by BioReliance (see **Table**
623 **5-9**) using the solubility protocol in **Appendix G2**. Despite the fact that the solubility of an
624 individual substance might be different in 3T3 and NHK growth media, the SMT selected the
625 same solvent (i.e., medium or DMSO) for both test methods, rather than having different
626 solvents for each method.
627
628 BioReliance occasionally achieved higher solubility values for the Phase I and II substances
629 than the three laboratories (e.g., see the results for arsenic trioxide, aminopterin, and
630 chloramphenicol in **Table 5-9**). The laboratories were using the solubility protocols in
631 **Appendices C3** through **C6** (for Phases Ib and II), which were somewhat different from the
632 protocol used by BioReliance. Although all the laboratories used the same protocols, they did
633 not always obtain similar results with respect to the solvent to be used (e.g., see the results
634 for aminopterin, cadmium chloride, and chloramphenicol in **Table 5-9**). In an attempt to
635 avoid the selection of a solvent for which one or more laboratories could not achieve the
636 desired solubility, the SMT used the solubility data from all the laboratories to determine the
637 solvents to be used for each chemical tested in Phase III. **Table 7-14** shows that cell culture
638 medium was used as the solvent for 38 substances and DMSO was used for 34 substances.
639
640 Five of the substances were insoluble in medium and DMSO in at least one testing
641 laboratory. Arsenic trioxide was insoluble at all laboratories. IIVS also found sodium oxalate,
642 strychnine, and triethylenemelamine insoluble in media and DMSO, and FAL found thallium
643 sulfate insoluble in both solvents. Therefore, the SMT used the results from the laboratories
644 that did achieve solubility to select the solvents to be used for testing these substances.
645
646

646

Table 7-14 Solvent Determinations by Laboratory

Reference Substance	Solvent Used for Testing[1]	ECBC	FAL	IIVS
Acetaminophen	DMSO	Medium	Medium	DMSO
Acetonitrile	Medium	Medium	Medium	Medium
Acetylsalicylic acid	DMSO	Medium	DMSO	Medium
Aminopterin	DMSO	DMSO	DMSO	DMSO
5-Aminosalicylic acid	Medium	Medium	Medium	Medium
Amitriptyline HCl	DMSO	DMSO	DMSO	DMSO
Arsenic III trioxide	Medium	ID	ID	ID
Atropine sulfate	Medium	Medium	Medium	Medium
Boric aid	Medium	Medium	Medium	Medium
Busulfan	DMSO	DMSO	DMSO	DMSO
Cadmium II chloride	DMSO	DMSO	DMSO	DMSO
Caffeine	Medium	Medium	Medium	Medium
Carbamazepine	DMSO	Medium	DMSO	DMSO
Carbon tetrachloride	DMSO	Medium	DMSO	Medium
Chloral hydrate	Medium	Medium	Medium	Medium
Chloramphenicol	DMSO	DMSO	DMSO	Medium
Citric acid	Medium	Medium	Medium	Medium
Colchicine	Medium	Medium	Medium	Medium
Cupric sulfate pentahydrate	Medium	Medium	Medium	Medium
Cycloheximide	Medium	Medium	Medium	Medium
Dibutyl phthalate	DMSO	DMSO	DMSO	DMSO
Dichlorvos	DMSO	Medium	DMSO	Medium
Diethyl phthalate	DMSO	DMSO	DMSO	DMSO
Digoxin	DMSO	DMSO	DMSO	DMSO
Dimethylformamide	Medium	Medium	Medium	Medium
Diquat dibromide monohydrate	Medium	Medium	Medium	Medium
Disulfoton	DMSO	DMSO	DMSO	DMSO
Endosulfan	DMSO	DMSO	DMSO	DMSO
Epinephrine bitartrate	Medium	Medium	Medium	Medium
Ethanol	Medium	Medium	Medium	Medium
Ethylene glycol	Medium	Medium	Medium	Medium
Fenpropathrin	DMSO	DMSO	DMSO	DMSO
Gibberellic acid	Medium	Medium	Medium	Medium
Glutethimide	DMSO	DMSO	DMSO	DMSO
Glycerol	Medium	Medium	Medium	Medium
Haloperidol	DMSO	DMSO	DMSO	DMSO
Hexachlorophene	DMSO	DMSO	DMSO	DMSO
Lactic acid	Medium	Medium	Medium	Medium
Lindane	DMSO	DMSO	DMSO	DMSO
Lithium I carbonate	Medium	Medium	Medium	Medium
Meprobamate	DMSO	Medium	Medium	DMSO
Mercury II chloride	DMSO	DMSO	DMSO	DMSO
Methanol	DMSO	Medium	Medium	DMSO
Nicotine	Medium	Medium	Medium	Medium
Paraquat	Medium	Medium	Medium	Medium
Parathion	DMSO	DMSO	DMSO	DMSO
Phenobarbital	DMSO	Medium	DMSO	DMSO
Phenol	Medium	Medium	Medium	Medium

Table 7-14 Solvent Determinations by Laboratory

Reference Substance	Solvent Used for Testing[1]	ECBC	FAL	IIVS
Phenylthiourea	DMSO	DMSO	Medium	DMSO
Physostigmine	DMSO	Medium	DMSO	DMSO
Potassium I chloride	Medium	Medium	Medium	Medium
Potassium cyanide	Medium	Medium	Medium	Medium
Procainamide HCl	Medium	Medium	Medium	Medium
2-Propanol	Medium	Medium	Medium	Medium
Propranolol HCl	DMSO	Medium	Medium	Medium
Propylparaben	DMSO	DMSO	DMSO	DMSO
Sodium arsenite	Medium	Medium	Medium	Medium
Sodium chloride	Medium	Medium	Medium	Medium
Sodium dichromate dihydrate	Medium	Medium	Medium	Medium
Sodium fluoride	Medium	Medium	Medium	Medium
Sodium hypochlorite	Medium	Medium	Medium	Medium
Sodium oxalate	Medium	Medium	Medium	ID
Sodium selenate	Medium	Medium	Medium	Medium
Strychnine	Medium	Medium	Medium	ID
Thallium I sulfate	Medium	Medium	ID	Medium
Trichloroacetic acid	Medium	Medium	Medium	Medium
1,1,1-Trichloroethane	Medium	Medium	Medium	Medium
Triethylenemelamine	DMSO	Medium	DMSO	ID
Triphenyltin hydroxide	DMSO	DMSO	DMSO	DMSO
Valproic acid	DMSO	Medium	DMSO	DMSO
Verapamil HCl	DMSO	DMSO	DMSO	DMSO
Xylene	DMSO	DMSO	DMSO	DMSO
DMSO Total	34	22	29	28
Medium Total	38	49	41	40

Abbreviations: DMSO=Dimethyl sulfoxide; ECBC=Edgewood Chemical Biological Center; FAL= Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; ID=Insufficient data to select solvent; Medium=Cell culture medium.
[1]Solvents selected by the SMT for use by all laboratories.

The testing laboratories selected the same solvent for 55 of the 72 reference substances (76%). Excluding the five substances that were found to be insoluble in both solvents by at least one laboratory, there were 12 substances on which the laboratories disagreed: acetaminophen, acetylsalicylic acid, carbamazepine, carbon tetrachloride, chloramphenicol, dichlorvos, meprobamate, methanol, phenobarbital, phenylthiourea, physostigmine, and valproic acid. Each laboratory reported relatively low solubility, ≤2 mg/mL, in medium for these substances. Because 2 mg/mL in medium is the departure point for the selection of medium or DMSO, small variations in solubility lead the laboratories to select different solvents. The solubility of acetaminophen, for example was reported as 2 mg/mL in culture media by ECBC and FAL, but <2 mg/mL by IIVS. IIVS found it soluble in 200 mg/mL

663 DMSO and selected DMSO as the solvent. ECBC and FAL selected culture media as the
664 solvent. The SMT selected DMSO as the solvent for acetaminophen to be used by all
665 laboratories so that they would all be assured of obtaining usable test results.

666 **7.5 Summary**

667 Intra- and inter-laboratory reproducibility were assessed by comparing the laboratory-
668 specific IC_{50}-LD_{50} regressions to the mean, across-laboratory regression for each method,
669 ANOVA, CV analysis, and comparison of maximum:minimum mean laboratory IC_{50} values.
670 ANOVA permitted statistical comparisons of laboratories and experimental averages, while
671 controlling for other factors. CV analysis compared the relative magnitudes of variability on
672 a standardized scale. Intralaboratory reproducibility was evaluated using the results from the
673 reference substances that yielded IC_{50} values from all three laboratories: 64 and 68 reference
674 substances in the 3T3 and the NHK NRU test methods, respectively. The analysis of
675 interlaboratory reproducibility, by evaluating the similarity of the laboratory specific IC_{50}-
676 LD_{50} regressions, showed that the laboratory regressions for both NRU test methods were
677 within the 95% confidence limits of the laboratory mean regressions.
678
679 The ANOVA showed significant interlaboratory differences for 23 substances in the 3T3
680 NRU test method and six in the NHK NRU test method. Intralaboratory CV values ranged
681 from 1-122% in the 3T3 test method and 1-129% in the NHK NRU test method. Mean
682 interlaboratory CV values were 26% for both NRU test methods, but NHK had a lower mean
683 interlaboratory CV (28% vs 47%). Interlaboratory CV values ranged from 3-135% in the 3T3
684 NRU test method and 1-91% in the NHK NRU test method. FAL had the highest mean
685 intralaboratory CV in both NRU test methods (33% in 3T3, 43% in NHK).
686
687 An analysis to determine the relationship between the chemical attributes and interlaboratory
688 CV indicated that chemical structure, physical form, solubility, and volatility had little effect
689 on CV. The CV seemed to be related, however, to GHS acute toxicity category, IC_{50}, and
690 boiling point. Mean interlaboratory CV values were larger for substances in the most toxic
691 GHS categories than for substances in the other toxicity categories, especially with the 3T3
692 NRU test method. The mean interlaboratory CV for substances in the $LD_{50} \leq 5$ mg/kg (72%)
693 and $5 < LD_{50} \leq 50$ mg/kg (78%) classes were larger than the mean overall interlaboratory CV

694 (47%) with the 3T3 NRU test method. The mean interlaboratory NHK CV was 37% for
695 substances with LD_{50} ≤5 mg/kg, and 41% for substances with 5< LD_{50} ≤50 mg/kg, while the
696 mean overall interlaboratory CV was 28%. A Spearman correlation analysis showed that the
697 IC_{50} was inversely correlated to interlaboratory CV for both the 3T3 (p=0.015) and NHK
698 (p=0.014) test methods, and that boiling point was positively correlated to interlaboratory CV
699 (p=0.007) (i.e., higher boiling points were associated with higher CV values) for the 3T3 but
700 not the NHK NRU test method (p=0.809).
701
702 The ANOVA results for the positive control IC_{50} in the 3T3 NRU test method showed that
703 there were significant differences among laboratories (p=0.006) but not among study phases
704 within laboratories (p >0.01). However, interlaboratory CV values, which ranged from 2% to
705 10% for the different study phases, were similar across laboratories and the intralaboratory
706 CV values ranged from 5% to 24%. The SLS IC_{50} values from the NHK NRU test method
707 were more variable than those from the 3T3 NRU test method. The ANOVA results for SLS
708 in the NHK NRU test method indicated that there were significant differences among
709 laboratories (p <0.001) and among study phases within laboratories (p ≤0.001). A change in
710 cell culture methods at FAL after Phase Ib decreased the SLS IC_{50} in subsequent phases, but
711 FAL's CV values still tended to be higher than in the other laboratories. Intralaboratory CV
712 values for the NHK SLS IC_{50} during the various study phases ranged from 11% to 51% and
713 interlaboratory CV values for SLS in the NHK NRU test method ranged from 8% to 39%.
714
715 Cell culture medium was used as the solvent for 38 substances and DMSO was used for 34
716 substances. Concordance among all three laboratories in selecting the solvent for the
717 reference substances was 76% (55/72).

8.0 3T3 AND NHK NRU TEST METHOD DATA QUALITY 8-3

8.1 Compliance with Good Laboratory Practice Regulations 8-3
8.1.1 Guidelines Followed for Cytotoxicity Testing 8-3
8.1.2 Quality Assurance (QA) for NRU Cytotoxicity Test Data 8-5
8.1.3 Guidelines Followed for Rodent Acute Oral LD_{50} Data Collection 8-7

8.2 Results of Data Quality Audits.....8-8
8.2.1 QA Statements 8-8
8.2.2 QA Statements from the Laboratories.....8-8

8.3 Effect of Deviations or Non-compliance with GLPs.....8-11
8.3.1 Laboratory Error Rates.....8-12
8.3.2 Failure Rates for Definitive and PC Tests.....8-13
8.3.3 Intralaboratory Reproducibility 8-15
8.3.4 Prediction of Globally Harmonized System (GHS) Acute Oral Toxicity Categories 8-15

8.4 Availability of Laboratory Notebooks 8-16

8.5 Summary.....8-17

[This Page Intentionally Left Blank]

8.0 3T3 AND NHK NRU TEST METHOD DATA QUALITY

This section of the BRD presents the extent of adherence to GLP regulations for generation of the validation study data. Data quality is described, along with deviations from the regulations and their effect (if any) on the quality of the data. Statistical analyses are provided to compare the data generation, collection, and reporting by the two GLP compliant laboratories and the one non-GLP compliant laboratory, as well as for the GLP-compliant laboratory that distributed the reference substances and performed solubility studies. Discussions of various quality assurance aspects of the study are included.

8.1 Compliance With Good Laboratory Practice Regulations

8.1.1 Guidelines Followed for Cytotoxicity Testing

8.1.1.1 *Good Laboratory Practices*

The SOW provided the following definition of U.S. Regulatory agency GLPs to each laboratory:

> "Regulations governing the conduct, procedures, and operations of toxicology laboratories; regulations to assure the quality and integrity of the data and to address such matters as organization and personnel, facilities, equipment, facility operations, test and control articles, and validation study protocol, and conduct (U.S. Food and Drug Administration, Title 21 CFR Part 58; U.S. Environmental Protection Agency, Title 40 CFR Part 160)."

IIVS, ECBC, and BioReliance performed testing under GLP guidelines. The details of GLP compliance and training are addressed in **Section 11**.

8.1.1.2 *Spirit of GLP*

The SMT determined a definition for "spirit of GLP" and provided the following to the laboratories:

> "Laboratories that are non GLP-compliant shall adhere to GLP principles and other method parameters as put forth in this Statement of Work and the Test Method Protocols (provided by NIEHS/NICEATM); documentation and accountability shall be equal to

GLP requirements; laboratories must make assurances that they are equal in performance criteria and that there is parity amongst the laboratories."

FAL performed testing in the "spirit of GLP" (see **Section 11.2.2**) by following the international GLP standards referenced in the ECVAM Workshop 37 Report (Cooper-Hannan 1999) and the OECD Principles of GLP (OECD 1998). The laboratory did not have their data and test procedures reviewed by an independent, quality assurance (QA) auditor. The SOW directed FAL to, at a minimum, routinely document their equipment monitoring and record keeping (see **Table 8-1**), and to archive all documents. The FAL already had most of the requested procedures and guidelines in place for routine laboratory procedures before initiation of this study. The various general laboratory-related activities were documented in workbooks and logbooks, and the information was made available to the SMT.

Table 8-1 SMT-Recommended Documentation for FAL Laboratory

Daily	Per Use	Periodic
Temperatures Laboratory (ambient), incubators, water baths, refrigerators, freezers	Cryogenic Storage Unit Liquid N_2 volume	Laboratory Supplies[1] Lot numbers and expiration dates for stock media formulations and components, NRU reagents, tissue culture plasticware
Humidity/CO_2 Cell culture incubators	Equipment Calibration Balances, pH meters, cell counters	Cells Quantity, and cryogenic storage conditions, for 3T3 and NHK cells
Visual Observations Cell Culture Growth	Reagents Lot numbers and expiration dates of medium/supplements	Equipment Calibration Incubators, laminar flow hoods, autoclaves, micropipettors, spectrophotometer plate readers, computers (software)

Abbreviations: SMT=Study Management Team; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; NRU=Neutral Red Uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes.
[1]Documentation for laboratory supplies begins when supplies are purchased and received by the laboratory

8.1.1.3 *Good Cell Culture Practices (GCCP)*

The SMT provided guidance in the SOW for implementing GLPs in a cell culture laboratory environment. The initial assumption by the SMT was that each laboratory had the basic cell

culture skills and knowledge (e.g., as described in Freshney 2000) to reliably perform the *in vitro* NRU cytotoxicity test methods. Reviews of historical laboratory documents, and scientific and professional exchanges with the laboratory personnel, assured the SMT that each laboratory had demonstrated, through previous validation studies and other experience, that the personnel were capable of providing quality scientific data through the use of good cell culture practices. A comparison of the SOW and the *in vitro* NRU cytotoxicity protocols showed that the guidelines developed for the NICEATM/ECVAM study were harmonious with the guidelines in the ECVAM Good Cell Culture Practices Reports (Hartung 2002; Coecke et al. 2005), and the OECD document on GLPs and *in vitro* studies (OECD 2004a).

8.1.2 Quality Assurance (QA) for NRU Cytotoxicity Test Data

8.1.2.1 *Coded Reference Substances*

BioReliance acquired 73 high purity chemicals (72 reference substances and one positive control substance) from reputable commercial sources. Sixty-four of the reference substances were ≥99% pure, and seven were between 90 and 99% pure. Lactic acid had the lowest purity, 89% (See **Appendix F1**). The substances were coded with unique identification numbers and provided to the testing laboratories in a blinded fashion. Procurement of chemicals and their preparation for distribution was performed under GLP guidelines and the SOW provided by the SMT (see **Appendix G**). **Section 3.4** provides detailed information on the acquisition and distribution of reference substances.

8.1.2.2 *Solubility Testing and Data Review*

All laboratories performed solubility tests on all reference substances using the solvents and procedures specified in the protocols provided by the SMT, and submitted solubility data to the SMT in the form of hard copy printouts and electronic worksheets. The Study Directors reviewed all laboratory procedures and all data produced at their respective laboratories, and the QA designee in each GLP-compliant laboratory reviewed all data in the their laboratory. The SMT Project Coordinators served as informal QA reviewers for FAL (i.e., reviewed all the raw data sheets). The errors and omissions detected were reported to FAL, and corrections were requested. The SMT reviewed all solubility data and NRU assay data produced by all of the laboratories.

The SMT reviews of the submitted data in Phases Ia and Ib revealed that, even after data review by the Study Directors, the data files contained an unacceptably high frequency of errors (see **Section 2.6.2.5**). The laboratories were alerted to the problem and personnel from all laboratories attended a weeklong training session at the IIVS laboratories in Gaithersburg, Maryland to enhance harmonization among the laboratories. Errors continued to be found in data files submitted for Phase III after the training, albeit less frequently; however, such errors generally resulted from the rush to rapidly complete the data files for submission to the SMT shortly after the conclusion of each test. The formal QA reviews of the files occurred later in each phase of the study.

The most common errors included typographical mistakes, transcriptional and data entry errors in the Microsoft® EXCEL® and the GraphPad PRISM® 3.0 templates, and incorrect labeling of files. The SMT reviewed every electronic file and hard copy printout throughout the study and alerted the Study Directors of the affected laboratories when errors were found. All data files were checked for consistency within the documents, and for compliance with the protocols. The SMT also documented errors on the hard copy printouts in the form of handwritten notations to the files (at NICEATM) and added these notations to the electronic data summary files compiled for data management. Files that were revised and/or corrected by the Study Director were resubmitted to the SMT and identified as corrected files.

8.1.2.3 *NRU Cytotoxicity Test Tallies*

The Study Directors periodically received individualized test tallies specific to their laboratories from NICEATM that detailed:

- The number of range finder tests performed by the laboratory
- The number of definitive tests performed, and the pass/fail status of each
- The number of positive control tests performed, and the pass/fail status of each
- The number of acceptable tests completed
- The test completion status for each chemical (i.e., whether one range finder test had been completed, and the number of acceptable definitive tests had been completed)

151 The laboratories compared the NICEATM tallies to their own records to verify their
152 consistency and accuracy. Discrepancies were resolved through direct communication
153 between the Study Director and the SMT.

154 8.1.3 Guidelines Followed for Rodent Acute Oral LD_{50} Data Collection

155 For the purposes of this validation study, the *in vitro* NRU test methods were proposed for
156 predicting starting doses for rodent acute oral toxicity test methods, rather than as
157 replacement tests for the *in vivo* test method. No *in vivo* tests were performed for this
158 validation study. All *in vivo* data (i.e., rat and mouse LD_{50} values) were collected by
159 NICEATM through reviews of the literature and from publicly available databases. All
160 relevant data and pertinent information were gathered and stored in an Excel® spreadsheet.

161 8.1.3.1 *Rodent Acute Oral LD_{50} Values Used in the Registry of Cytotoxicity (RC)*

162 The RC is a database of acute oral LD_{50} values for rats and mice obtained primarily from the
163 1983/84 RTECS® database compiled by NIOSH, and IC_{50} values from *in vitro* cytotoxicity
164 assays using multiple cell lines and cytotoxicity endpoints for chemicals with known
165 molecular weights (Halle 1998, 2003). Collection and reporting methods used for generating
166 the data in RTECS® were not a part of the data collection hierarchy employed by NIOSH,
167 and the data in this database were not evaluated for quality and accuracy. Many of the values
168 come from secondary sources with no citation to the original report. GLP guidelines were not
169 used to determine acceptable data for the database. The only criterion used by NIOSH for
170 reporting acute oral toxicity data in RTECS® was that the LD_{50} value was the most toxic
171 LD_{50} value for a chemical that could be found in the literature, regardless of the number of
172 other values available, or their distribution.

173 8.1.3.2 *Rodent Acute Oral LD_{50} Values Collected by NICEATM from Other Sources*

174 One critical aspect of the validation study design was the establishment of a rat acute oral
175 LD_{50} reference value for each of the 72 reference substances (see **Section 4**). These reference
176 values were used to evaluate the extent to which the two *in vitro* NRU test methods could
177 predict rat acute oral LD_{50} values. Primary rat acute oral LD_{50} studies were located through
178 searching electronic databases, published articles, and secondary references. Rat data were
179 not available for three of the reference substances and mouse acute oral LD_{50} values were

180 used. Only seven of the 455 LD_{50} values collected from the literature were produced under
181 GLP guidelines.

182 8.2 Results of Data Quality Audits

183 The QA unit or designee in each GLP laboratory provided a systematic and critical
184 comparison of the data provided in the laboratory's study reports to the raw data in the
185 laboratory records. The SOW provided to each laboratory contained the following guidance
186 regarding QA statements:

187 "The Final Reports for all phases of the Validation Study shall be audited by the Quality
188 Assurance unit of the Testing Facility for GLP compliance and a QA Statement shall be
189 provided by the Testing Facility. Each Final Report shall identify: 1) the phases and data
190 inspected, 2) dates of inspection, and 3) dates findings were reported to the Study
191 Director and Testing Facility management. The QA Statement shall identify whether the
192 methods and results described in the Final Report accurately reflect the raw data
193 produced during the Validation Study."

194 8.2.1 QA Statements

195 The QA statements from the GLP-compliant laboratories addressed the reviews of:

196 - Protocols
197 - Laboratory standard operating procedures (SOPs)
198 - Laboratory operations, in general
199 - 3T3 and NHK NRU experiment data
200 - The report submitted to the SMT
201

202 The QA statements from the GLP laboratories affirm that the methods described in the
203 protocols are the methods that the laboratory personnel used, and that the data reported to the
204 SMT accurately reflect the raw data obtained by the laboratory. See **Section 8.2.2** for
205 information about the QA statements for the non-GLP laboratory.

206 8.2.2 QA Statements from the Laboratories

207 8.2.2.1 *BioReliance QA Statements*

208 The Study Director/Laboratory Director provided the following statement in all of the final
209 reports:

"The solubility studies, acquisition, preparation, and distribution of the test chemicals were conducted in compliance with GLP. Although not audited (per SOW), the work described in this report for Phase X (i.e., Ia, Ib, and II) fully and accurately reflects to the best of my knowledge the raw data generated in the study."

8.2.2.2 *FAL QA Statements*

The Study Director for FAL performed the final review of all data and reports before sending them to the SMT, and provided the following two statements in the final reports provided to the SMT.

- "The laboratory worked under the principles of GLP whilst not being a GLP-compliant laboratory."
- "The report accurately reflects the work undertaken and the results obtained at the FRAME Alternatives Laboratory."

Formal QA statements were not provided to FAL because the SMT performed informal QA reviews.

8.2.2.3 *ECBC QA Statements*

The QA statements reported the particular study phase and laboratory procedures that were examined for GLP compliance. In addition, the laboratory's statement noted that the scope of work, associated protocols, and quality control acceptance criteria were updated or changed during the study, which made the assessment of the procedures and data for conformance to the SOPs more difficult. However, compliance with the requirements and intent of GLP guidelines was continually assessed during the review of the SOPs and the observance of operations. The QA reviews found the ECBC protocols to be in compliance with the NICEATM/ECVAM study protocols. The aspects of the studies inspected by the QA designee were:

- Review of protocols and laboratory SOPs
- Review of waste handling procedures
- Review of laboratory operations
- Certification of new personnel
- Review of data
- Review of the final report for each testing phase

The QA designee also observed the preparation of reference substances, 96-well plate configuration, application of reference substance, annotation to the workbook, and appropriate sterile technique while performing the testing. The number of inspections of laboratory operations was reduced in the latter phases of the study because the same personnel conducted the testing throughout the entire study.

<u>ECBC Review Dates of the Study Phases</u>

- Phase Ia: July 2002 through May 2003
- Phase Ib: July 2002 through January 2003
- Phase II: May 2003 through February 2004
- Phase III: November 2003 through March 2005

8.2.2.4 *IIVS QA Statements*

Because the IIVS QA unit is small, it carried out reviews of different aspects of the procedures at different times. The IIVS QA Statement reads:

> "This study has been divided into a series of in-process phases. Using a random sampling approach, Quality Assurance monitors each of these phases over a series of studies. Procedures, documentation, equipment records, etc., are examined to assure that the study is performed in accordance with the U.S. FDA Good Laboratory Practice regulations (21 CFR 58), the U.S. EPA GLP Standards (40 CFR 792 and 40 CFR 160) and the OECD Principles of Good Laboratory Practice and to assure that the study is conducted according to the protocol and relevant Standard Operating Procedures."

The aspects of the studies inspected by the QA designee were as follows:

- Protocol and initial paperwork
- Reading of the plates (definitive test)
- Dilution of the test articles (definitive test)
- Treatment of the cells
- Termination of treatment and addition of the NR dye (definitive test)
- Cell concentration determination and seeding of the plates (third definitive test)

- Termination of treatment and addition of the NR dye
- Washing the cells
- Draft report and data
- Final report

<u>IIVS Review Dates of Various Aspects of the Test Phases</u>

- Phase Ia: August 2002 Final Report Review: October 2005
- Phase Ib: January 2003 Final Report Review: October 2005
- Phase II: July-August 2003 Final Report Review: October 2005
- Phase III: January-November 2004 Final Report Review: October 2005

8.2.2.5 *Other QA Information*

Data generated by the laboratories and reviewed by their respective Study Directors were submitted to the SMT. Often, the data were provided electronically within days of the end of testing. The SMT was active as a secondary QA reviewer of all information provided by the Study Directors. If the SMT found discrepancies, the Project Coordinators corresponded with the appropriate Study Director to identify and rectify the error. The Study Director made corrections/adjustments to the discrepancies in data reporting and presented the changes to the SMT. The SMT did not initiate any external data quality audits.

The quality of the reference substances was assured in the form of certificates of analysis provided by the chemical manufacturer to BioReliance at the time of purchase. The SMT and the laboratories obtained certificates of analysis from CAMBREX for Clonetics® NHK culture medium and supplements. In addition, the SMT obtained quality control data directly from CAMBREX technical departments concerning the NHK medium's ability to support keratinocyte growth.

8.3 Effect of Deviations or Non-compliance with GLPs

Rates for several types of errors (i.e., documentation, testing methods, and data management) were determined by the SMT. Many of the errors (particularly in Phases Ia and Ib) were the result of minor mistakes (e.g., typographical, mislabeling) and did not affect the quality of the data.

8.3.1 Laboratory Error Rates

The SMT was concerned about the number of errors that were seen in documentation and testing methods during Phases Ia and Ib, and compiled the detected errors from each laboratory. The types of errors found included errors in documentation (e.g., reference substance identification did not match on all associated data sheets; IC_{20} and IC_{80} values were transposed in the EXCEL® template; a test acceptance criterion flag in a data sheet was incorrect) and in testing (e.g., wrong dilution scheme was used for the PC; wrong SLS IC_{50} was used as the PC IC_{50}). Error rates were compiled as the number of tests with errors per total number of tests. As shown in **Table 2-3**, FAL had the highest error rates: 93% for the 3T3 NRU test method and 41% for the NHK NRU test method. The highest error rates in the other laboratories were 10% for the 3T3 NRU test method and 23% for the NHK NRU test method (both ECBC).

There were relatively few errors detected in the Phase III data files. The SMT did not compile the typographical and transcriptional errors found, but reported them directly to the appropriate Study Director so that the data sheets could be immediately corrected. The SMT did not detect errors in the raw optical density data from the 96-well plates provided in each data file. The laboratories and the SMT corrected typographical and transcriptional errors (e.g., incorrect $\log IC_{50}$ value entered) in the EXCEL® templates. The EXCEL® template formulas were used for the statistical analyses.

An assessment of error rates was performed specifically for Phase III for one particular clerical error – the transfer of statistical results (e.g., ICx values) from the GraphPad PRISM® 3.0 template to the Microsoft® EXCEL® template. It was often necessary for the SMT to revise the EXCEL® data files provided by the laboratories because the incorrect values had been transferred to EXCEL®. **Table 8-2** summarizes the Phase III error rates resulting from the transfer of data from PRISM® to EXCEL®.

Table 8-2 Phase III Error Rates in the Transfer of Data to the EXCEL® Template

Laboratory	Number of Errors Detected	Number of Definitive Tests	Percentage of Tests with Detected Errors
ECBC	49	402	12
FAL	171	513	33
IIVS	25	419	6

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences.

8.3.2 Failure Rates for Definitive and PC Tests

Table 8-3 presents the test failure (i.e., did not meet test acceptance criteria) rates experienced in Phase III. Approximately 25% of all 3T3 definitive tests and 18% of all NHK definitive tests failed. If a definitive test (see **Section 2.3.2.2** for the definition of a definitive test) failed, the laboratory repeated the test and attempted to obtain three acceptable definitive tests for each reference substance in each cell type (see **Section 2.5** for criteria for repeating tests). The PC tests failed 0 to 18% of the time with a combined average failure rate of 8% for both cell types. FAL had the highest individual laboratory test failure rates for 3T3 definitive tests (30%), NHK definitive tests (32%), and NHK PC tests (18%). ECBC had the highest failure rate for 3T3 PC tests (11%). IIVS had no PC test failures.

346 **Table 8-3 Definitive Test and Positive Control (PC) Test Failure Rates in Phase III**
347

Test Type	3T3 NRU Test Method				NHK NRU Test Method				Total
	ECBC	FAL	IIVS	Total	ECBC	FAL	IIVS	Total	
Definitive Tests - Acceptable	169	177	176	522	173	175	174	522	1044
Definitive Tests - Total	215	257	225	697	187	256	194	637	1334
% Failed Definitive Tests	21	30	22	25	8	32	10	18	22
PC Tests - Acceptable	66	40	16	122	58	37	20	115	237
PC Tests - Total	74	42	17	133	59	45	20	124	257
% Failed PC Tests	11	5	6	8	2	18	0	7	8
Definitive Tests Failed Only Because PC Tests Failed	14	6	14	34	0	22	0	22	56
% Definitive Tests Failed Only Because PC Tests Failed	7	2	6	5	0	9	0	4	4

348 Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in
349 Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NRU=Neutral red uptake;
350 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes.
351
352
353 The Phase III guidelines required each laboratory to provide three acceptable definitive tests
354 for each substance for both cell types (3 x 60 x 2 = 360 definitive tests). PC tests were run
355 concurrently with the definitive tests, and more than one reference substance was usually
356 tested in conjunction with each PC test. Because of test failures, each laboratory performed
357 additional testing to obtain the three acceptable definitive tests required for each substance.
358
359 **Table 8-4** presents the success rates for each laboratory for Phase III testing and a total for
360 all the laboratories combined.
361
362

Table 8-4 Combined Definitive and Positive Control (PC) Test Success Rates for the 3T3 and NHK Methods in Phase III

Test Type	ECBC	FAL	IIVS	Total
Acceptable Definitive Tests/ Total Definitive Tests	342/402	352/513	350/419	1044/1334
% Acceptable Definitive Tests	85%	69%	84%	78%
Acceptable PC Tests/Total PC Tests	124/133	77/87	36/37	237/257
% Acceptable PC Tests	93%	89%	97%	92%

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; NRU=Neutral red uptake; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes.

8.3.3 Intralaboratory Reproducibility

CV values for each method were determined for each reference substance in each laboratory using the IC_{50} values from the acceptable definitive tests, as described in **Section 5.5.1**. **Table 8-5** presents the average CV values for the substances tested in each of the study phases, and for the entire study.

Table 8-5 Coefficients of Variation for Definitive Tests

Cell Type	Labs	Phases I & II		Phase III		All Phases	
		Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV	Number of Reference Substances	Average % CV
3T3	ECBC	12	17	57	24	69	23
	FAL	11	28	55	33	66	33
	IIVS	11	20	56	22	68	21
NHK	ECBC	12	24	57	22	69	23
	FAL	12	31	57	45	69	42
	IIVS	12	14	58	14	70	14

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; CV=Coefficient of variation.

8.3.4 Prediction of Globally Harmonized System (GHS) Acute Oral Toxicity Categories

Predicted LD_{50} values were determined using the *in vitro* NRU IC_{50} values in the IC_{50}-LD_{50} regressions presented in **Table 6-4.** The predicted LD_{50} values were used to assign each

substance to a predicted GHS acute oral toxicity category (UN 2005). The accuracy of the 3T3 and NHK NRU test methods for predicting GHS acute oral toxicity categories was determined by comparison with categorization based on *in vivo* rodent oral LD_{50} values (in mg/kg) in **Table 4-2**. Using the RC rat-only millimole regression, the accuracy of the predictions and the extent of underprediction or overprediction are shown for each laboratory in **Table 8-6**. The laboratories were generally in agreement with each other in their predictions. Although FAL had the highest error rates and CV values, their predictions of GHS toxicity category were consistent with the other laboratories. The laboratories determined category matches for 25 to 30% of the reference substances for the 3T3 NRU test method and 29 to 31% of the reference substances for the NHK NRU test method. For the 3T3 NRU test method, toxicity was overpredicted for 38% of the reference substances and underpredicted for 33 to 38% of them. For the NHK NRU test method, toxicity was overpredicted for 35 to 38% of the reference substances and underpredicted for 32 to 34% of them. (See **Appendix J** for additional laboratory comparisons for the other *in vitro – in vivo* regressions evaluated in **Section 6**.)

8.4 Availability of Laboratory Notebooks

All laboratories maintained laboratory notebooks using a template provided by IIVS, and provided copies of the notebooks to the SMT (archived at NICEATM) after completion of each testing phase. The notebooks contained information from all aspects of testing including, but not limited to:

- Environmental conditions
- Reagent identification

Preparation of 96-well plates

- Preparation of reference substances
- Treatment of cell cultures
- Visual observations of cell cultures
- NRU assays
- Data analysis

Table 8-6 GHS Acute Oral Toxicity Category Predictions by Laboratory[1]

	Labs	Total Reference Substances	Category Match	Toxicity Overpredicted	Toxicity Underpredicted
3T3	ECBC	64	30%	38%	33%
	FAL	64	25%	38%	38%
	IIVS	64	27%	38%	36%
NHK	ECBC	68	31%	35%	34%
	FAL	68	29%	38%	32%
	IIVS	68	31%	37%	32%

Abbreviations: ECBC=Edgewood Chemical Biological Center; FAL=Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory; IIVS=Institute for *In Vitro* Sciences; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; GHS=Globally Harmonized System for Classification and Labelling of Chemicals (UN 2005).

[1]3T3 and NHK NRU test method IC_{50} data (geometric mean of within laboratory replicates) used with the RC rat-only millimole regression, log LD_{50} (mmol/kg) = 0.439 x log IC_{50} (mM) +0.621, to assign GHS category. *In vivo* category was based on reference rodent oral LD_{50} values (mg/kg) in **Table 4-2**. For each method, the reference substances evaluated were those for which all three laboratories obtained IC_{50} values.

8.5 Summary

- The determinations of test method and data collection errors showed that FAL consistently had the highest error levels; however, the laboratory's GHS acute oral toxicity category predictions were comparable to the other laboratories' results.
- The laboratories reported no significant deviations from the protocols, and deviations that did occur during the testing phases were generally quickly acknowledged and addressed by the Study Directors. If a deviation occurred that would affect the data (e.g., improper concentration of DMSO solvent), the Study Director would reject the test, notify the SMT, and perform an additional test. Improper transfer of data to either the EXCEL® or PRISM® templates, which would affect the data summaries and analyses, were recognized, documented, and rectified by the Study Director and/or the SMT.

- The SMT reviewed all data sheets to ensure that data were not inadvertently attributed to the incorrect data summary files, and that the correct data were used in all statistical analyses.
- An electronic copy of all data for this 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.

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS OF *IN VITRO* CYTOTOXICITY TEST METHODS AND THEIR ABILITY TO PREDICT IN VIVO ACUTE TOXICITY AND OTHER TOXIC EFFECTS.....9-3

9.1 Relevant Studies.....9-3
9.1.1 Correlation of NRU Cytotoxicity Values with Rodent Lethality.....9-3
9.1.2 Use of *In Vitro* Cytotoxicity Data to Reduce the Use of Animals in Acute Oral Toxicity Testing.....9-9
9.1.3 Other Evaluations of 3T3 or NHK NRU Methods.....9-14

9.2 Independent Scientific Reviews.....9-17
9.2.1 *In Vitro* Acute Toxicity Testing for the Classification and Labeling of Chemicals.....9-17
9.2.2 Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing.....9-19
9.2.3 Validation of 3T3 NRU Assay for Phototoxicity.....9-23

9.3 Studies Using *In Vitro* Cytotoxicity test Methods with Established Performance Standards.....9-25
9.3.1 *Guidance Document* (ICCVAM 2001b).....9-25
9.3.2 King and Jones (2003).....9-25
9.3.3 A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005).....9-26

9.4 Summary.....9-27

[This Page Intentionally Left Blank]

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS OF IN VITRO CYTOTOXICITY TEST METHODS AND THEIR ABILITY TO PREDICT IN VIVO ACUTE TOXICITY AND OTHER TOXIC EFFECTS

In vitro cytotoxicity methods based on NRU have been evaluated for a number of uses. This section reviews studies that used *in vitro* NRU cytotoxicity methods to:

- Predict acute rodent oral toxicity
- Predict starting doses for acute systemic toxicity tests
- Predict other *in vivo* toxicity endpoints, including phototoxicity and eye irritation.

Section 9.1 describes studies that evaluated *in vitro* cytotoxicity test methods that measured NRU for its ability to predict acute systemic toxicity in rodents, and other *in vivo* endpoints. Also reviewed are studies that evaluated the use of *in vitro* cytotoxicity results to reduce animal use in acute toxicity testing. **Section 9.2** reviews independent evaluations of the use of *in vitro* cytotoxicity methods to predict acute oral toxicity, and to determine starting doses for acute systemic toxicity assays. Also discussed is a 3T3 NRU test method that has been validated and accepted for regulatory use for detecting phototoxic potential using a protocol similar to that used in the NICEATM/ECVAM validation study. The conclusions of these reports will be compared to the conclusions reached in this study, wherever possible. **Section 9.3** reviews published studies that used the *Guidance Document* approach (ICCVAM 2001b), which established the current test method performance standard.

9.1 Relevant Studies

9.1.1 Correlation of NRU Cytotoxicity Values with Rodent Lethality

This section reviews five, published *in vitro* cytotoxicity studies that correlated cytotoxicity values (i.e., IC_{20} or IC_{50}) from NRU cytotoxicity test methods that used various cell types, to rat and/or mouse acute LD_{50} values from various exposure routes. In those sections, *italics* are used to identify reference substances tested in the reviewed studies that were also tested in the NICEATM/ECVAM validation study. **Table 9-1** characterizes the substances tested in the reviewed studies by providing the ranges of their rat oral LD_{50} values. Also shown for comparison are the mouse and/or rat oral LD_{50} ranges for the NICEATM/ECVAM validation

study and the RC. The table shows that the substances tested by Peloux et al. (1992), Fautrel et al. (1993), and Rasmussen (1999), covered a wide range of rat acute LD_{50} values. The substances used by Roguet et al. (1993) and Creppy et al. (2004) covered a much smaller range. **Table 9-2** characterizes the test substances by chemical class based on NLM Medical Subject Heading (MeSH®) descriptors.

Table 9-1 Rat Acute Oral LD_{50} Ranges for Test Substances Used in Previous *In vitro* NRU Cytotoxicity Studies and the NICEATM/ECVAM Study[1]

Study/Database	N	Rat Acute Oral LD_{50} Range (mg/kg)[2]
Peloux et al. (1992)	30	2 – 14500
Fautrel et al. (1993)	31	2 – 14500
Roguet et al. (1993)	28	0.04 – 176
Rasmussen (1999)	20	1 – 10298
Creppy et al. (2004)	2	48 – 924[5]
NICEATM/ECVAM Validation[3]	72	2 – 19770
RC[4]	347	1 – 31015

Abbreviations: N=Number of substances in the study/database; RC=Registry of Cytotoxicity.
[1]Studies reviewed in **Section 9.1.1**.
[2]Values cited in the studies or from references provided by the studies.
[3]Current study summarized in this BRD.
[4]The RC includes both rat and mouse LD_{50} values.
[5]Upper limit of range is an LD_{50} calculated from the *in vitro* NRU IC_{50} because there was no *in vivo* value available for that substance.

Table 9-2 Chemical Classes Represented by the Substances Used in Published Studies for Correlation of *In vitro* NRU Cytotoxicity with Rodent Acute Lethality

Chemical Class[1]	Study[2]	Chemical Class[1]	Study[2]	Chemical Class[1]	Study[2]
Alcohols	1, 2, 3, 4	Fluorine	3, 4	Nitriles	1, 2
Amides	1, 2, 3	Heterocyclics	1, 2, 3, 4, 5	Nitrogen	3, 4
Amines	1, 2	Hydrocarbons	1, 2, 3, 4, 5	Organophosphates	3, 4
Arsenicals	3, 4	Iron	3	Phenols	3, 4
Carboxylic Acids	1, 2, 3, 4	Lactones	1, 2	Polycyclics	3
Chlorine	3, 4	Lithium	1, 2, 3, 4	Potassium	3, 4
Copper	3, 4	Mercury	3, 4	Sodium	3, 4
Ethers	1, 2	Metals	3, 4	Sulfur	1, 2, 3, 4

Study references: 1=Peloux et al. (1992) (24/25 substances were organic compounds); 2=Fautrel et al. (1993) (30/31 substances were organic compounds); 3=Roguet et al. (1993) (22/30 substances were organic compounds); 4=Rasmussen (1993) (13/20 substances were organic compounds); 5=Creppy et al. (2004) (2/2 substances were organic compounds).
[1]Classification by NLM Medical Subject Heading (MeSH®) descriptors.
[2]Studies reviewed in **Section 9.1.1.**

98 9.1.1.1 *Peloux et al. (1992)*

99 The authors used several different *in vitro* cytotoxicity methods with primary rat hepatocytes
100 to determine the correlation with rat/mouse intraperitoneal (i.p.) or intravenous (i.v.) LD_{50}
101 values for the 25 substances tested. The *in vitro* cytotoxicity methods, which used 20-hour
102 test substance exposure durations, assessed the following endpoints: NRU, total protein
103 content, lactate dehydrogenase (LDH) release, and tetrazolium salt (MTT) reduction. MTT is
104 metabolized by mitochondrial succinate dehydrogenase of viable cells to yield a purple
105 formazan reaction product. The IC_{50} values obtained using the four endpoints were highly
106 correlated (r = 0.973 to 0.999) to each other. When performing the IC_{50}-LD_{50} regressions,
107 Peloux et al. (1992) used the lowest reported published LD_{50} value for acute rat or mouse
108 studies that administered the test substances using the i.p. or i.v. routes. The IC_{50} values
109 obtained using NRU as the endpoint had the highest correlation coefficient, r = 0.877, to the
110 rat/mouse i.p./i.v. LD_{50} values. The total protein assay yielded r = 0.872, the MTT reduction
111 assay yielded r = 0.808, and the LDH release assay yielded r = 0.789.
112
113 Peloux et al. (1992) followed the recommendations of Fry et al. (1988, 1990) and used
114 parenteral LD_{50} values rather than oral LD_{50} values for comparison with *in vitro* values. Fry
115 et al. (1988, 1990) recommended the use of the i.p./i.v. LD_{50} values for comparisons because
116 they proposed that cells *in vivo* receive a more direct test substance exposure via these routes
117 than through the oral route. They had posited that *in vitro* cell cultures would mirror this
118 (direct) toxicity because they also receive direct exposure to test substances via the cell
119 culture medium. The authors also noted that the oral route of exposure presents confounding
120 variables such as, 1) only a fraction of a test substance would be available in the systemic
121 circulation due to limited absorption or pre-systemic metabolism, and 2), the level of the
122 substance in the systemic circulation decreases due to elimination mechanisms (e.g.,
123 metabolism, excretion). Fry et al. (1990) had reported a correlation of only r = 0.49 for *in*
124 *vivo/in vitro* comparisons of oral LD_{50} and IC_{50} values (from a total protein assay) and a
125 correlation of r = 0.68 for i.p. LD_{50} and ID_{50} values[1].
126
127

[1] ID_{50}: index of cytotoxicity; concentrations (µg/mL) producing a 50% reduction in protein value.

127 9.1.1.2 *Fautrel et al. (1993)*

128 Six laboratories tested the cytotoxicity of 31 substances in primary rat hepatocyte cultures
129 using a 24-hour exposure followed by measurement of NRU. The investigators performed
130 linear regression analyses for the prediction of rat i.v., i.p., and oral LD_{50} values from the
131 NRU IC_{50} values. The regressions for the various *in vivo* administration routes did not use the
132 same substances because LD_{50} values were not available for all of the tested substances in all
133 of the routes. Oral, i.v., and i.p. LD_{50} values were available for 27, 24, and 18 substances,
134 respectively, and IC_{50} values were obtained for 15, 14, and 11 of these substances,
135 respectively. The regression for the i.v. data was statistically significant ($r = 0.88$, $n = 11$),
136 but the i.p. ($r = 0.48$, $n = 14$) and oral regressions ($r = 0.17$, $n = 15$) were not. The finding that
137 the i.v. LD_{50} values corresponded more closely with the *in vitro* cytotoxicity data than did the
138 oral LD_{50} was thought to be the result of having fewer pharmacokinetic variables (i.e.,
139 absorption, distribution, etc.) to consider following i.v. administration.

140 9.1.1.3 *Roguet et al. (1993)*

141 Roguet et al. (1993) tested the cytotoxicity of 28 MEIC substances in primary rat hepatocytes
142 exposed for 21 hours, followed by the measurement of NRU. A correlation of the NRU IC_{50}
143 values to oral LD_{50} values obtained from the unpublished data of B. Ekwall et al. (personal
144 communication) yielded a statistically significant linear correlation ($p < 0.001$) with $r = 0.80$
145 when the *in vivo* and *in vitro* data were in molar units. [NOTE: The LD_{50} values subsequently
146 published by Ekwall et al. (1998) were from the 1997 edition of RTECS®.] The authors
147 reported that the toxicities of thioridazine, malathion, and *copper sulfate* were overestimated,
148 and the toxicity of *potassium cyanide* was underestimated by the correlation, but their criteria
149 for over- and under- estimation were not provided.

150

151 The *in vivo* toxicity of *potassium cyanide* was also underpredicted in the
152 NICEATM/ECVAM validation study. **Table 6-2** shows that *potassium cyanide* was an
153 outlier for which toxicity was underpredicted when using the IC_{50} values from both the 3T3
154 and NHK NRU test methods in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log
155 IC_{50} mM + 0.625). The GHS category predictions using both NRU test methods and the RC
156 rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621), and the RC

157 rat-only weight regression (i.e., log LD_{50} =0.372 log IC_{50} + 2.024), were also higher (i.e., less
158 toxic) than the *in vivo* category (see **Appendix L2**).

159 9.1.1.4 *Rasmussen (1999)*

160 Twenty MEIC substances were tested for cytotoxicity using NRU release from 3T3 cells
161 following 24-hr exposure, with and without the addition of a Aroclor-induced rat liver
162 microsomal preparation (S9 mix). Similar to the present validation study, Rasmussen (1999)
163 was *xylene* was non-toxic to the cells, even though it was dissolved in ethanol instead of
164 DMSO. In the presence of S9, the cytotoxicities of malathion, 2,4-dichlorophenoxyacetic
165 acid, *propranolol*, thioridazine, *lithium* sulfate, *copper sulfate*, and *thallium sulfate,* were
166 significantly decreased ($p < 0.05$), while the cytotoxicities of *1,1,1-trichloroethane*, *phenol,*
167 *nicotine*, and *paraquat* were significantly increased ($p < 0.05$).
168
169 Because the NICEATM/ECVAM validation study used cells with little or no xenobiotic
170 metabolizing capability, it could be expected that these systems would overpredict the
171 toxicity of substances that would be inactivated by the addition of a metabolizing system, or
172 to underpredict the toxicity of substances that are metabolized to more toxic substances.
173 None of the four substances in common for which toxicity was decreased by the addition of
174 S9 were overpredicted in the NICEATM/ECVAM study. However, the toxicities of two of
175 the four substances in common for which toxicity was increased by the addition of S9, were
176 underpredicted in the NICEATM/ECVAM study. **Table 6-2** shows that *nicotine* and
177 *paraquat* were outliers whose toxicity was underpredicted when using the 3T3 and NHK IC_{50}
178 values in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log IC_{50} mM + 0.625). The
179 GHS category predictions using both NRU test methods and the RC rat-only millimole
180 regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621), and the RC rat-only weight
181 regression (log LD_{50} mg/kg = 0.357 log IC_{50} μg/mL + 2.194), were also higher than the *in*
182 *vivo* category (see **Appendix L2**).
183
184 Although both the IC_{20} and IC_{50} values were determined in the Rasmussen (1999) study, only
185 the IC_{20} values were used for correlations with the rat acute oral LD_{50} values from RTECS®.
186 The units of the LD_{50} values were not reported, but the correlations were assumed to be in
187 molar units because the IC_{20} and IC_{50} values were reported in μM units. Significant

188 correlations ($p < 0.001$) between IC_{20} and LD_{50} values were obtained with and without rat
189 liver microsomes. The correlation of IC_{20} with LD_{50} was slightly higher with the S9 mix ($r =$
190 0.72 vs. 0.68 for oral LD_{50} values, and 0.82 vs. 0.78 for i.p. LD_{50} values).
191
192 Although the presence of S9 increased the cytotoxicity of some substances to the 3T3 cells, it
193 decreased the toxicity of others, and yielded only a small improvement in the correlation to *in*
194 *vivo* data. Rasmussen (1999) concluded that the toxicity of the S9 mix (0.32 mg protein/mL),
195 itself, was insignificant because it reduced cell survival by less than 10 % compared with
196 cells without S9. However, others have shown that S9 microsomal mixes could produce
197 significant cytotoxic effects. Kohn (1993) showed that an S9 mix containing 0.07 mg
198 protein/mL was cytotoxic to all types of murine neurons in culture when the cells were
199 exposed for four days or longer. Non-neuronal cells tolerated higher concentration exposures
200 of S9, but exhibited cytoplasmic inclusions when exposed to S9 at 0.35 mg protein/mL. Dal
201 Negro et al. (2006) reported 100% cell death of human monocyte-derived U-937 cells when
202 the S9 fraction (1 mg protein/mL) and co-factors were applied to the cells for a 72-hour
203 incubation. Both of these studies used longer exposure durations, and/or higher protein
204 concentrations, than the Rasmussen (1999) study.

205 9.1.1.5 *Creppy et al. (2004)*

206 Creppy et al. (2004) used a 48-hour NRU assay to determine the cytotoxicity of ochratoxin A
207 (OTA) and fumonisin B1 (FB1) on cultured C6 glioma (rat brain), Caco-2 (human intestinal),
208 and Vero (green monkey kidney) cells. The IC_{50} determined in the NRU assay was used in
209 the RC millimole regression to predict rodent acute oral LD_{50} values. The predicted LD_{50} for
210 OTA using the C6 glioma cells was similar to mouse LD_{50} values generated from four *in vivo*
211 mouse studies, but the LD_{50} values predicted by the other cell lines were about 50 times
212 greater. The authors found the relative insensitivity of the Vero cells surprising because OTA
213 is a kidney toxin. There were no available *in vivo* rodent oral LD_{50} values with which to
214 compare the predicted LD_{50} of FB1, which ranged from 671 to 924 mg/kg for the three cell
215 types tested.
216

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

9.1.2.1 *Halle et al. (1997): Animal Savings with the ATC Method Using Cytotoxicity Data*

This study assessed the animal savings that would be produced by using IC_{50} data in an IC_{50}-LD_{50} regression to determine a starting dose for ATC testing. No cytotoxicity testing was performed for this study. Instead, the authors used the IC_{50} values from the RC database and the RC millimole regression to predict the LD_{50} for 347 RC substances. The predicted LD_{50} values were then used to determine the starting doses for simulated ATC testing.

At the time of the Halle et al. (1997) study, the ATC method (1996 version from OECD) was designed to classify substances using three classes of acute oral toxicity and an unclassified group, as defined by the acute oral toxicity classification system of the EU (see **Table 9-3**). As a result, the fixed doses for the ATC testing were 25, 200, and 2000 mg/kg. The authors used the LD_{50} predicted by the RC IC_{50} and the RC millimole regression for the 347 RC substances as a starting point to estimate the number of ATC dose steps, and number of animals, that would be needed to classify the substances in the EU category associated with the rodent oral LD_{50} (i.e., rat or mouse values from RTECS®). The method required the simulated ATC testing for each substance to start at the fixed ATC dose nearest to the predicted LD_{50}. The outcome of the simulated testing of three animals per fixed dose was determined by the *in vivo* LD_{50}. If the test dose was lower than the *in vivo* LD_{50}, animals were assumed to live and, if the test dose was higher than the LD_{50}, the animals were assumed to die. Testing of the substance would proceed with higher (when the animals lived) or lower fixed doses (when the animals died) until the substance was placed into the EU toxicity category indicated by the *in vivo* rodent oral LD_{50}.

The method of Halle et al. (1997) can be illustrated with digoxin, which has an *in vivo* mouse LD_{50} of 18 mg/kg (from RTECS®). The predicted LD_{50} of 414 mg/kg was calculated using the RC IC_{50} in the RC millimole regression (log LD_{50} [mmol/kg] = 0.435 x log IC_{50} (mM) + 0.625). Simulated ATC testing would start at the nearest fixed dose, 200 mg/kg. The three animals tested were assumed to die, and then three more animals would be tested at 25 mg/kg. The animals tested at 25 mg/kg were assumed to die and digoxin would be classified

247 in category 1 for LD_{50} ≤25 mg/kg. Thus, the classification of digoxin using the 4-category
248 system required six animals.
249
250 **Table 9-3 EU[1] Classes of Acute Oral Toxicity**
251

Category	LD_{50} (mg/kg)
1	LD_{50} ≤25
2	25< LD_{50} ≤200
3	200< LD_{50} ≤2000
Unclassified	LD_{50} >2000

252 Abbreviations: EU=European Union
253 [1]Anon (1993)
254
255 Using such simulations of ATC testing, Halle et al. (1997) estimated that 2139 animals
256 would be used to test the 347 substances:
257 • Three hundred twenty-eight would require testing with two doses using three
258 test animals each.
259 • Nineteen would require testing with three doses using three animals each.
260
261 Halle et al. (1997) cited Schlede et al. (1995) in reporting that the average number of animals
262 required to classify substances using the ATC method was 9.11 animals per test. Using this
263 average, ATC testing of the 347 RC substances would require 3161 animals. Thus, Halle et
264 al. (1997) estimated that there would be a 32% reduction ([3161-2139]/3161) in the number
265 of test animals used when the LD_{50} prediction from the RC millimole regression was used
266 with the 1996 version of the ATC method, in lieu of the animal classification procedure
267 (Halle et al. 1997).
268
269 The simulated average animal savings for the ATC in the NICEATM/ECVAM validation
270 study at dose-response slopes of 2.0 and 8.3 was 4.8% to 10.2% (0.51 to 1.09 animals) for
271 the 3T3 (67 reference substances) and NHK (68 reference substances) NRU test methods
272 (see **Section 10.3.3.2**), depending on the regression evaluated. This is considerably lower
273 than the average savings of 32% estimated by Halle et al. (1997). However, there are a
274 number of differences between the evaluation performed by Halle et al. (1997) and the
275 NICEATM/ECVAM study that contribute to the difference in calculated animal savings:

- The NICEATM/ECVAM study used six GHS acute toxicity categories for classification whereas Halle et al. (1997) used the EU toxicity classification scheme, which had only four toxicity categories. The accuracy of category prediction by any method would be higher with fewer categories.
- The NICEATM/ECVAM study used experimentally derived *in vitro* cytotoxicity data from a standardized protocol to estimate starting doses (using two regressions based on the RC substances with rat LD_{50} data), whereas Halle et al. (1997) used IC_{50} data from the RC database.
- The reference substances tested in the NICEATM/ECVAM study poorly fit the RC millimole regression. Nearly half of the reference substances were outliers (30/67 in the 3T3 NRU test method, and 31/68 in the NHK NRU test method) (see **Table 6-2**). The RC database had 95/347 (27.4%) substances outside of the prediction intervals.
- The NICEATM/ECVAM study used computer simulations of ATC testing, which incorporated assumptions about mortality distributions, to determine animals used, whereas Halle et al. (1997) used simplified assumptions (i.e., all animals lived when test dose was less than the *in vivo* LD_{50} and all animals died when test dose was greater than the *in vivo* LD_{50}).
- The NICEATM/ECVAM study determined animal savings by comparing animal use with starting doses determined by the *in vitro* data, to animals used at the default starting dose of 300 mg/kg. Halle et al. (1997) used the average animal use for the ATC for comparison to animal use with simulated testing.

9.1.2.2 *Spielmann et al. (1999): Animal Savings Using Cytotoxicity Data with the UDP Method*

Spielmann et al. (1999) recommended that an *in vitro* cytotoxicity procedure as a range finding test for the *in vivo* toxicity test to reduce the number of animals used in acute toxicity tests. The authors identified nine substances in both the RC database and an evaluation of acute toxicity methods by Lipnick et al. (1995). They then compared the LD_{50} values from Lipnick et al. (1995) to LD_{50} predictions calculated when using the RC IC_{50} values in the RC millimole regression formula ($\log LD_{50}$ [mmol/kg] = 0.435 x $\log IC_{50}$ [mM] + 0.625). For seven of the nine substances, the LD_{50} prediction was within an order of magnitude of the

307 experimental LD_{50} (OECD 1987) reported by Lipnick et al. (1995). Spielmann et al. (1999)
308 concluded that the RC millimole regression provided an adequate prediction of LD_{50}, and
309 that *in vitro* cytotoxicity data could be used to predict starting doses for the UDP. The
310 authors recommended using the IC_{50}, with the RC millimole regression, to calculate a
311 starting dose (i.e., an estimated LD_{50}) for the UDP, FDP, or ATC method whenever an IC_{50}
312 was available.
313
314 If no IC_{50} was available, Spielmann et al. (1997) recommended determining cytotoxicity
315 using a standard cell line and specific cytotoxic endpoint (e.g., NRU, total protein, MTT
316 reduction). They recommended testing 10 to 20 RC substances to demonstrate that the *in*
317 *vitro* cytotoxicity test methods provide results that are consistent with the RC millimole
318 regression. The resulting IC_{50} values would then be used to calculate a new regression (using
319 the LD_{50} values reported in the RC), which would be compared to the RC millimole
320 regression. If the new regression fit into the acceptance interval (± log 5 of the fitted
321 regression line) of the RC millimole regression, the RC millimole regression would be used
322 to predict starting doses for the UDP. If the new regression is parallel to the RC millimole
323 regression, but outside the ± log 5 acceptance interval, then the new regression would be
324 used for the prediction of the starting dose.
325
326 Spielmann et al. (1999) contended that the RC millimole regression provides a sufficient
327 prediction of LD_{50} values from IC_{50} values for substances that do not require metabolic
328 activation and are not very toxic (i.e., LD_{50} > 200 mg/kg). The authors acknowledged that the
329 fit of substances with LD_{50} <200 mg/kg to the RC millimole regression is not good, and
330 attributed the poor fit of these substances to the need for metabolic activation to a more toxic
331 substance. They suggested that the prediction of starting doses using cytotoxicity data can be
332 used with the UDP and ATC methods, but not with the FDP because dosing is not sequential
333 (which contradicted a claim made earlier in the paper that the approach could be used with
334 the FDP). They did not estimate the number of animals that might be saved with this
335 approach, but did recommend that the approach be validated experimentally using several
336 established cell lines with a limited number of representative substances from the RC.
337

338 9.1.2.3 *EPA (2004): U.S. EPA HPV Challenge Program Submission*

339 In response to the EPA HPV Chemical Challenge Program, PPG Industries, Inc., the
340 manufacturer of Propanoic acid, 2-hydroxy-, compd. with 3-[2-(dimethylamino)ethyl] 1-(2-
341 ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CAS No. 68227-46-3], and the
342 sponsor of this compound, submitted data to the EPA. This is an isolated intermediate used to
343 produce a resin component of paint products. PPG provided the following types of data in
344 their submission (http://www.**epa**.gov/chemrtk/prop2hyd/c13863rt3.pdf) to the EPA:
345 physical-chemical, environmental fate and pathway, ecotoxicity, and toxicology. The acute
346 mammalian toxicology data were generated using both *in vitro* and *in vivo* methods.
347
348 An *in vitro* NRU cytotoxicity test was conducted with 3T3 cells to estimate a starting dose
349 for the *in vivo* acute UDP oral toxicity test (OECD 2001a) (see **Appendix M** for the OECD
350 UDP test guideline). The use of this *in vitro* NRU test method was intended to minimize the
351 number of animals used for *in vivo* testing. The estimated LD_{50} of the compound as
352 determined by the NRU assay was 489 mg/kg. Therefore, the starting dose for the UDP study
353 was set at 175 mg/kg, which is the first default dose below the estimated LD_{50} value; this is
354 also the default starting dose for the UDP, and is used when no information on which to base
355 a starting dose is available. A total of fifteen female rats received the compound at 175, 550,
356 or 2000 mg/kg. Five of nine rats treated at 2000 mg/kg died prematurely on Days 2 and 3,
357 and by Day 15, 2/4 surviving animals at this dose had lost up to 25% of their Day 1 body
358 weights. The LD_{50} was estimated to be 2000 mg/kg, with a 95% confidence interval of 1123-
359 5700 mg/kg. Thus, the *in vitro* NRU test method overpredicted the toxicity of the compound
360 by estimating an LD_{50} value that was lower than that determined in the UDP test. The report
361 authors reported that a greater than predicted number of animals was used for the UDP
362 testing because the estimated LD_{50}, 489 mg/kg and, consequently, the starting dose, was
363 much lower than the *in vivo* LD_{50} of 2000 mg/kg. However, because the UDP started with the
364 default starting dose of 175 mg/kg, the claim that more animals were used is incorrect,
365 because animal use with the default starting dose is the baseline against which other animal
366 use should be compared.
367
368

9.1.3 Other Evaluations of 3T3 or NHK NRU Test Methods

This section briefly reviews five studies that evaluated NRU test methods for purposes other than the prediction of starting doses for acute oral toxicity assays. NRU test methods using either 3T3 or NHK cells have been evaluated for use as alternatives to the Draize eye irritation test, to measure phototoxicity, and to predict acute lethality in humans. Except for the 3T3 NRU phototoxicity assay, NRU methods have not been scientifically validated by an independent review for any of these purposes or accepted for regulatory use. The use of the validated 3T3 NRU test method to determine phototoxic potential is addressed in **Section 9.2**.

The *in vitro* NRU protocols evaluated in the five reviewed studies are similar to those used in the NICEATM/ECVAM validation study, all of which were based on the method of Borenfreund and Puerner (1985). The major difference is that most studies used a 24-hour test substance exposure duration for the 3T3 NRU test method, while the NICEATM/ECVAM 3T3 study used a 48-hour exposure duration. The major difference between the NHK protocols used in the reviewed studies and the protocol used in the NICEATM/ECVAM study is that the cell culture medium was changed at the time of test substance application in the NICEATM/ECVAM study.

9.1.3.1 *Draize Eye Irritation*

Triglia et al. (1989)

Four laboratories collaborated in an interlaboratory validation study to test the NHK NRU assay marketed by Clonetics® Corporation[2] for its intra- and inter-laboratory reproducibility and ability to predict *in vivo* ocular irritancy. Each laboratory tested 11 blind-coded surfactant-based substances and compared the IC_{50} values to *in vivo* Draize ocular irritancy scores.

The test exhibited the following performance characteristics for the comparison of *in vitro* and *in vivo* data:

- specificity (percentage of non-irritants correctly detected) = 93%

[2] Clonetics® Corporation sponsored this study. It was not clear in the publication if Clonetics® Corporation participated as one of the testing laboratories.

398 • sensitivity (percentage of true irritants correctly detected) = 80%
399 • predictive values (probability that an unknown agent will be properly classified)
400 – positive predictive value = 90%
401 – negative predictive value = 87%
402
403 The authors reported that there was excellent correlation among the laboratories, and good
404 correlation between the *in vitro* IC_{50} values and *in vivo* Draize scores (Spearman Rank
405 correlation coefficients between *in vivo* and *in vitro* data for the laboratories ranged from
406 0.67-0.76). The authors also concluded that the NRU test could not replace the Draize test,
407 but may be an effective screening tool for use in a battery of *in vitro* alternatives
408
409 *Sina et al. (1995)*
410 Sina et al (1995) evaluated the NHK NRU test method along with six other *in vitro* methods
411 to determine whether they could be used as complimentary tests in a battery approach to
412 estimate ocular irritation. The NRU data correlated poorly with Draize ocular scores for the
413 33 pharmaceutical intermediates tested. The Spearman correlation coefficient for the IC_{50} and
414 maximum average Draize score (MAS) was -0.10, and the Pearson correlation coefficient
415 was -0.04.
416
417 *Brantom et al. (1997)*
418 This study examined the potential of 10 alternative methods to predict the eye irritation
419 potential of cosmetic ingredients. Four laboratories tested 55 coded substances (23 single
420 ingredients and 32 formulations) using the 3T3 NRU test method, and used the resulting IC_{50}
421 values to predict modified maximum average scores (MMAS) for the Draize test.
422
423 An endpoint was generated for each test by interpolation from a plot of percent cell survival
424 versus test substance concentration. A prediction model (PM) was developed from data of 30
425 single ingredients (29 surfactants and one substance not classified by the authors) to equate
426 the IC_{50} value to an MMAS.
427

428 The interlaboratory CV for the IC_{50} values was 37.3 ± 29.8% (7.5 ± 6.8, log transformed).
429 Most of the mean IC_{50} values from a single laboratory differed by plus or minus an order of
430 magnitude from the means of all the laboratories for each substance, which the authors
431 interpreted as "no significant outliers". Correlations of NRU-predicted MMAS scores with *in*
432 *vivo* MMAS scores yielded Pearson's r values ranging from 0.25 to 0.32 for the four
433 laboratories.
434
435 Although the authors concluded the interlaboratory reproducibility was good, the IC_{50} values
436 did not predict the MMAS. The r values for the *in vitro*/*in vivo* correlations were low (0.246
437 to 0.316) and the tests all underpredicted irritants and overpredicted non-irritants. Four
438 substances were outside of the 95% confidence intervals and the authors concluded that the
439 3T3 NRU test method had wide applicability to test the remaining 51 coded substances
440 according to the limitations in the prediction model, but that it was not effective as a stand-
441 alone replacement for the Draize test across the entire irritation scale. The authors did not
442 identify the test substances.
443
444 *Harbell et al. (1997)*
445 This publication reported the results of the evaluation of 12 *in vitro* cytotoxicity assays to
446 predict ocular irritation. Data were voluntarily submitted to the US Interagency Regulatory
447 Alternatives Group (IRAG), composed of members from CPSC, EPA, and FDA. The NHK
448 NRU test method was one of the tests evaluated by six laboratories testing surfactants and
449 surfactant-containing formulations (the 3T3 NRU test method was not tested). Two
450 laboratories submitted results for the same test substances, but the other four submitted data
451 for various sets of substances and formulations.
452
453 The correlation of results from the two laboratories that independently tested the same
454 substances was r = 0.99. Correlations between the IC_{50} data and *in vivo* maximum average
455 Draize score (MAS) ranged from -0.92 to -0.54. The IRAG concluded that the assays were
456 suitable as a screening and adjunct assay to assess eye irritation over the range of toxicities
457 found in personal care and household products, and recommended that its use be limited to
458 water-soluble materials. Although the method was also evaluated for surfactants, IRAG

459 recommended that the evaluation continue for its performance in predicting eye irritation for
460 various product classes (e.g., fabric softeners, shampoos). In addition, the substance's
461 physical form should be considered because the *in vitro* toxicity of a solution of the test
462 substance will not necessarily predict toxicity of the parent, solid substance *in vivo*.

463 9.1.3.2 *Predicting Human Lethal Blood Concentrations*

464 *Seibert et al. (1992)*

465 This single laboratory study was designed to evaluate various aspects of cellular toxicity in
466 four *in vitro* test systems for their relevance and reliability with respect to acute systemic
467 toxicity, in particular, human lethal blood concentrations. The 3T3 NRU test method was one
468 of four methods evaluated with 10 MEIC substances.
469
470 The authors stated that final conclusions on the relevance of the *in vitro* systems for *in vivo*
471 data could not be determined because the variations in lethal blood concentrations were
472 unknown so that limits for over or underprediction of human *in vivo* toxicity using
473 experimental models could not be defined. In addition, the ability of *in vitro* toxicity to
474 predict *in vivo* toxicity may depend on toxicokinetic factors that were not considered in the *in*
475 *vitro* systems.

476 **9.2 Independent Scientific Reviews**

477 This section summarizes independent scientific reviews of the use of *in vitro* cytotoxicity
478 methods for the prediction of rodent acute oral toxicity, and for the reduction of animal use in
479 acute toxicity testing. The conclusions of these reviews are compared to the conclusions of
480 the current study. Also discussed is the 3T3 NRU phototoxicity method, because it is similar
481 to the 3T3 NRU test method used in the current validation study and has been validated by
482 ECVAM and is the subject of OECD Test Guideline 432 (OECD 2004).

483 9.2.1 <u>*In Vitro* Acute Toxicity Testing for the Classification and Labelling of Chemicals</u>

484 9.2.1.1 *Seibert et al. (1996): ECVAM Workshop 16*

485 ECVAM sponsored an Acute Toxicity Testing *In vitro* workshop in 1994 to review the
486 current status of various *in vitro* methods and to determine their potential uses for reducing,
487 refining, and/or replacing the use of laboratory animals for acute systemic toxicity testing.
488 The workshop participants reviewed various types of toxicity, *in vitro* cytotoxicity testing

489 schemes and strategies, inclusion of biokinetic parameters, biotransformation, biodistribution
490 *in vitro* and *in vivo*, and a proposed acute toxicity testing scheme for the classification of
491 substances.
492
493 The workshop participants agreed that some studies showed good correlations between *in*
494 *vitro* cytotoxicity data and LD_{50} values. They also acknowledged that *in vitro* basal
495 cytotoxicity tests could not address all the different of mechanisms of acute systemic toxicity.
496 Additional approaches to replacing animals would have to incorporate the three main types
497 of cellular level toxic effects that can lead to in acute systemic toxicity (i.e., basal
498 cytotoxicity, selective toxicity, and cell-specific function toxicity). The participants
499 determined that it is also important that any alternative method take into account the active
500 concentration and meaningful dose of a test substance in an *in vitro* cell culture system.
501 Quantitative comparisons of test substance concentrations must be made to evaluate the
502 effects of the test substances regarding the three types of cytotoxicity.
503
504 The biokinetics of a test substance (determined by its absorption, distribution, metabolism,
505 and elimination) must be considered when making predictions of *in vivo* toxicity using *in*
506 *vitro* toxicity data. Various methods can be used to convert *in vitro* effective concentrations
507 of a test substance to equivalent body doses. Test substance factors, such as physicochemical
508 characteristics (e.g., pKa, lipophilicity, volatility), estimates of protein binding, and *in vitro*
509 characteristics (e.g., cell concentration, cell protein concentration, ratio of cell/medium
510 volumes, medium albumin concentration), are needed for such conversions.
511 An *in vitro* tiered testing scheme was proposed by the workshop participants for using *in*
512 *vitro* methods to determine the acute oral toxicity of a substance:
513 - Stage 1: Basal cytotoxicity test
514 - Stage 2: Hepatocyte-specific cytotoxicity test to assess the role of
515 biotransformation in producing toxicity
516 - Stage 3: Test system that evaluates non-hepatocyte-specific selective
517 cytotoxicity (i.e., effects on cell-specific functions)
518

This testing scheme was proposed as an approach to classify substances by their *in vitro* toxicity. The lowest IC_{50} value determined at any of the testing stages would be used to classify a substance (i.e., very toxic, toxic, harmful, and no label). The workshop participants recommended that a feasibility study be conducted to determine the practicability, relevance, and reliability of this tiered testing scheme. As noted in **Section 6.4**, the participants suggested that an *in vitro* basal cytotoxicity test could only be used as an adjunct test, and not a stand-alone test, for classifying substances for acute oral toxicity. However, the participants also suggested that *in vitro* tests could be used to identify starting doses for acute toxicity testing to reduce the number of animals used.

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

9.2.2.1 *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the ATC Method*

Participants at Workshop 2000 examined the influence of starting dose on animal use in the ATC method (ICCVAM 2001a; section 2.2.3, pp.12-14; no testing was performed at the Workshop). The participants made inferences from the 1996 version of the ATC method that was based on the EU toxicity classification system (**Table 9-1**). The fixed doses for testing were 25, 200, and 2000 mg/kg. Normally, classification of a substance requires testing three animals in two to four dosing steps (i.e., six to 12 animals). The number of dosing steps increases with increasing distance between the true toxicity class and the starting dose. They estimated that one to three dosing steps could be avoided (i.e., three to nine animals saved) if the optimum starting dose could be predicted by *in vitro* cytotoxicity testing.

The predicted savings of one to three dosing steps was made under ideal conditions. The Workshop 2000 report (ICCVAM 2001a) provides a biometric analysis at a dose-mortality slope of 2.0 that shows that the greatest animal savings would occur for substances with very high and very low toxicity. Three animals are needed to classify a substance in the <25 mg/kg class if the true LD_{50} is 1 mg/kg and 25 mg/kg is the starting dose, but six animals are needed if the test starts from the default starting dose of 200 mg/kg (i.e., an animal savings of 50%). For a substance with a true LD_{50} of 10000 mg/kg, 11.3 animals on average are needed when the default starting dose is used, but only 7.7 animals would be needed at the 2000

mg/kg starting dose (i.e., an animal savings of 31%). For substances with a true LD_{50} of 2000 mg/kg, no animals would be saved by starting at the 2000 mg/kg dose (compared to starting at the default starting dose of 200 mg/kg).

Although these analyses were performed assuming the 1996 ATC method used starting doses of 25, 200, 2000 mg/kg, the Workshop 2000 participants noted that the animal savings that would be produced by improving the starting dose would not be significant for the current ATC method that uses GHS doses of 5, 50, 300, and 2000 mg/kg (or up to 5000 mg/kg) (OECD 2001c; see **Appendix M** for the current ATC test guideline). The Workshop 2000 participants did not predict the animal savings when *in vitro* cytotoxicity methods are used to estimate starting doses for the ATC, other than the biometric analysis described above.

The NICEATM/ECVAM study yielded patterns of animal savings with the ATC that were similar to those discussed at the 2000 Workshop (i.e., animal savings were greater for substances with a lower or higher LD_{50} than the default starting dose; see **Section 10.3.3.3**). Depending on the NRU test method and regression evaluated, the average animal savings per test (for the 67 or 68 reference substances evaluated) predicted by the NICEATM/ECVAM validation study at a dose-mortality slope of 2.0 were:

- 22.6 to 30.4 % (2.21 to 2.96 animals) for substances in the $LD_{50} \leq 5$ mg/kg category
- 10.2 to 13.0 % (1.17 to 1.51 animals) for substances in the $5 < LD_{50} \leq 50$ mg/kg category
- 3.8 to 4.3 % (0.42 to 0.47 animals) for substances in the $50 < LD_{50} \leq 300$ mg/kg category
- -9.5 to -6.1% (-0.93 to -0.60 animals) for substances in the $300 < LD_{50} \leq 2000$ mg/kg category
- -0.03 to 12.7% (-0.30 to 1.43 animals) for substances in the $2000 < LD_{50} \leq 5000$ mg/kg category
- 17.1 to 25.5% (2.03 to 3.02 animals) for substances with $LD_{50} > 5000$ mg/kg

The major differences between the evaluation reviewed by the Workshop 2000 participants and the NICEATM/ECVAM study were:

- The NICEATM/ECVAM study used the GHS toxicity categories for classification whereas the Workshop participants used the EU classification scheme, which has fewer toxicity categories. The accuracy of category prediction is higher with fewer categories.
- The NICEATM/ECVAM study used *in vitro* cytotoxicity data to estimate starting doses using two regressions based on the RC substances with rat LD_{50} data, whereas the Workshop 2000 participants used the fixed ATC doses as starting doses.
- The NICEATM/ECVAM study used computer simulations of ATC testing for individual substances whereas Workshop 2000 participants used an evaluation that estimated animal use based on fixed *in vivo* LD_{50} values and the fixed ATC doses.

9.2.2.2 *ICCVAM (2001a): Estimation of Animal Savings Using Cytotoxicity Data with the UDP Method*

Workshop 2000 participants examined the effect of starting dose on animal usage in the UDP assay by making inferences from the computer simulations of animal use shown in the peer-review BRD (ICCVAM 2000). When the rule that requires testing to stop when four animals have been tested after the first reversal is used, and no other stopping rules are considered, the animal use is relatively insensitive to the slope of the dose-mortality curve. The number of animals required when the starting dose equals the true LD_{50} is approximately six. However, approximately nine animals are required when the starting dose is 1% of the true LD_{50}. Thus, animal use is 30% less when the starting dose is the true LD_{50} compared to a starting dose that is 1% of the true LD_{50} (ICCVAM 2001a, section 2.2.4, pg. 16). When UDP testing stops based on the likelihood-ratio stopping rule, the animal use depends principally on the slope of the dose-mortality curve. The Workshop 2000 participants estimated that 25 to 40% of the animals would be saved when the starting dose is equal to the true LD_{50}, compared to the savings at a starting dose 1% of the true LD_{50}.

According to the UDP BRD (ICCVAM 2000) used by the Workshop participants, UDP simulations at a mortality-response slope of 2.0 showed that an average of 12.4 animals per test were used when the starting dose was 1% of the true LD_{50}, but an average of 8.7 animals was used when the starting dose was the true LD_{50} (i.e., a 30% reduction). At a slope of 8.3, an average of 11 animals per test were used when the starting dose was 1% of the true LD_{50}, but an average of only six animals were used when the starting dose was the true LD_{50} (i.e., a 46% reduction). The animal savings predicted by Workshop 2000 participants was 25 to 40% based on starting at the true LD_{50} in comparison to starting at a dose that is 1% of the true LD_{50}.

Depending on the regression evaluated, the average animal savings predicted in the NICEATM/ECVAM validation study at dose-response slopes of 2.0 and 8.3 were 5.8 to 7.8% (0.49 to 0.66 animals) using the 3T3 (67 reference substances) and NHK (68 reference substances) NRU test methods (see **Section 10.2.3**). When averaged for the reference substances in each GHS category, the highest mean animal savings at a mortality-response slope of 2.0 was obtained for reference substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg categories. Animal savings were 11.3 to 16.7% (1.28 to 1.65 animals) using the 3T3 and NHK NRU test methods for the two regressions. The average animal savings for the substances in these categories at a dose-mortality slope of 8.3 were 12.1 to 21.0% (1.11 to 1.63 animals) for both methods and regressions. The major differences between the evaluation performed by the Workshop 2000 participants and the NICEATM/ECVAM study were that:

- The default starting dose used for the NICEATM/ECVAM simulations was 175 mg/kg (see **Section 10.2.2**), rather than 1% of the true LD_{50} assumed by the Workshop 2000 participants.
- The NRU IC_{50} was used in various regressions of *in vitro* data against *in vivo* data to estimate starting doses. This estimation was not always close to the true LD_{50}, which was the value used by the Workshop 2000 participants. For example, LD_{50} values predicted by the NICEATM/ECVAM study for phenylthiourea were approximately 540 mg/kg by the 3T3 IC_{50} and approximately 904 mg/kg by the NHK IC_{50} using the RC rat-only millimole

641 regression (see **Appendix N**). The true *in vivo* LD_{50} for phenylthiourea is 3
642 mg/kg. Workshop 2000 participants used a best-case scenario when they
643 assumed that *in vitro* cytotoxicity precisely predicted the true LD_{50}.

644 9.2.3 Validation of the 3T3 NRU Assay for Phototoxicity

645 An NRU assay using 3T3 cells was validated by ECVAM, and accepted for regulatory use,
646 to detect the phototoxic potential of test substances. The 3T3 NRU test for phototoxicity
647 requires a 60-minute exposure to the test substance, a 50-minute exposure to ultraviolet
648 (UVA, 315-400 nm) light, followed by removal of test substance and incubation for another
649 24 hours in fresh medium (Spielmann et al. 1998). NR medium is then added, and NRU is
650 measured after a 3-hour incubation. Phototoxic potential is assessed by comparing the
651 differences in cytotoxicity between test plates containing the test substance that have not
652 been exposed to UVA and comparable test plates exposed to UVA.
653
654 Two different models, employing the Photoinhibition Factor (PIF) and the Mean Photo
655 Effect (MPE), were validated for the prediction of *in vivo* phototoxic potential. The accuracy
656 of the models for classifying the phototoxic potential of the 30 substances tested in nine
657 laboratories was 88% for the PIF, and 92% for the MPE, when compared with *in vivo*
658 classifications. Interlaboratory variability for classification (i.e., phototoxic vs. non-
659 phototoxic) was assessed using a bootstrapping approach. For each substance, the
660 classification based on a single experiment was compared to the classification based on the
661 mean PIF or mean MPE. The interlaboratory variability for classification was 0 to 18.8%
662 using PIF and 0 to 20% using MPE.
663
664 The ECVAM Scientific Advisory Committee confirmed the scientific validity of the method
665 in 1997 (ECVAM 1997) and its regulatory acceptance was noted in Annex V of Council
666 Directive 67/548/EEC part B.41 on phototoxicity, in 2000. An OECD Test Guideline, 432,
667 was finalized in 2004 (OECD 2004). The 3T3 NRU phototoxicity test is used in a tiered
668 testing approach to determine the phototoxic potential of test substances.
669
670 The performance of the 3T3 NRU phototoxicity assay could not be compared with the
671 performance of the 3T3 NRU test method used in this validation study because different

classification schemes were used (i.e., a two-category classification for the phototoxicity vs. a six-class scheme for acute oral toxicity). The ECVAM measurements of interlaboratory variability also used different techniques and were not comparable to those used for the NICEATM/ECVAM study.

9.2.3.1 *NHK NRU Phototoxicity Assay*

FAL participated in the European Union/European Cosmetic, Toiletry and Perfumery Association (EU/COLIPA) study (30 substances tested using NHK and 3T3 cells) and the ECVAM/COLIPA study (20 substances tested using NHK cells) (Clothier et al. 1999). The studies showed that the NHK NRU test method could be used to predict phototoxic potential. The accuracy for predicting *in vivo* results was similar to that of the 3T3 NRU phototoxicity test (see **Table 9-4**). The NHK NRU phototoxicity test uses the same test substance exposure duration (approximately 2 hours) as the 3T3 NRU test method, but the duration of culture after UV exposure is 72 hours rather than 24 hours. NRU was measured after a 45-minute incubation with NR.

Although the NHK NRU phototoxicity method achieved good concordance with *in vivo* phototoxicity, it has not yet been validated for regulatory use.

Table 9-4 Correct Identification of *In vivo* Phototoxicants by the NHK NRU Phototoxicity Assay

Study	3T3 NRU Phototoxicity Method	NHK NRU Phototoxicity Method
EU/COLIPA (Spielmann et al. 1998)	29/30 (97%)[1]	28/30 (93%)[1]
ECVAM/COLIPA	NA	18/20 (90%)[1] 19/20 (95%)[2]
Combined Study Data	45/45 (100%)[2]	44/45 (98%)[2]

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral Red Uptake; EU=European Union; ECVAM=European Centre for the Validation of Alternative Methods; COLIPA=The European Cosmetic Toiletry and Perfumery Association; NA=not available.
[1]Mean Photo Effect (MPF) prediction model.
[2]Photoinhibition Factor (PIF) prediction model.

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

The procedure provided in the *Guidance Document* for evaluating basal cytotoxicity assays for use in predicting starting doses for acute oral toxicity assays provides the existing performance standards for the 3T3 and NHK NRU test methods (ICCVAM 2001b).

9.3.1 *Guidance Document* (ICCVAM 2001b)

In addition to guidance for evaluating *in vitro* basal cytotoxicity methods for use in predicting starting doses for rodent acute oral toxicity assays, the *Guidance Document* provided results from testing 11 reference substances using the recommended 3T3 and NHK NRU protocols (ICCVAM 2001b). The 11 substances were chosen from the RC database so as to have a close fit to the RC millimole regression and to cover a wide range of cytotoxicity. The major differences between the *Guidance Document* protocols and the protocols used in this validation study are the reduced NR concentrations (from 50 μg/mL to 25 μg/mL in the 3T3 method, and from 50 μg/mL to 33 μg/mL in the NHK NRU test method), the increased duration of test substance exposure in the 3T3 NRU test method, from 24 to 48 hours, and the lack of a refeeding step in the NHK NRU test method just prior to substance application (see **Section 2.3** for further detail). Despite these differences, the *Guidance Document* shows that the test results for the 11 substances in both the 3T3 and NHK NRU test methods were similar to the results in the RC database. The calculated regressions for the 11 *Guidance Document* substances were:

- $\log LD_{50} = 0.506 \log IC_{50} + 0.475$ (R^2=0.985) for the 3T3 NRU test method
- $\log LD_{50} = 0.498 \log IC_{50} + 0.551$ (R^2=0.936) for the NHK NRU test method
- $\log LD_{50} = 0.435 \log IC_{50} + 0.625$ for the RC millimole regression

The 3T3 and NHK NRU regressions were compared with the RC millimole regression (347 substances) to show that the regression lines, as well as all 11 substance data points, were within the acceptance interval (± 0.5 log around the regression) of the RC millimole regression (see **Appendix D1**, *Guidance Document* **Figures 3** and **4, pg.13**).

9.3.2 King and Jones (2003)

This study also tested the 11 substances recommend in the *Guidance Document* using the

recommended 3T3 NRU protocol. The IC_{50} - LD_{50} regression obtained was comparable to the RC millimole regression and to the 11 substance regression provided in the *Guidance Document* (ICCVAM 2001b). The regression was log LD_{50} = 0.552 log IC_{50} + 0.503 (R^2=0.929) and the RC millimole regression was log LD_{50}=0.435 log IC_{50} + 0.625. The 11-substance regression fit within the acceptance interval (± 0.5 log) of the RC millimole regression.

King and Jones (2003) also showed that a 3T3 NRU test method that was adapted for high throughput testing by using three test sample concentrations yielded approximately the same IC_{50} as an eight concentration-response. A regression used to compare the IC_{50} values using the two different concentration-response approaches yielded R^2 = 0.945.

9.3.3 A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005)

The A-Cute-Tox Project is an Integrated Project under the EU 6th framework program that started in January 2005, with a termination date of January 2010. It was initiated in response to the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Directive and the 7th amendment of the Cosmetics Directive, which call for the broad replacement of animal experiments for finished products by 2003, and for ingredients by 2009. The project is an extension of the NICEATM/ECVAM validation study and the EDIT program, which is the continuation of the MEIC program. The partnership is made up of the EDIT Consortium, ECVAM, and 35 other European toxicity research group partners.

The aim of the project is to develop a simple and robust *in vitro* testing strategy for prediction of human acute oral toxicity, which could replace the animal acute oral toxicity tests currently used for regulatory purposes. The objectives of A-Cute-Tox are:

- Compilation, critical evaluation, and generation of high quality *in vitro* and *in vivo* data for comparative analysis.
- Identifying factors (e.g., kinetics, metabolism, and organ specificity) that influence the correlation between *in vitro* toxicity (concentration) and *in vivo* toxicity (dosage), and to define an algorithm that accounts for these effects.

762 • Explore innovative tools and cellular systems to identify new toxicity end-
763 points and strategies to better anticipate animal and human toxicity.
764 • To design a simple, robust and reliable *in vitro* test strategy associated with
765 the prediction model for acute toxicity that is amenable to high-throughput
766 testing.
767
768 The project has been divided into the following workpackages that will be implemented by
769 various configurations of research partners:
770 • WP1: Generation of a "high quality" *in vivo* database (through literature
771 searches and historical data) and establishment of a depository list of
772 reference substances
773 • WP2: Generation of a "high quality" *in vitro* database (including data from the
774 NICEATM/ECVAM study, EDIT studies, and MEIC studies)
775 • WP3: Iterative amendment of the testing strategy
776 • WP4: New end-points and new cell systems
777 • WP5: Alerts and correctors in toxicity screening (I): Role of absorption,
778 distribution, and excretion
779 • WP6: Alerts and correctors in toxicity screening (II): Role of metabolism
780 • WP7: Alerts and correctors in toxicity screening (III): Role of target organ
781 toxicity (i.e., neuro-, nephro-, hepato-toxicity)
782 • WP8: Technical optimisation of the amended test strategy
783 • WP9: Pre-validation of the test strategy
784
785 A-Cute-Tox aims to extend the NICEATM/ECVAM and EDIT/MEIC approaches toward a
786 full replacement test strategy by improving the prediction of acute oral toxicity using *in*
787 *vitro* methods, and then validating the testing procedure.

788 **9.4 Summary**

789 • *In vitro* NRU cytotoxicity test methods using various cell types have been
790 evaluated for their correlation with rodent lethality endpoints (e.g., rat/mouse
791 i.v., i.p., and oral toxicity). Peloux et al. (1992) and Fautrel et al. (1993)

showed good correlations (r=0.877 and 0.88, respectively) of *in vitro* cytotoxicity with rodent i.p./i.v. and i.v. toxicity data, respectively.

- 3T3 and NHK NRU test 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).
- 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 to 40% (ICCVAM 2001a), as compared with the average animal savings of 5.3 to 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 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 to 10.2%.

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT) 10-3

10.1 Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Rodent Acute Oral Toxicity Assays 10-4

10.2 Reduction and Refinement of Animal Use for the UDP 10-5
10.2.1 *In Vivo* Testing Using the UDP 10-5
10.2.2 Computer Simulation Modeling of the UDP 10-7
10.2.3 Animal Savings in the UDP When Using 3T3- and NHK-Based Starting Doses 10-9
10.2.4 Refinement of Animal Use for the UDP When Using 3T3- and NHK-Based Starting Doses 10-19
10.2.5 Accuracy of UDP Outcomes Using the IC_{50}-Based Starting Doses 10-19

10.3 Reduction and Refinement of Animal Use in the ATC Method 10-21
10.3.1 *In Vivo* Testing Using the ATC Method 10-21
10.3.2 Computer Simulation Modeling of the ATC Method 10-21
10.3.3 Animal Savings for the ATC Method When Using 3T3- and NHK-Based Starting Doses 10-24
10.3.4 Refinement of Animal Use in the ATC Method When Using 3T3- and NHK-Based Starting Doses 10-34
10.3.5 Accuracy of the ATC Method Outcomes Using IC_{50}-Based Starting Doses 10-34

10.4 The Impact of Accuracy on Animals Savings 10-36

10.5 The Impact of Prevalence on Animals Savings 10-39

10.6 Summary 10-39

[This Page Intentionally Left Blank]

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

As demonstrated in **Section 6**, *in vitro* basal cytotoxicity methods cannot be used as replacement assays[1] for rodent acute oral toxicity test methods for hazard classification. However, as described in this section, these methods can be used to reduce[2] and refine[3] animal use in the UDP or ATC acute oral toxicity assays, as shown by the computer simulations of such testing. Although the use of *in vitro* cytotoxicity data to determine starting doses for the FDP may reduce the use of animals for the FDP, even though death is not the primary endpoint, such an evaluation will not be provided in this document.

The test guidelines recommend using information on structurally-related substances and the results of any other toxicity tests (EPA 2002b) for the test substance, including *in vitro* cytotoxicity results, to approximate the LD_{50} and the slope of the dose-mortality curve (OECD 2001a; OECD 2001d; EPA 2002a). However, for the purposes of the reduction and refinement evaluation conducted in this section, it was assumed that no information other than 3T3 and NHK NRU IC_{50} data would be available. To determine the extent of animal reduction or refinement that would occur in the UDP and the ATC method when using a starting dose based on 3T3 or NHK NRU IC_{50} values rather than the default starting dose, computer models were used to simulate the *in vivo* testing of the reference substances used in the validation study.

Section 10.1 lists the regressions that were run with IC_{50} data from the 3T3 and NHK NRU test methods to determine starting doses for the UDP and the ATC. **Sections 10.2.1** and **10.3.1** summarize the animal testing procedures in the current test guidelines for the UDP and the ATC, respectively. The procedures for using computer simulation of the animal testing of the reference substances are described in **Sections 10.2.2** and **10.3.2**. The computer simulations were used to determine the numbers of animals used and the numbers of animals

[1] Replacement alternative: a new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).
[2] Reduction alternative: a new or modified test method that reduces the number of animals required.
[3] Refinement alternative: a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

that "died" for each test. The modeling was performed using five different dose-mortality slopes[4] (i.e., 8.3, 4.0, 2.0, 0.8, and 0.5) because such slope information was not available for all of the reference substances used. To simplify the presentation of results, the animal use figures provided in **Sections 10.2.3**, **10.2.4**, **10.3.3**, and **10.3.4** include the data for only two of the slopes, 8.3 and 2.0. The slope of 2.0 is the default used for the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3 is shown to represent substances, such as pesticides, with higher slopes. The results for the other three slopes were calculated, and are provided in **Appendices N** and **Q**. The numbers of animals used are summarized to show the mean number of animals tested when the default starting dose is used and the mean number of animals used when the starting dose was determined from the 3T3 or NHK NRU IC_{50} values. The difference in animal use between the default starting doses and the IC_{50}-based starting doses is referred to as the animal savings. Differences were tested for statistical significance (at $p < 0.05$) using a one-sided Wilcoxon signed ranked test based on the number of substances evaluated. **Sections 10.2** and **10.3** summarize mean animal use by the total number of substances tested and by the number of substances in each GHS category. **Sections 10.2.4** and **10.3.4** provide the mean number of animal deaths compared to the mean number of animals used for each default and IC_{50}-based starting dose to determine whether the IC_{50}-based starting doses lead to a reduction in the number of animals used and the number that die (i.e., refinement). **Sections 10.2.5** and **10.3.5** discuss concordance for the reference substance outcomes of simulated testing using the IC_{50}-based starting doses, with the outcomes of the default starting doses.

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

The IC_{50} values developed from the 3T3 and NHK NRU test methods were used to predict starting doses for rodent acute oral toxicity tests using the following linear regressions of IC_{50}-LD_{50} values (from **Section 6.3**):

- The RC rat-only millimole regression: $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621

[4] The dose-mortality slope is the slope of the dose-response curve for mortality.

- The RC rat-only weight regression: log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024

The IC_{50} values from each *in vitro* NRU test method, regression, and simulated acute oral toxicity test method, were evaluated. The criteria for the use of a reference substance for this evaluation were that it must have:

- Replicate IC_{50} values from at least one laboratory
- A rat acute oral LD_{50} reference value (from **Table 4-2**)

Sixty-seven and 68 reference substances were evaluated for the 3T3 and the NHK NRU test methods, respectively. Of the 72 reference substances tested, epinephrine bitartrate, colchicine, and propylparaben were excluded because they did not have associated rat oral LD_{50} data. Carbon tetrachloride and methanol were excluded from the 3T3 evaluations, and carbon tetrachloride was excluded from the NHK evaluations, because none of the laboratories achieved sufficient toxicity in any test for the calculation of an IC_{50}.

10.2 Reduction and Refinement of Animal Use for the UDP

10.2.1 *In Vivo* Testing Using the UDP

This section describes the general dosing procedure for the UDP (OECD 2001a; EPA 2002a). Although doses, interval between doses, and dose progression, may be adjusted as necessary, the procedures described reflect the default guidance. Guidance on the types of animals that can be used, animal housing, clinical observations, etc., are outside the scope of the current discussion and are provided in the test guidelines (see **Appendix M**).

10.2.1.1 *Main Test*

The UDP is based on a staircase design in which single animals are dosed, in sequence, at 48-hour intervals. The effect on the first animal determines the dose of the next animal. If the first animal dies or is in a moribund state within 48-hr after dosing, the dose administered to the next animal is lowered by dividing the original dose by one-half log (i.e., 3.2, which is the default dose progression). If the first animal survives, the dose administered to the next animal is increased by one-half log times the original dose. A dose progression of one-half log unit corresponds to a dose-mortality slope of 2.0. The default dose progression can be

133 adjusted if the analyst has prior information upon which to estimate a slope.
134
135 The starting dose recommended by the guideline is one dose progression step below the
136 analyst's best estimate of the LD_{50}, because, in the UDP test method, the LD_{50} estimate tends
137 to move toward the starting dose. A default starting dose of 175 mg/kg is used if there is no
138 information on which to base a starting dose. The default dosing scheme, using the dose
139 progression of 3.2, is 1.75, 5.5, 17.5, 55, 175, 550, 1750, and 5000 mg/kg (EPA 2002a) or
140 1.75, 5.5, 17.5, 55, 175, 550, and 2000 mg/kg (OECD 2001a). The difference between the
141 two reflects the different maximum doses in the different guidelines. Dosing single animals,
142 upward or downward, in sequence proceeds until the first of three conditions, referred to as
143 stopping rules, is met:

144 • Three consecutive animals survive at the upper dose limit (2000 or 5000
145 mg/kg)
146 • Five reversals[5] occur in any six consecutive animals tested
147 • Four or more animals have followed the first reversal, and the likelihood-
148 ratios specified by the guideline exceed the critical value. For a wide variety
149 of LD_{50} values and dose-mortality slopes, this rule is satisfied with four to six
150 animals after the first reversal. Three likelihood values are calculated: a
151 likelihood at an LD_{50} point estimate (called the rough estimate or dose-
152 averaging estimate); a likelihood at a value below the point estimate (the point
153 estimate divided by 2.5); and a likelihood at a value above the point estimate
154 (the point estimate multiplied by 2.5). The ratios of the likelihoods are
155 examined to determine whether they exceed a critical value.
156
157 If none of these conditions is met, the dosing stops after 15 animals have been used.

158 10.2.1.2 *Limit Test*

159 The UDP guidelines include a limit test using three to five animals dosed sequentially at
160 2000 mg/kg (OECD 2001a) or 5000 mg/kg (EPA 2002a). The EPA guideline for testing at a

[5] Reversal: a situation where a nonresponse (i.e., animal lives) is observed at some dose, and a response is observed at the next dose tested (i.e. animal dies), or vice versa. Reversal is created by a pair of responses. (See **Appendices M1** and **M2**)

limit dose calls for proceeding to the main test if the first animal dosed at 5000 mg/kg dies (EPA 2002a). If the first animal lives, two more animals are dosed, in sequence, with 5000 mg/kg. If both animals live, then testing is terminated, and the substance is designated as having an LD_{50} >5000 mg/kg. If one or both animals die, then two more animals are dosed in sequence. As soon as a total of three animals survive, the test is terminated, with the conclusion that LD_{50} >5000 mg/kg. However, the main test is conducted if three animals die.

The OECD guideline for testing at a limit dose calls for proceeding to the main test if the first animal dosed at 2000 mg/kg dies (OECD 2001a). If the animal lives, four more animals are sequentially dosed. The main test is performed if three animals die. If three or more animals survive, testing is terminated with the conclusion that the LD_{50} >2000 mg/kg.

10.2.2 Computer Simulation Modeling of the UDP

Ten thousand UDP testing simulations were run for each substance, *in vitro* NRU test method, and dose-mortality slope. Because the analysis assumed there was no information upon which to estimate a dose-mortality slope, the modeling used the default dose progression factor of 3.2, and 5000 mg/kg as the upper limit dose because this upper limit is specified in the EPA guideline (EPA 2002a). If the starting dose estimated from the *in vitro* IC_{50} value was ≥4000 mg/kg, then the limit test, rather than the main test, was performed. If, during the dose progression, the next highest dose to be administered was approximately 4000 mg/kg or greater, then the limit dose of 5000 mg/kg was administered. If a dose one step below the IC_{50}-estimated LD_{50} was selected to be used as the starting dose, the other doses administered corresponded to the default doses specified in the guidelines (OECD 2001a; EPA 2002a). The simulation modeling procedures also used a lower limit of 1 mg/kg. Thus, a dose of 1 mg/kg was administered if the dose progression fell below 1 mg/kg. To estimate animal use by the default method, a starting dose of 175 mg/kg was used; the other doses administered after the default starting dose corresponded to the doses specified in the guidelines (OECD 2001a; EPA 2002a).

The simulation was performed using SAS® version 8 (SAS 1999) and implemented the distributional assumptions underlying the dose-mortality relationship. The lowest dose at which an animal dies in response to the administration of a toxic substance varies from

animal to animal. For an entire population of animals, mortality is assumed to have a log-normal distribution, with the mean equal to the log of the true LD_{50}. Sigma (σ), the variability of the simulated population, is the inverse of the slope of the dose-mortality curve. Because of a lack of information concerning the actual dose-mortality curves, the simulations assumed several different values of the slope, but no corresponding changes were made in the dose progression. Dose-mortality slopes of 0.5, 0.8, 2.0, 4.0, and 8.3 were used because these were used in the simulation modeling used to evaluate the current version of the UDP guidelines (ICCVAM 2001c).

To model the variability of the IC_{50} values within and among laboratories, the values for each reference substance were log-transformed to normalize their distribution. The mean and variance of these log-transformed values were used to generate a log-normal distribution from which an IC_{50} value was randomly selected. This IC_{50} value was used with the regressions to determine starting doses using two different methods. One method used the LD_{50} estimated from the IC_{50} and the regression as the starting dose, while the other used the closest default dose that was lower than the estimated LD_{50}. The latter method is recommended by the EPA and OECD test guidelines (EPA 2002a; OECD 2001a), and the results from that simulation are in **Section 10.2**. Additionally, the UDP is only usable for regulatory purposes if the starting dose is set below the expected LD_{50}. **Appendix Q** contains the results obtained when the LD_{50} that was estimated by the IC_{50} and the regression was used as the starting dose.

The simulation procedure used the following steps for each reference substance:

1. The LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value and the choices of assumed slope were entered as the true slopes for the dose-mortality curve.
2. An IC_{50} value was selected from a distribution identified by the mean and variance of the IC_{50} values for each chemical to reflect the variation in IC_{50} values produced by the different laboratories (see **Tables 5-4** and **5-5** for mean IC_{50} values and standard deviations for the 3T3 and NHK NRU test methods, respectively).

3. The IC_{50} value from Step 2 was used in the regression model being evaluated to predict a LD_{50} value to use as the starting dose.
4. The dosing simulation was run three times: once with the default starting dose of 175 mg/kg, once at the next default dose below the LD_{50} estimated by the regression being evaluated, and once at a dose equal to that of the LD_{50} estimated by the regression being evaluated.
5. For each simulated trial, the animals are dosed sequentially; therefore for each animal (*i*) there is a corresponding dose (*i*) that is administered to the animal. For the first animal in each trial, it is the starting dose for that trial. For each subsequent animal, the dose is dependent on the previous dose and the previous animal's response, as described in **Section 10.2.1**. For animal (*i*), the probability of a response is computed with the cumulative log-normal distribution at the dose administered. That is,

$$P(response) = P(x < \log[dose(i)]) \text{ where } x \sim N(\mu, \sigma),$$

where μ is the log of the true LD_{50} value, and σ is the inverse of the assumed slope of the dose-mortality curve. One observation is then sampled from a binomial distribution with this calculated probability of success to determine whether the animal lives or dies.
6. Dosing simulation is stopped as soon as one of the stopping rules is satisfied.

Steps 2-6 were repeated 10,000 times in order to compute an average animal use for each method evaluated.

10.2.3 Animal Savings in the UDP When Using 3T3- and NHK-Based Starting Doses

10.2.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*

As described in **Section 10.2.2**, the simulation modeling of animal use for the UDP assumed five different dose-mortality slopes in order to assess animal use under various conditions of population variability. **Table 10-1** shows that the number of animals used for the UDP decreases with increasing slope for both the default starting dose and the IC_{50}-determined starting dose when based on the RC rat-only millimole regression. The IC_{50}-determined starting dose was the next default dose lower than the regression-estimated LD_{50}. For

253 example, because the LD_{50} predicted for cadmium chloride by the 3T3 NRU IC_{50} with the
254 RC rat-only millimole regression was 16 mg/kg, the starting dose was 1.75 mg/kg (i.e., the
255 next default dose below the predicted LD_{50}). This approach is consistent with the UDP
256 guidelines (OECD 2001a; EPA 2002a) as a means of reducing the number of animals that
257 might experience pain and suffering from a treatment. This approach also overcomes the
258 nonconservative bias of the UDP, which tends to yield an LD_{50} close to the starting dose.
259
260 **Table 10-1** shows that, for each dose-mortality slope, the mean number of animals saved was
261 statistically significant ($p<0.05$) when compared to mean number of animals needed when
262 the default starting dose was used. When expressed as a percentage of the number of animals
263 used when the default starting dose is used, animal savings also generally increased with
264 increasing slope of the dose-response. The animal savings is the same at all slopes tested, but
265 the fewer animals are used at the steeper slopes, which increases the relative percentages of
266 animals saved.
267
268 To simplify the presentation of animal savings and the comparison of the various regressions
269 and starting doses, the results of subsequent analyses presented in **Section 10.2.3** are limited
270 to the dose-mortality slopes of 2.0 and 8.3. The slope of 2.0 is the default used for the
271 calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3 is
272 shown to represent substances, such as pesticides, with higher slopes. Animal savings results
273 for the other dose-mortality slopes were calculated, and are presented in **Appendices N1-N3.**
274 Although using the next lower default dose to the *in vitro*-determined LD_{50} value overcomes
275 the bias of the UDP toward the starting dose (OECD 2001a, EPA 2002a) and is the
276 appropriate approach for regulatory use, animal savings results using the estimated LD_{50} as
277 the starting dose were also calculated (see **Appendix Q**).

278 10.2.3.2 *Mean Animal Use for UDP Simulations – Comparison of Regressions and*
279 *Predictions from the 3T3 and NHK NRU Test Methods*

280 **Table 10-2** shows the mean animal use for the simulated UDP testing of the reference
281 substances described in **Section 10.1**. Mean animal use is shown for the default starting dose
282 and for starting doses that were one default dose lower than the LD_{50} predicted from the *in*
283 *vitro* NRU methods and the regressions evaluated in **Section 6.4** for the prediction of GHS

284 category. The difference in animal use between the two starting doses is the mean animal
285 savings produced by using the starting dose based on the *in vitro* NRU methods. All
286 differences (i.e., mean animal savings) were statistically significant ($p < 0.05$) by a one-sided
287 Wilcoxon signed rank test. Mean animal savings ranged from 0.49 to 0.66 (6.2% to 7.0%)
288 animals per test depending upon the *in vitro* NRU test method, regression, and dose-mortality
289 slope. The lowest mean animal savings were obtained for the RC rat-only millimole
290 regression (0.49 [6.2%] to 0.54 [5.8%] animals for the different test methods and dose-
291 mortality slopes), and the greatest mean animal savings were obtained with the RC rat-only
292 weight regression (0.54 [6.8%] to 0.66 [7.0%] animals per test).
293
294 **Table 10-1 Change in Animal Use[1] with Dose-Mortality Slope for the UDP[2]**
295

Dose-Mortality Slope	With Default Starting Dose[1,3]	With IC_{50}-Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	10.01 ± 0.10	9.48 ± 0.11	0.53* (5.3%)
0.8	9.95 ± 0.13	9.34 ± 0.14	0.61* (6.1%)
2.0	9.35 ± 0.16	8.80 ± 0.17	0.54* (5.8%)
4.0	8.68± 0.18	8.15 ± 0.19	0.52* (6.0%)
8.3	7.95 ± 0.18	7.42 ± 0.20	0.53* (6.6%)
NHK NRU Test Method			
0.5	10.01 ± 0.09	9.53 ± 0.12	0.49* (4.9%)
0.8	9.96 ± 0.13	9.41 ± 0.15	0.55* (5.5%)
2.0	9.36 ± 0.16	8.86 ± 0.18	0.50* (5.3%)
4.0	8.66 ± 0.17	8.18 ± 0.20	0.48* (5.6%)
8.3	7.92 ± 0.18	7.43 ± 0.20	0.49* (6.2%)

296 Abbreviations: UDP=Up-and-Down Procedure; 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human
297 keratinocytes; RC=Registry of Cytotoxicity.
298 *Statistically significant ($p<0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in
299 parentheses.
300 [1]Mean numbers of animals ± standard errors for 10,000 simulations for each of the 67 substances in the 3T3
301 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole
302 animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit
303 dose =5000 mg/kg.
304 [2]OECD (2001a); EPA (2002a).
305 [3]Default starting dose = 175 mg/kg.
306 [4]The starting dose = next lower default dose to the predicted LD_{50}, which was calculated from the IC_{50} value in
307 the RC rat-only millimole regression: $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621. The IC_{50} value for
308 each reference substance was randomly selected from the distribution of values obtained during the testing with
309 each method.
310 [5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based
311 starting dose.
312
313

The animal savings using the *in vitro* NRU test methods with the RC rat-only regressions apply only to the reference substances evaluated in this validation study, and are based on substances pre-selected for their known *in vivo* toxicities and fit to the RC regression and may not be broadly applicable to other substances. **Table 3-4** shows that 22 (38%) of the 58 RC substances selected for testing were known to have a poor fit to the RC millimole regression (i.e., the *in vivo* LD_{50} was outside the RC acceptance interval for the predicted LD_{50}). **Table 6-2** shows that 43% (30/70 for the 3T3) and 44% (31/71 for the NHK) of the reference substances that produced IC_{50} values were outliers. The RC rat-only millimole regression evaluated here is very similar to the RC millimole regression (see **Table 6-4**). Substances with better fits to the regression are more likely to yield greater animal savings.

10.2.3.3 *Animal Savings in the UDP by Toxicity Category Using 3T3- and NHK-Based Starting Doses*

Tables 10-3 and **10-4** show mean animal use and mean animal savings for the UDP when the default starting dose and the IC_{50}-predicted starting doses were used, and when the reference substances are grouped by GHS category (UN 2005). The data come from the same analyses as the data provided in **Table 10-2**. The IC_{50}-predicted starting doses were based on the:

- RC rat-only millimole regression (**Table 10-3**)
- RC rat-only weight regression (**Table 10-4**)

These analyses showed that:

- For each *in vitro* NRU test method and regression, animal savings were statistically significant for substances in the $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg toxicity categories.
- For substances with $5 < LD_{50} \leq 50$ mg/kg and $50 < LD_{50} \leq 300$ mg/kg, both *in vitro* NRU test methods with each regression used slightly more animals than the default-dose method, but the differences were not statistically significant.

Table 10-2 Mean Animal Use[1] in the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the Different Regressions

Assay/Regression	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[5]	Animals Saved[5]
3T3 NRU Test Method	**Dose-mortality Slope = 2.0**			**Dose-mortality Slope = 8.3**		
RC rat-only millimole[6]	9.35 ± 0.16	8.80 ± 0.17	0.54* (5.8%)	7.95 ± 0.18	7.42 ± 0.20	0.53* (6.6%)
RC rat-only weight[7]	9.36 ± 0.16	8.70 ± 0.16	0.66* (7.0%)	7.94 ± 0.18	7.32 ± 0.19	0.62* (7.8%)
NHK NRU Test Method	**Dose-mortality Slope = 2.0**			**Dose-mortality Slope = 8.3**		
RC rat-only millimole[6]	9.36 ± 0.16	8.86 ± 0.18	0.50* (5.3%)	7.92 ± 0.18	7.43 ± 0.20	0.49* (6.2%)
RC rat-only weight[7]	9.36 ± 0.16	8.80 ± 0.17	0.56* (6.0%)	7.92 ± 0.18	7.38 ± 0.20	0.54* (6.8%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.
*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean numbers of animals ± standard errors for 10,000 simulations for each of the 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =5000mg/kg.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose =175 mg/kg.
[4]The starting dose = one default dose lower than the predicted acute oral LD_{50} calculated using the IC_{50} value in the specified regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the *in vitro* testing with each test method.
[5]Difference between mean animal use with default starting dose and mean animal use with the IC_{50}-based starting dose.
[6]$\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.
[7]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.

Animal Savings for the UDP by GHS Acute Oral Toxicity Category Using 3T3- and NHK-Based Starting Doses with the RC Rat-Only Millimole Regression

Table 10-3 shows the animal savings by GHS category when the IC_{50} values are used with the RC rat-only millimole regression. Mean animal savings were statistically significant ($p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, test methods, and dose-mortality slopes:

- The use of the NHK NRU test method at both dose-mortality slopes for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced savings of 0.49 (6.5%) to 0.52 (6.1%) animals per test.
- The use of the 3T3 NRU test method at the 8.3 dose-mortality slope for substances with $300 < LD_{50} \leq 2000$ mg/kg that produced a saving of 0.31 (4.1%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with $2000 < LD_{50} \leq 5000$ mg/kg that produced savings of 1.11 (12.1%) to 1.28 (11.9%) animals per test.
- The use of both *in vitro* NRU test methods and both dose-mortality slopes for substances with an $LD_{50} > 5000$ mg/kg that produced savings of 1.47 (14.8%) to 1.58 (20.3%) animals per test.

The mean animal savings, or expenditures, using the 3T3 and NHK NRU test methods were similar for most toxicity categories at both dose-mortality slopes, with the mean savings with the 3T3 slightly higher than with the NHK. For the dose-mortality slope of 2.0, the mean animal savings with the 3T3 NRU test method ranged from -0.42 (-5.5%) to 1.58 (16.0%) animals per test for the various toxicity categories, and savings for the NHK NRU test method ranged from -0.34 (-3.5%) to 1.47 (14.8%) animals per test. For the dose-mortality slope of 8.3, animal savings, or expenditures, for the 3T3 NRU test method ranged from -0.29 (-4.3%) to 1.58 (20.3%) animals per test and savings for the NHK NRU test method ranged from -0.33 (-3.9%) to 1.47 (19.2%) animals per test. Animal savings were also obtained for highly toxic substances ($LD_{50} \leq 5$ mg/kg) with both the 3T3 (0.96 [9.9%] to 1.19 [10.0%] animals per test) and NHK (0.71 [7.3%] to 0.75 [6.7%] animals per test) NRU test methods, but the savings were not statistically significant.

388

389 No mean animal savings (≤-0.28 animal per test) were observed for substances with 50<
390 LD_{50} ≤300 mg/kg by either the 3T3 or the NHK NRU test method. This category includes the
391 default starting dose of 175 mg/kg. Animal savings were not expected for this category
392 because savings were determined by comparing animal use with the IC_{50}-based starting dose
393 with animal use at the default starting dose. No animal savings (-0.07 to -0.34 animals per
394 test) were observed for substances with 5< LD_{50} ≤50 mg/kg for either NRU test method.
395 None of these differences in animal use was statistically significant.
396
397 The animal savings from the future use of these *in vitro* NRU test methods with the RC rat-
398 only millimole regression will depend on the proportion of test substances that will fall into
399 each of the GHS categories.
400
401 *Animal Savings for the UDP by GHS Category Using 3T3- and NHK-Based Starting Doses*
402 *with the RC Rat-Only Weight Regression*
403 **Table 10-4** shows the mean animal savings by GHS acute oral toxicity category when the
404 IC_{50} values are used with the RC rat-only weight regression. A comparison of mean animal
405 savings, by category, with the RC rat-only millimole regression, indicates that, in most cases,
406 animal savings were slightly higher for the RC rat-only weight regression than for the
407 millimole regression. In the RC rat-only weight regression, the mean differences between
408 animal use for the default starting dose and mean animal use with the IC_{50}-determined
409 starting dose were statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test
410 for the following GHS categories, NRU test methods, and dose-mortality slopes:

411 • The use of the 3T3 NRU test method at the 8.3 mortality-slope for substances
412 with 300< LD_{50} ≤2000 mg/kg that produced a savings of 0.28 (3.8%) animals
413 per test.
414 • The use of both *in vitro* NRU test methods at both dose-mortality slopes for
415 substances with 2000< LD_{50} ≤5000 mg/kg that produced savings of 1.28
416 (14.0%) to 1.64 (15.2%) animals per dose.

- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with LD_{50} >5000 mg/kg that produced savings of 1.53 (20.0%) to 1.65 (16.7%) animals per test.

For the dose-mortality slope of 2.0, the mean animal savings (for the various GHS categories) with the 3T3 NRU test method ranged from -0.25 (-3.3%) to 1.65 (16.7%) animals per test, and from -0.24 (-3.1%) to 1.54 (15.6%) animals per test using the NHK NRU test method. At the dose-mortality slope of 8.3, animal savings with the 3T3 NRU test method ranged from -0.18 (-2.7%) to 1.63 (21.0%) animals per test, and savings for the NHK NRU test method ranged from -0.18 (-2.7%) to 1.53 (20.0%) animals per test. Animal savings were also obtained for highly toxic substances (LD_{50} ≤5 mg/kg) with both the 3T3 (0.96 0.78 [8.0%] to 0.90 [8.0%] animals per test) and NHK (0.69 [7.1%] to 0.72 [6.4%] animals per test) NRU test methods, but these savings were not statistically significant.

There were no mean animal savings (≤-0.18 animals per test) for substances with 50 < LD_{50} ≤300 mg/kg with either *in vitro* NRU test method. This category includes the default starting dose of 175 mg/kg. Animal savings were not expected for this category because savings were determined by comparing animal use with the IC_{50}-based starting dose with animal use at the default starting dose. For the NHK NRU test method, there were no animal savings (-0.07 to -0.13 animals per test) when used for substances with 5 < LD_{50} ≤50 mg/kg. None of these small changes in animal use were statistically significant.

The animal savings from testing new substances with these *in vitro* NRU test methods using the RC rat-only weight regression will depend on the proportion of test substances that fall into each of the GHS categories.

Table 10-3 Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

		Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**
		3T3 NRU Test Method					
LD_{50} ≤5 mg/kg	6	11.32 ± 0.20	10.19 ± 0.70	1.14 (10.0%)	9.70 ± 0.28	8.74 ±0.43	0.96 (9.9%)
5< LD_{50} ≤50 mg/kg	11	9.68 ± 0.23	9.74 ± 0.45	-0.07 (-0.7%)	8.46 ± 0.28	8.54 ± 0.47	-0.08 (-1.0%)
50< LD_{50} ≤300 mg/kg	12	7.76 ± 0.10	8.18 ± 0.21	-0.42 (-5.5%)	6.61 ± 0.19	6.90 ± 0.19	-0.29 (-4.3%)
300< LD_{50} ≤2000 mg/kg	16	8.53 ± 0.21	8.14 ± 0.21	0.38 (4.5%)	7.46 ± 0.24	7.15 ± 0.19	0.31* (4.1%)
2000< LD_{50} ≤5000 mg/kg	10	10.73 ± 0.10	9.46 ± 0.15	1.28* (11.9%)	9.17 ± 0.23	7.96 ± 0.31	1.21* (13.2%)
LD_{50} >5000 mg/kg	12	9.87 ± 0.34	8.29 ± 0.49	1.58* (16.0%)	7.76 ± 0.59	6.18 ± 0.69	1.58* (20.3%)
		NHK NRU Test Method					
LD_{50} ≤5 mg/kg	6	11.21 ± 0.24	10.47 ± 0.71	0.75 (6.7%)	9.66 ± 0.27	8.95 ± 0.52	0.71 (7.3%)
5< LD_{50} ≤50 mg/kg	11	9.65 ± 0.16	9.99 ± 0. 45	-0.34 (-3.5%)	8.43 ± 0.26	8.77 ± 0.49	-0.33 (-3.9%)
50< LD_{50} ≤300 mg/kg	12	7.78 ± 0.11	8.12 ± 0.21	-0.34 (-4.4%)	6.57 ± 0.19	6.85 ± 0.19	-0.28 (-4.2%)
300< LD_{50} ≤2000 mg/kg	16	8.55 ± 0.22	8.03 ± 0.23	0.52* (6.1%)	7.49 ± 0.25	7.00 ± 0.20	0.49* (6.5%)
2000< LD_{50} ≤5000 mg/kg	10	10.75 ± 0.08	9.54 ± 0.20	1.21* (11.3%)	9.17 ± 0.23	8.06 ± 0.29	1.11* (12.1%)
LD_{50} >5000 mg/kg	13	9.87 ± 0.32	8.41 ± 0.44	1.47* (14.8%)	7.66 ± 0.59	6.18 ± 0.69	1.47* (19.2%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.

*Statistically significant ($p<0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses.

[1]Mean numbers of animals used ± standard errors for 10,000 simulations for each substance with an upper limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Substances were categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**.

[2]OECD (2001a); EPA (2002a).

[3]UN (2005).

[4]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

Table 10-4 Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

		Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose**	**Animals Saved[7]**
		3T3 NRU Test Method					
LD_{50} ≤5 mg/kg	6	11.29 ± 0. 20	10.38 ± 0.62	0.90 (8.0%)	9.70 ± 0.28	8.92 ± 0.37	0.78 (8.0%)
5 < LD_{50} ≤50 mg/kg	11	9.71 ± 0.22	9.58 ± 0.42	0.13 (1.3%)	8.47 ± 0.28	8.41 ± 0.44	0.06 (0.8%)
50 < LD_{50} ≤300 mg/kg	12	7.74 ± 0.10	7.99 ± 0.18	-0.25 (-3.3%)	6.58 ± 0.19	6.76 ± 0.18	-0.18 (-2.7%)
300 < LD_{50} ≤2000 mg/kg	16	8.52 ± 0.21	8.16 ± 0.19	0.35 (4.1%)	7.46 ± 0.24	7.17 ± 0.16	0.28* (3.8%)
2000 < LD_{50} ≤5000 mg/kg	10	10.78 ± 0.11	9.14 ± 0.24	1.64* (15.2%)	9.20 ± 0.24	7.61 ± 0.37	1.59* (17.3%)
LD_{50} >5000 mg/kg	12	9.87 ± 0.34	8.23 ± 0.48	1.65* (16.7%)	7.76 ± 0.59	6.14 ± 0.69	1.63* (21.0%)
		NHK NRU Test Method					
LD_{50} ≤5 mg/kg	6	11.21 ± 0.24	10.49 ± 0.71	0.72 (6.4%)	9.66 ± 0.27	8.97 ± 0.52	0.69 (7.1%)
5 < LD_{50} ≤50 mg/kg	11	9.70 ± 0.18	9.78 ± 0.41	-0.07 (-0.8%)	8.45 ± 0.27	8.59 ± 0.44	-0.13 (-1.6%)
50 < LD_{50} ≤300 mg/kg	12	7.75 ± 0.11	7.99 ± 0.21	-0.24 (-3.1%)	6.58 ± 0.19	6.76 ± 0.18	-0.18 (-2.7%)
300 < LD_{50} ≤2000 mg/kg	16	8.54 ± 0.21	8.20 ± 0.22	0.34 (3.9%)	7.48 ± 0.23	7.17 ± 0.16	0.31 (4.1%)
2000 < LD_{50} ≤5000 mg/kg	10	10.77 ± 0.08	9.40 ± 0.25	1.38*(12.8%)	9.18 ± 0.23	7.90 ± 0.33	1.28* (14.0%)
LD_{50} >5000 mg/kg	13	9.88 ± 0.32	8.34 ± 0.44	1.54*(15.6%)	7.66 ± 0.56	6.12 ± 0.63	1.53* (20.0%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.

*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.

[1]Mean number of animals used ± standard errors for 10,000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances for the 3T3 NRU test method and 68 substances for the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**.

[2]OECD (2001a); EPA (2002a).

[3]UN (2005).

[4]The RC rat-only weight regression is log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} values for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

.

477 10.2.4 Refinement of Animal Use for the UDP When Using 3T3- and NHK-Based
478 Starting Doses

479 A procedure refines animal use when it lessens or eliminates pain or distress in animals or
480 enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3-
481 and NHK-based starting doses refines animal use by reducing the number of animals that die
482 and experience accompanying pain and distress during UDP testing, compared to the number
483 of animals that die when the default starting dose of 175 mg/kg is used. **Table 10-5** reports
484 the results for the UDP simulation modeling using the 5000 mg/kg limit dose. For every
485 regression evaluated, the mean numbers of deaths when using the IC_{50}-based starting doses
486 were equal to, or slightly lower, than the mean number of deaths when using the default
487 starting dose. The percentage of deaths, however, was slightly higher for the IC_{50}-based
488 starting doses than for the default starting dose because the total number of animals used was
489 lower for the IC_{50}-based starting doses. Thus, fewer animals were used when using an IC_{50}-
490 based starting dose compared with use of the default starting dose, but the same numbers of
491 animals died.

492 10.2.5 Accuracy of UDP Outcomes Using the IC_{50}-Based Starting Doses

493 For each of the reference substances, the outcome of the simulated UDP testing, the
494 simulated LD_{50}, was used to classify the substance into a GHS acute oral toxicity category.
495 The accuracy of GHS toxicity category assignments using the IC_{50}-based starting doses was
496 determined by calculating the proportion of reference substances for which the GHS acute
497 oral toxicity category obtained using the IC_{50}-based starting dose matched the categories
498 obtained using the default starting dose.
499
500 The concordance between the GHS categories determined using the 3T3 and NHK NRU test
501 methods with the RC rat-only millimole regression, and those determined using the UDP
502 default starting dose, was 96% for 3T3 and 97% for NHK. The discordant reference
503 substances were acetaminophen and sodium dichromate dihydrate in the 3T3 NRU test
504 method, and acetaminophen, caffeine, and sodium dichromate dihydrate in the NHK NRU
505 test method. The use of the IC_{50}-based starting dose from both *in vitro* NRU test methods
506 resulted in a higher GHS category (i.e., higher simulated LD_{50}) for acetaminophen (simulated
507 LD_{50} = 2047 vs. 1765 mg/kg for 3T3, and LD_{50} = 2174 vs. 1755 mg/kg for NHK), and a

lower GHS category for sodium dichromate dihydrate (simulated LD_{50} = 44 vs. 52 mg/kg for 3T3 and LD_{50} = 45 vs. 52 mg/kg for NHK) than when using the default starting dose. The NHK-based starting dose resulted in a lower GHS category for caffeine (simulated LD_{50} = 280 vs. 357 mg/kg).

Table 10-5 Animal Deaths[1] in the UDP[2] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods

Assay/Regression	With Default Starting Dose[3]			With IC_{50}-Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	9.35	4.11	44.0%	8.80	4.09	46.5%
RC rat-only weight[6]	9.36	4.11	43.9%	8.70	4.05	46.6%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	7.95	3.44	43.3%	7.42	3.43	46.2%
RC rat-only weight[6]	7.94	3.43	43.2%	7.32	3.39	46.3%
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	9.36	4.08	43.6%	8.86	4.07	45.9%
RC rat-only weight[6]	9.36	4.08	43.6%	8.80	4.02	45.7%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	7.92	3.39	42.8%	7.43	3.39	45.6%
RC rat-only weight[6]	7.92	3.39	42.8%	7.38	3.35	45.4%

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.
[1]Numbers are mean numbers of animals used for 10,000 simulations for each substance. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose = 5000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test methods.
[2]OECD (2001a); EPA (2002a).
[3]Default starting dose = 175 mg/kg.
[4]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value in the regression specified. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[5]$\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.
[6]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024.

The concordance of GHS acute toxicity category predictions with those determined using the default starting dose was 97% for the 3T3 and NHK NRU test methods when the RC rat-only weight regression was used. The discordant reference substances were caffeine and sodium dichromate dihydrate. The simulated LD_{50} outcome for caffeine was lowered from 338 mg/kg for the default starting dose to 272 mg/kg for the 3T3-based starting dose, and from 339

536 mg/kg to 270 mg/kg for the NHK-based starting dose. The simulated LD_{50} outcome for
537 sodium dichromate dihydrate was lowered from 51 mg/kg for the default starting dose to 48
538 mg/kg for the 3T3-based starting dose, and from 51 mg/kg to 49 mg/kg for the NHK-based
539 starting dose.
540
541 Thus, the use of the IC_{50}-based starting doses did not significantly alter the outcome of the
542 simulated UDP tests compared with the outcome obtained using the default starting doses.

543 10.3 Reduction and Refinement of Animal Use in the ATC Method

544 10.3.1 *In Vivo* Testing Using the ATC Method

545 This section describes the general dosing procedure for the conduct of the ATC procedure
546 (OECD 2001d). The ATC is used to assign a test substance to the appropriate GHS category
547 for classification and labeling. This is done by estimating the range of the LD_{50} values for the
548 test substance, rather than calculating a point estimate of the LD_{50}. The time between
549 administration of test substance doses is determined by the onset, duration, and severity of
550 toxic signs. Guidance on the types of animals to use, animal housing, clinical observations,
551 etc., which are outside the scope of the current discussion, are provided in the test guideline
552 (See **Appendix M**).

553 10.3.1.1 *Main Test*

554 The ATC method uses a stepwise administration of test substances to three animals at a time,
555 at one of a number of fixed doses: 5, 50, 300, and 2000 mg/kg (and 5000 mg/kg, if
556 necessary). The starting dose is selected so that at least some of the animals die at that dose.
557 If no information on which to base a starting dose is available, a default starting dose of 300
558 mg/kg is used. The next step is determined by the starting dose and the outcome of the three
559 animals tested at the starting dose and may be a decision to stop testing, test additional
560 animals at the same dose, test at the next higher dose, or test at the next lower dose. For
561 example, if two to three animals die or are in a moribund state after receiving the 300 mg/kg
562 starting dose, the next step is to administer 50 mg/kg to three more animals. However, if no,
563 or one, animal dies at 300 mg/kg, three additional animals are tested at that dose. Most
564 substances require two to four dosing steps before they can be classified, and testing can be
565 stopped. See **Appendix M** for the outcome-based testing sequence for each starting dose.

566 10.3.1.2 *Limit Test*

567 For test substances that are likely to be nontoxic, the ATC guideline includes a limit test in
568 which six animals (three animals per step [see **Appendix M3**]) are tested at the limit dose of
569 2000 mg/kg or 5000 mg/kg (OECD 2001d).

570 10.3.2 Computer Simulation Modeling of the ATC Method

571 The simulation for the ATC method was performed using MATLAB® (The MathWorks, Inc.
572 1996-2004) computational software, which is functionally comparable with SAS® version 8.
573 Two thousand simulations of ATC testing were run for each substance, *in vitro* NRU test
574 method, and dose-mortality slope, using an upper limit dose of 2000 mg/kg. The simulation
575 implements the distributional assumptions underlying the dose-mortality response. The
576 lowest dose at which an animal dies in response to the administration of a toxic substance
577 varies from animal to animal. For an entire population of animals, mortality is assumed to
578 have a log-normal distribution with the mean equal to the log of the true LD_{50}. Sigma (σ), the
579 variability of the simulated population, is the inverse of the slope of the dose-mortality curve.
580 For any given dose, the probability that an animal will die is computed by the cumulative
581 log-normal distribution:
582

583 $$\text{Probability (death)} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log dose} e^{\frac{-(t-\log trueLD_{50})^2}{2\sigma^2}} dt$$

584

585 Because of a lack of information regarding the real dose-mortality curves, the simulations
586 assumed several different values of the slope (i.e., the inverse of σ). Dose-mortality slopes of
587 0.5, 0.8, 2.0, 4.0, and 8.3 were chosen, so as to be comparable to the slopes chosen for
588 simulation modeling of the UDP (see **Section 10.2.2**).
589

590 To model the variability of the IC_{50} values within and among laboratories, the values for each
591 substance were log-transformed to normalize their distribution. The mean and variance of
592 these log-transformed values were used to generate a log-normal distribution from which to
593 randomly select an IC_{50} value.
594

595 The simulation procedure used the following steps for each substance:

1. The rodent acute oral LD_{50} value (in mg/kg) from **Table 4-2** was entered as the true LD_{50} value and the choices of assumed slope were entered as the true slope for the dose-mortality curve.
2. An IC_{50} value was selected from a distribution identified by the mean and variance of the IC_{50} values computed from the data to reflect that different laboratories produce different IC_{50} values in different situations (see **Tables 5-4** and **5-5** for mean IC_{50} values and standard deviations for the 3T3 and NHK NRU test methods, respectively).
3. The IC_{50} value from Step 2 was used in the regression model being evaluated to compute a predicted LD_{50} value to use as the starting dose.
4. The dosing simulation (of 2000 iterations) was run twice: once with the default starting dose of 300 mg/kg and once with a starting dose equal to the next fixed dose below the LD_{50}, which was estimated by the regression being evaluated (i.e., the IC_{50}-based starting dose). If the IC_{50}-based starting dose was greater than the 2000 mg/kg limit dose, then testing proceeded using the 2000 mg/kg limit test rather than the main test.
5. For every dose group of three animals, one observation was sampled from a binomial distribution with the probability of death calculated by the probability equation for a population of three. The sampled value, referred to as N1, indicates the number of animals, 0, 1, 2, or 3, in the dosing group that die.
6. If N1 $\leq$1, step 4 is repeated with the same dose. The resulting sampled value from the binomial distribution is referred to as N2.
7. If N2 $\leq$1 and the dose is the highest dose tested, or the dose has already been decreased, a toxicity category is assigned and testing is terminated. If the dose is not the highest dose tested, or if the dose has not been decreased, the next higher fixed dose is administered and step 4 is repeated.
8. If N1 >1 or N2 >2, and the dose is the lowest dose tested, or if the dose has already been increased, a toxicity category is assigned and testing is terminated. If the dose is not the lowest dose tested, or if the dose has not

626 already been increased, the next lower fixed dose is administered and step 4 is
627 repeated.

628 10.3.3 Animal Savings for the ATC Method When Using 3T3- and NHK-Based Starting
629 Doses

630 10.3.3.1 *The Effect of the Dose-Mortality Slope on Animal Use*

631 As described in **Section 10.3.2**, the simulation modeling of animal use for the ATC used five
632 different dose-mortality slopes to assess animal use under various conditions of population
633 variability. **Table 10-6** shows how animal use for the simulated ATC changes with dose-
634 mortality slope, and the mean animal use when using the default starting dose of 300 mg/kg
635 and a starting dose that was one fixed dose lower than that predicted by the 3T3 and NHK
636 NRU IC_{50} values with the RC rat-only millimole regression. The mean number of animals
637 used for the ATC method decreased slightly with increasing slope for both the default
638 starting dose and the IC_{50}-based starting dose.
639
640 The mean numbers of animals saved at all dose-mortality slopes were statistically significant
641 ($p < 0.05$ by one-sided Wilcoxon signed rank tests) when compared with mean animal use
642 with the default dose, and tended to decrease with increasing slope. To simplify the
643 presentation of animal savings and comparisons of the various regressions and starting doses,
644 subsequent results in **Section 10.3.3** are shown only for dose-mortality slopes of 2.0 and 8.3.
645 As stated earlier, these slopes are shown here because the slope of 2.0 is the default used for
646 the calculation of LD_{50} by the UDP method (OECD 2001a; EPA 2002a) and the slope of 8.3
647 is shown to represent substances, such as pesticides, with higher slopes. Results for the other
648 dose-mortality slopes were computed, and are presented in **Appendices N4-N6.**
649

649 **Table 10-6 Change in Animal Use[1] with Dose-Mortality Slope in the ATC Method[2]**
650

Dose-Mortality Slope	Default Starting Dose[1,3]	IC_{50}- Based Starting Dose[1,4]	Animals Saved[5]
3T3 NRU Test Method			
0.5	11.25 ± 0.05	10.56 ± 0.17	0.69* (6.1%)
0.8	11.10 ± 0.07	10.46 ± 0.19	0.64* (5.8%)
2.0	10.89 ± 0.12	10.27 ± 0.24	0.62* (5.7%)
4.0	10.73 ± 0.15	10.15 ± 0.26	0.58* (5.4%)
8.3	10.64 ± 0.17	10.13 ± 0.27	0.51* (4.8%)
NHK NRU Test Method			
0.5	11.25 ± 0.05	10.43 ± 0.16	0.82* (7.3%)
0.8	11.10 ± 0.07	10.31 ± 0.18	0.79* (7.1%)
2.0	10.91 ± 0.11	10.11 ± 0.24	0.80* (7.3%)
4.0	10.75 ± 0.15	9.98 ± 0.27	0.77* (7.1%)
8.3	10.67 ± 0.17	9.96 ± 0.29	0.70* (6.6%)

651 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human
652 keratinocytes; RC=Registry of Cytotoxicity.
653 *Statistically significant ($p < 0.05$) by a one-sided Wilcoxon rank test. Percent difference is shown in parentheses.
654 [1]Mean numbers of animals used ± standard errors for 2000 simulations each for 67 substances in the 3T3 NRU test
655 method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the
656 results over a large number of simulations produced fractional numbers. Upper limit dose = 2000 mg/kg.
657 [2]OECD (2001d).
658 [3]Default starting dose = 300 mg/kg.
659 [4]Next fixed dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-
660 only millimole regression: log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621. The IC_{50} value for each reference substance
661 was randomly selected from the distribution of values obtained during the testing with each method.
662 [5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.
663
664

665 10.3.3.2 *Mean Animal Use for ATC Simulations – Comparison of Regressions and*
666 *Predictions from the 3T3 and NHK NRU Test Methods*

667 **Table 10-7** shows the mean animal use for testing the reference substances using the
668 simulated ATC method, when the starting dose was the default starting dose and when the
669 starting dose was one fixed dose lower than that determined by the 3T3 and NHK-predicted
670 LD_{50}, and the regressions evaluated in **Section 6.4** for prediction of GHS category. The mean
671 difference in animal use between the two starting doses is the mean animal savings. All mean
672 animal savings were statistically significant ($p < 0.05$ using one-sided Wilcoxon signed rank
673 tests), and ranged from 0.51 (4.8%) to 1.09 (10.2%) animals per test depending upon the
674 NRU test method, regression, and dose-mortality slope. The lowest mean animal savings
675 were obtained for the RC rat-only millimole regression (0.51 [4.8%] to 0.80 [7.3%] animals

per test), and the highest were obtained with the RC rat-only weight regression (0.91 [8.6%] to 1.09 [10.2%] animals per test).

The animal savings obtained using the *in vitro* NRU test methods with the RC rat-only regressions apply only to the reference substances evaluated in this validation study, and are based on substances pre-selected for their known *in vivo* toxicities and fit to the RC regression and may not be broadly applicable to other substances. T**able 3-4** shows that 22 (38%) of the 58 RC substances selected for testing were known to have a poor fit the RC millimole regression (i.e., the predicted LD_{50} was outside the RC acceptance interval). **Table 6-2** shows that 43% (30/70 in the 3T3) and 44% (31/71 in the NHK) of the reference substances that yielded IC_{50} values were outliers. The findings of the RC rat-only millimole regression evaluated here are similar to the RC millimole regression (see **Table 6-4**) findings. Substances that better fit the regression are likely to yield greater animal savings.

691 **Table 10-7 Animal Use[1] for the ATC[2] Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with**
692 **Both Regressions**
693

Method/Regression	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[4]	Animals Saved[5]	With Default Starting Dose[3]	With IC_{50}-Based Starting Dose[5]	Animals Saved[5]
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**			**Dose-Mortality Slope = 8.3**		
RC rat-only millimole[6]	10.89 ± 0.12	10.27 ± 24	0.62* (5.7%)	10.64 ± 0.17	10.13 ± 0.27	0.51* (4.8%)
RC rat-only weight[7]	10.89 ± 0.12	9.85 ± 0.24	1.04* (9.6%)	10.64 ± 0.17	9.55 ± 0.29	1.09* (10.2%)
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**			**Dose-Mortality Slope = 8.3**		
RC rat-only millimole[6]	10.91 ± 0.11	10.11 ± 0.24	0.80* (7.3%)	10.67 ± 0.17	9.96 ± 0.29	0.70* (6.6%)
RC rat-only weight[7]	10.91 ± 0.11	9.95 ± 0.24	0.96* (8.8%)	10.67 ± 0.17	9.75 ± 0.30	0.91* (8.6%)

694 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.
695 *Statistically significant ($p<0.05$) using a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
696 [1]Mean numbers of animals used ± standard errors for 2000 simulations each for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU
697 test method. Upper limit dose =2000 mg/kg.
698 [2]OECD (2001d).
699 [3]Default starting dose =300 mg/kg.
700 [4]Starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the regression specified. The IC_{50}
701 value for each reference substance was randomly selected from the distribution of values obtained during the testing with each test method.
702 [5]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.
703 [6]log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
704 [7]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.

705 10.3.3.3 *Animal Savings in the ATC Method by GHS Acute Oral Toxicity Category Using*
706 *the 3T3- and NHK -Based Starting Doses*

707 **Tables 10-8** and **10-9** show mean animal use and mean animal savings for the ATC when
708 used with the *in vitro* NRU test methods, organized by GHS category (UN 2005), and when
709 based on the:

710 - RC rat-only millimole regression (**Table 10-8**)
711 - RC rat-only weight regression (**Table 10-9**)

712 The following data come from the same analyses as the data provided in **Table 10-7**.
713
714 The analyses showed that:

715 - For each *in vitro* NRU test method and regression, the highest mean animal
716 savings were generally in the LD_{50} ≤5 mg/kg and LD_{50} >5000 mg/kg toxicity
717 categories.
718 - For each NRU test method and regression, the lowest mean animal savings
719 were in the 300 < LD_{50} ≤2000 mg/kg toxicity category.

720
721 *Animal Savings in the ATC Method by GHS Category Using the 3T3- and NHK-Based*
722 *Starting Doses with the RC Rat-Only Millimole Regression*

723 **Table 10-8** shows the mean animal savings in the ATC method by GHS category for the *in*
724 *vitro* NRU test methods used with the RC rat-only millimole regression. Mean differences
725 between animal use for the default starting dose and with the IC_{50}-determined starting dose
726 were statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test for the
727 following GHS toxicity categories, NRU test methods, and dose-mortality slopes:

728 - The use of both *in vitro NRU* test methods at both dose-mortality slopes for
729 substances with 5 < LD_{50} ≤50 mg/kg produced savings of 1.15 (9.8%) to 1.33
730 (11.4%) animals per test
731 - The use of the 3T3 NRU test method at both dose-mortality slopes for
732 substances with 300 < LD_{50} ≤2000 mg/kg produced savings of -0.92 (-9.5%)
733 to -1.30 (-14.0%) animals per test

- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with LD_{50} >5000 mg/kg produced savings of 2.03 (17.1%) to 2.66 (22.2%) animals per test

At the dose-mortality slope of 2.0, the mean animal savings with the 3T3 NRU test method ranged from -0.92 (-9.5%) to 2.68 (27.4%) animals per test, and the animal savings with the NHK NRU test method ranged from -0.60 (-6.1%) to 2.96 (30.4%) animals per test. At the dose-mortality slope of 8.3, the mean animal savings with the 3T3 NRU test method ranged from -1.30 (-14.0%) to 2.70 (29.7%) animals per test, and the animal savings with the NHK NRU test method ranged from -0.85 (-9.2%) to 2.99 (33.0%) animals per test.

At both the 2.0 and 8.3 dose-mortality slopes, the mean animal savings using the 3T3 NRU test method were lower than the corresponding savings using the NHK NRU test method, for substances in four of the six toxicity categories: $LD_{50} \leq 5$ mg/kg; $5 < LD_{50} \leq 50$ mg/kg; $300 < LD_{50} \leq 2000$ mg/kg; and $LD_{50} > 5000$ mg/kg. The mean animal savings per test were higher with the 3T3 NRU test method than the NHK NRU test method for substances in the $2000 < LD_{50} \leq 5000$ mg/kg category at both dose-mortality slopes. For substances in the $50 < LD_{50} \leq 300$ mg/kg category, the mean animal savings using the 3T3 NRU test method was greater than the savings using the NHK NRU test method, when the dose-mortality slope equaled 8.3. When the 3T3 NRU test method was used, the highest mean animal savings occurred when testing substances in the $LD_{50} \leq 5$ mg/kg category (2.68 [27.4%] animals per test at dose-mortality slope = 2.0, and 2.70 [29.7%] at dose-mortality slope = 8.3). When the NHK NRU test method was used, the highest mean animal savings occurred when testing substances in the $LD_{50} \leq 5$ mg/kg category (2.96 [30.4%] animals per test at dose-mortality slope = 2.0, and 2.99 [33.0%] animals per dose at dose-mortality slope = 8.3). However, the animal savings were not statistically significant with either *in vitro* NRU test method.

Table 10-8 Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

		Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**WithIC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**
		3T3 NRU Test Method					
LD_{50} ≤5 mg/kg	6	9.77 ± 0.17	7.09 ± 1.09	2.68 (27.4%)	9.08 ± 0.08	6.38 ± 1.09	2.70 (29.7%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.39 ± 0.52	1.17* (10.2%)	11.75 ± 0.16	10.60 ± 0.43	1.15* (9.8%)
50 < LD_{50} ≤300 mg/kg	12	10.81 ± 0.20	10.39 ± 0.17	0.42 (3.9%)	9.42 ± 0.26	9.27 ± 0.11	0.15 (1.6%)
300 < LD_{50} ≤2000 mg/kg	16	9.75 ± 0.07	10.67 ± 0.48	-0.92* (-9.5%)	9.26 ± 0.10	10.56 ± 0.62	-1.30* (-14.0%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	11.14 ± 0.08	0.08 (0.7%)	11.88 ± 0.10	11.77 ± 0.10	0.11 (0.9%)
LD_{50} >5000 mg/kg	12	11.85 ± 0.04	9.82 ± 0.78	2.03* (17.1%)	12.00 ± 0.000	9.81 ± 0.84	2.19* (18.3%)
		NHK NRU Test Method					
LD_{50} ≤5 mg/kg	6	9.74 ± 0.16	6.78 ± 1.31	2.96 (30.4%)	9.09 ± 0.08	6.09 ± 1.23	2.99 (33.0%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.38 ± 0.35	1.18* (10.2%)	11.76 ± 0.17	10.42 ± 0.45	1.33* (11.4%)
50 < LD_{50} ≤300 mg/kg	12	10.83 ± 0.21	10.39 ± 0.29	0.44 (4.0%)	9.44 ± 0.26	9.63 ± 0.49	-0.20 (-2.1%)
300 < LD_{50} ≤2000 mg/kg	16	9.77 ± 0.06	10.37 ± 0.49	-0.60 (-6.1%)	9.26 ± 0.10	10.11 ± 0.63	-0.85 (-9.2%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	11.25 ± 0.12	-0.03 (-0.3%)	11.87 ± 0.10	11.89 ± 0.15	-0.02 (-0.2%)
LD_{50} >5000 mg/kg	13	11.86 ± 0.03	9.43 ± 0.73	2.43* (20.5%)	12.00 ± 0.000	9.34 ± 0.80	2.66* (22.2%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.
*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.
[1]Mean number of animals used ± standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral LD_{50} reference values in mg/kg from **Table 4-2**. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers.
[2]OECD (2001d).
[3]GHS for acute oral toxicity (UN 2005).
[4]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[5]Default starting dose =300 mg/kg.
[6] The starting dose was the next fixed dose lower than the predicted LD_{50} using the IC_{50} for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.
.

The smallest mean animal savings (≤0.44) in both *in vitro* NRU test methods were observed for substances with LD_{50} values between 50 and 5000 mg/kg. Because the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the 50 < LD_{50} ≤300 mg/kg and 300 < LD_{50} ≤2000 mg/kg categories. The mean animal savings from both *in vitro* NRU test methods and both dose-mortality slopes for the substances in the 50 < LD_{50} ≤300 mg/kg category were -0.20 to 0.44 animals per test. There were no animal savings for substances in the 300 < LD_{50} ≤2000 mg/kg category using either NRU test method or dose-mortality slope. More animals were used when the starting doses were based on the IC_{50} than using the default starting dose (-0.60 to -1.30 animals per test).

The animal savings in the various GHS acute oral toxicity categories using the *in vitro* NRU test methods with the RC rat-only millimole regression applies only to the reference substances evaluated in this validation study, and may not be broadly applicable to other substances. The animal savings for future testing using the *in vitro* NRU test methods with the RC rat-only millimole regression will depend on the prevalence of test substances in each of the GHS acute oral toxicity categories.

Animal Savings with the ATC Method by GHS Category Using 3T3- and NHK-Based Starting Doses with the RC Rat-Only Weight Regression

Table 10-9 shows the animal savings for the simulation ATC method by GHS category for the *in vitro* NRU methods used with the RC rat-only weight regression. Mean animal savings were statistically significant ($p < 0.05$) by a one-tailed Wilcoxon signed rank test for the following GHS toxicity categories, NRU test methods, and dose-mortality slopes.

- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with 5 < LD_{50} ≤50 mg/kg produced savings of 1.25 (10.8%) to 1.51 (13.0%) animals per test.
- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with 50 < LD_{50} ≤300 mg/kg produced savings of 0.47 (4.3%) animals per test.

- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with 300 < LD_{50} ≤2000 mg/kg produced savings of -0.93 (-9.5%) animals per test.
- The use of the 3T3 NRU test method at the 2.0 dose-mortality slope for substances with 2000 < LD_{50} ≤5000 mg/kg produced savings of 1.43 (12.7%) animals per test.
- The use of both *in vitro* NRU test methods at both dose-mortality slopes for substances with LD_{50} >5000 mg/kg produced savings of 2.94 (24.8%) to 3.33 (27.7%) animals per test.

The mean animal savings with the 3T3 and NHK NRU test methods were similar for most acute oral toxicity categories at both dose-mortality slopes; the mean savings for the 3T3 NRU test method was slightly higher than for the NHK NRU test method. At the dose-mortality slope of 2.0, the mean animal savings for the 3T3 NRU test method (for the various toxicity categories) ranged from -0.93 (-9.5%) to 3.02 (25.5%) animals per test, and savings for the NHK NRU test method ranged from -0.69 (-7.1%) to 2.94 (24.8%) animals per test. At the dose-mortality slope of 8.3, animal savings with the 3T3 NRU test method ranged from -1.39 (-15.0%) to 3.33 (27.7%) animals per test, and savings with the NHK NRU test method ranged from -0.97 (-10.4%) to 3.25 (27.1%) animals per test.

There were no mean animal savings (≤-0.69 animals) for substances with 300 < LD_{50} ≤2000 when either *in vitro* NRU test method was used. The mean animal savings for the substances in the 50 < LD_{50} ≤300 mg/kg category using both *in vitro* NRU test methods and dose-mortality slopes were also relatively small (-0.20 to 0.47 animals per test). Because the default starting dose was 300 mg/kg, little change in mean animal use was expected for substances in the 50< LD_{50} ≤300 mg/kg and 300 < LD_{50} ≤2000 mg/kg categories.

Table 10-9 Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

		Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**
		3T3 NRU Test Method					
LD_{50} ≤5 mg/kg	6	9.77 ± 0.17	7.56 ± 1.03	2.21 (22.6%)	9.08 ± 0.08	6.85 ± 0.99	2.24 (24.6%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.06 ± 0.38	1.51* (13.0%)	11.75 ± 0.16	10.27 ± 0.33	1.48* (12.6%)
50 < LD_{50} ≤300 mg/kg	12	10.81 ± 0.20	10.35 ± 0.18	0.47* (4.3%)	9.42 ± 0.26	9.20 ± 0.10	0.22 (2.4%)
300 < LD_{50} ≤2000 mg/kg	16	9.75 ± 0.07	10.67 ± 0.50	-0.93* (-9.5%)	9.26 ± 0.10	10.65 ± 0.66	-1.39 (-15.0%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	9.80 ± 0.51	1.43* (12.7%)	11.88 ± 0.10	9.44 ± 0.88	2.43 (20.5%)
LD_{50} >5000 mg/kg	12	11.85 ± 0.04	8.83 ± 0.83	3.02* (25.5%)	12.00 ± 0.00	8.67 ± 0.91	3.33* (27.7%)
		NHK NRU Test Method					
LD_{50} ≤5 mg/kg	6	9.74 ± 0.16	6.87 ± 1.28	2.87 (29.4%)	9.09 ± 0.08	6.18 ± 1.20	2.91 (32.0%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.31 ± 0.19	1.25* (10.8%)	11.76 ± 0.17	10.40 ± 0.33	1.36* (11.5%)
50 < LD_{50} ≤300 mg/kg	12	10.83 ± 0.21	10.41 ± 0.28	0.42 (3.8%)	9.44 ± 0.26	9.63 ± 0.49	-0.20 (-2.1%)
300 < LD_{50} ≤2000 mg/kg	16	9.77 ± 0.62	10.46 ± 0.50	-0.69 (-7.1%)	9.26 ± 0.10	10.23 ± 0.65	-0.97 (-10.4%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.09	10.69 ± 0.37	0.53 (4.7%)	11.87 ± 0.10	11.03 ± 0.60	0.84 (7.1%)
LD_{50} >5000 mg/kg	13	11.86 ± 0.03	8.91 ± 0.78	2.94* (24.8%)	12.00 ± 0.00	8.75 ± 0.85	3.25* (27.1%)

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.

*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

[1]Mean number of animals used ± standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the rat acute oral reference LD_{50} values in mg/kg from **Table 4-2**.

[2]OECD (2001d).

[3]GHS for acute oral toxicity (UN 2005).

[4]From **Table 6-2;** $\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024

[5]Default starting dose = 300 mg/kg.

[6] The starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

The animal savings in the various GHS categories using the two *in vitro NRU* test methods with the RC rat-only weight regression applies only to the reference substances evaluated in this validation study, and may not be broadly applicable to other substances.

10.3.4 Refinement of Animal Use in the ATC Method When Using 3T3- and NHK-Based Starting Doses

A procedure refines animal use when it lessens or eliminates pain or distress in animals, or enhances animal well-being (ICCVAM 2003). This section evaluates whether the use of 3T3- and NHK-based starting doses refines animal use by reducing the number of animals that die when the IC_{50}-predicted starting doses are used, compared to the number of animals that die when using the default ATC starting dose of 300 mg/kg. **Table 10-10** reports the results for the ATC simulation modeling using the 2000 mg/kg limit dose. For every regression evaluated, the mean number of deaths when using the 3T3- and NHK-based starting doses was less than the mean number of deaths when using the default starting dose, by approximately 0.4 to 0.5 deaths per test. For the RC rat-only millimole regression and the RC rat-only weight regression, the percentage of deaths (compared with the numbers of animals used) was also slightly lower with the *in vitro*-based starting dose compared with the default starting dose. In general, fewer animals were used with the *in vitro*-based starting dose, and fewer animals died.

10.3.5 Accuracy of the ATC Method Outcomes Using the IC_{50}-Based Starting Doses

The accuracy of the outcome of the simulated ATC testing using the simulated GHS acute oral toxicity category and the IC_{50}-based starting dose was determined by calculating the proportion of reference substances for which the simulated GHS category for the IC_{50}-based starting dose matched the simulated GHS category for the default starting dose.

When RC rat-only millimole regression with the 3T3 and NHK NRU test methods was used, the concordance of simulated GHS categories for the IC_{50}-based starting doses with those for the default starting dose was 99% for both *in vitro* NRU test methods. The discordant reference substance in the 3T3 NRU test method was caffeine. The simulated GHS category using the 3T3 -based starting dose was $50 < LD_{50} \leq 300$ mg/kg, and the default starting dose was $300 < LD_{50} \leq 2000$ mg/kg.

Table 10-10 Animal Deaths[1] for the ATC[2] Method Using Starting Doses Based on the 3T3 and NHK NRU Test Methods

Method/Regression	Default Starting Dose[3]			IC_{50}- Based Starting Dose[4]		
	Used	Dead	% Deaths	Used	Dead	% Deaths
3T3 NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	10.89	3.77	34.6%	10.27	3.31	32.2%
RC rat-only weight[6]	10.89	3.77	34.6%	9.85	3.27	33.2%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	10.64	3.20	30.1%	10.13	2.77	27.3%
RC rat-only weight[6]	10.64	3.20	30.1%	9.55	2.73	28.6%
NHK NRU Test Method	**Dose-Mortality Slope = 2.0**					
RC rat-only millimole[5]	10.91	3.72	34.1%	10.11	3.19	31.6%
RC rat-only weight[6]	10.91	3.72	34.1%	9.95	3.21	32.3%
	Dose-Mortality Slope = 8.3					
RC rat-only millimole[5]	10.67	3.15	29.5%	9.96	2.67	26.8%
RC rat-only weight[6]	10.67	3.15	29.5%	9.75	2.67	27.4%

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.
[1]Mean numbers of animals used for 2000 simulations for each of 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Upper limit dose =2000 mg/kg.
[2]OECD (2001d).
[3]Default starting dose =300 mg/kg.
[4] The starting dose was one fixed dose lower than the predicted LD_{50} calculated by using the IC_{50} for each reference substance in the regression evaluated. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.
[5]log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.
[6]log LD_{50} (mg/kg) = 0.372 log IC_{50} (μg/mL) + 2.024.

The discordant reference substance in the NHK NRU test method was sodium dichromate dihydrate. The simulated GHS acute oral toxicity category using the NHK-based starting dose was $5 < LD_{50} \leq 50$ mg/kg and the default starting dose was $50 < LD_{50} \leq 300$ mg/kg; i.e., both discordant substances were predicted to have a starting dose one category below the actual category.

When the RC rat-only weight regression was used with the 3T3 and NHK NRU test methods, the concordance of simulated GHS acute toxicity category predictions with those determined using the default starting dose was 99% and 97% for the 3T3 and the NHK NRU test methods, respectively. The discordant reference substance in the 3T3 NRU test method was

909 caffeine. The simulated GHS acute oral toxicity category for caffeine using the 3T3-based
910 starting dose was 50 < LD_{50} ≤300 mg/kg and the default starting dose was 300 < LD_{50} ≤2000
911 mg/kg. The discordant reference substances in the NHK NRU test method were caffeine and
912 sodium dichromate dihydrate. The simulated GHS acute oral toxicity category for caffeine
913 using the NHK-based starting dose was 50 < LD_{50} ≤300 mg/kg and the default starting dose
914 was 300 < LD_{50} ≤2000 mg/kg. The simulated GHS acute oral toxicity category for sodium
915 dichromate dihydrate using the NHK-based starting dose was 5 < LD_{50} ≤50 mg/kg while that
916 for the default starting dose was 50 < LD_{50} ≤300 mg/kg. Similar to what was seen with the
917 RC millimole regression, the predicted starting doses for the discordant substances were one
918 GHS category below the actual category.
919
920 Thus, the use of the IC_{50}-based starting doses did not significantly alter the outcomes of the
921 simulated ATC tests compared with the outcome based on the default starting dose.

922 **10.4 The Impact of Accuracy on Animals Savings**

923 Two types of accuracy analyses were performed for the validation study. The first analyses
924 determined the accuracy of using the NRU IC_{50} values with an IC_{50}-LD_{50} regression to
925 predict LD_{50} values. It calculated the concordance for GHS acute oral toxicity category by
926 comparing the GHS categorization yielded by the NRU-predicted LD_{50} values (using the *in*
927 *vitro* NRU IC_{50} values in the regressions presented in **Table 6-4)** with the GHS
928 categorization based on rat acute oral LD_{50} data (see **Section 6.4**). The second analysis
929 determined the accuracy of the simulation outcomes using the IC_{50}-based starting doses (see
930 **Sections 10.2.5** and **10.3.5**). It calculated the concordance for the GHS acute oral toxicity
931 category outcomes obtained using the IC_{50}-based starting doses with the GHS category
932 outcomes obtained using the default starting dose. The magnitude of animal savings did not
933 correlate with either determination of accuracy and the accuracy determinations for IC_{50}-
934 based predictions and IC_{50}-based outcomes for GHS category did not correlate with one
935 another.
936
937 Animal savings did not correlate with the accuracy of the GHS acute oral toxicity category
938 predictions based on the LD_{50} values calculated using the IC_{50} values in the RC rat-only
939 regressions (see **Section 6.4.1**). Substances in categories with the lowest accuracy produced

940 the highest animal savings. For example, using the RC rat-only millimole regression with the
941 *in vitro* NRU IC_{50} values yielded very low accuracy (0 to 17%) for GHS acute oral toxicity
942 category prediction for substances with LD_{50} >5000 mg/kg (see **Table 6-5**), but the highest
943 animal savings of 14.8 to 20.3% occurred in this category (see **Table 10-3**). Animal savings
944 were small, 4.5 to 6.5%, for substances with $300 \leq LD_{50} \leq 2000$ mg/kg, but the accuracy of
945 75-81% for GHS acute oral toxicity category prediction was relatively high. The reason that
946 animal savings is unrelated to the accuracy of prediction of GHS acute oral toxicity category
947 based on the LD_{50} values calculated using IC_{50} values in the RC rat-only regressions is
948 because two different standards are used for comparison in the two analyses:

949 • GHS acute oral toxicity category prediction are compared with the GHS
950 categories derived from the *in vivo* reference LD_{50}
951 • The number of animals used (to determine animal savings) was compared
952 with the animal use at the default starting dose of 175 mg/kg for the UDP or
953 300 mg/kg for the ATC

954 Despite the relatively poor GHS accuracy for the low toxicity chemicals (the toxicity of
955 almost all were overpredicted by one GHS category), animal savings were greatest due to the
956 fact that testing goes to the limit dose faster.
957
958 The accuracy of GHS toxicity category assignments using the IC_{50}-based starting doses for
959 UDP and ATC test simulations was determined by calculating the proportion of reference
960 substances for which the GHS acute oral toxicity category obtained using the IC_{50}-based
961 starting dose matched the categories obtained using the default starting dose (see **Sections**
962 **10.2.5** and **10.3.5**). The accuracy of these GHS toxicity category assignments based on the
963 simulation outcomes does not correlate with animal savings using the IC_{50} values in the RC
964 rat-only regressions (see **Section 6.4.1** and **6.4.2**). For example, the accuracy of GHS acute
965 oral toxicity category outcomes for the ATC test method when using the RC rat-only
966 millimole regression was 100% for the 3T3 NRU test method for substances with $300 \leq LD_{50}$
967 ≤ 2000 mg/kg (see **Appendix N3**). In contrast, the animal savings for those substances was
968 negative at -6.1 to -14.0% (i.e., more animals were used compared with the default starting
969 dose) (see **Table 10-8**). The reason the outcome-based GHS acute oral toxicity category

predictions is unrelated to animal savings is that two different parameters are being measured in the two analyses:

- The accuracy of GHS acute oral toxicity outcomes using the IC_{50}-based starting doses measured outcome (i.e., GHS category based on LD_{50} outcome for the UDP and GHS category for the ATC)
- The animal savings analysis measured the number of animals used at the IC_{50}-based starting dose and the default starting dose of 175 mg/kg for the UDP or 300 mg/kg for the ATC

Thus, the measurements for the two analyses are different: outcome (i.e., GHS category) and number of animals used to achieve the outcome.

In addition, accuracy of the GHS toxicity category assignments based on the simulation outcomes does not correlate with the accuracy of the GHS acute oral toxicity category predictions using the IC_{50} values in the RC rat-only regressions (see **Section 6.4.1** and **6.4.2**). For example, the overall accuracy of GHS acute oral toxicity category outcomes for the ATC test method when using the RC rat-only millimole regression was 99% for both *in vitro* NRU test methods (see **Section 10.3.5** and **Appendix N3**). In contrast, the overall accuracy of GHS acute oral toxicity category predictions using the using the IC_{50} values in the RC rat-only millimole regression was 31% for the 3T3 NRU test method and 29% for the NHK NRU test method (see **Table 6.5**). The reason the outcome-based GHS acute oral toxicity category predictions differed from the accuracy of the GHS acute oral toxicity category predictions based on the calculation of LD_{50} using the IC_{50} in the IC_{50}-LD_{50} regression is because two different standards are used for comparison in the two analyses:

- GHS acute oral toxicity outcomes for the IC_{50}-based starting doses were compared with the GHS category outcomes using the default starting doses
- GHS acute oral toxicity category predictions using the IC_{50} values in the RC rat-only regressions were compared with the GHS category derived from the *in vivo* reference LD_{50}

Thus, despite that the IC_{50} values and IC_{50}-LD_{50} regressions predicted GHS acute oral toxicity categories poorly, the GHS acute oral toxicity category outcomes using the IC_{50}-

1000 based starting doses were practically the same as the GHS acute oral toxicity category
1001 outcomes using the default starting dose.

1002 **10.5 The Impact of Prevalence on Animals Savings**

1003 As stated several times in this section, the animal savings using the 3T3 and NHK NRU test
1004 methods to determine the staring dose for rodent acute oral toxicity test methods will depend
1005 on the proportion of test substances that fall into each of the GHS acute toxicity hazard
1006 categories. Although the prevalence of substances among the different categories will
1007 depend, to a large extent, on the mandate of a particular regulatory agency, Spielmann et al.
1008 (1999) indicated that 76% (845/1115) of the industrial substances submitted to the Federal
1009 Institute for Health Protection of Consumers and Veterinary Medicine in Berlin, Germany,
1010 since 1982 had LD_{50} >2000 mg/kg. The extent to which these substances represent the
1011 population of substances in commerce is not known. However, if the results of the validation
1012 study are broadly applicable to substances to be tested in the future, and if such substances
1013 are relatively nontoxic, the selection of starting doses using the *in vitro* NRU test methods
1014 may save a considerable number of animals since animal savings for the validation study
1015 were highest for the least toxic substances.

1016 **10.6 Summary**

1017 Computer simulation modeling of UDP testing using the default dose progression shows that,
1018 for the subset of reference substances evaluated, the prediction of starting doses using the
1019 3T3 and NHK NRU test methods with the RC rat-only millimole regression resulted in a
1020 statistically significant ($p < 0.05$) decrease in the number of animals used by an average of
1021 0.49 (6.2%) to 0.54 (5.8%) animals per test, depending upon the *in vitro NRU* test method
1022 and the dose-mortality slope (2.0 or 8.3) used. The mean animal savings improved slightly, to
1023 0.54 (6.8%) to 0.66 (7.0%) animals per test, when the RC rat-only weight regression was
1024 used.
1025
1026 When reference substances were grouped by GHS category, there were no mean animal
1027 savings by simulated UDP testing for substances with $50 < LD_{50} \leq 300$ mg/kg. The highest,
1028 and statistically significant, animal savings were observed with both *in vitro* NRU test
1029 methods when testing substances with $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg.

1030 When using the RC rat-only millimole regression, animal savings for these categories ranged
1031 from 1.28 (11.9%) to 1.58 (20.3%) animals per test. The use of the RC rat-only weight
1032 regression improved animal savings slightly for the substances in these toxicity categories to
1033 1.28 (14.0%) to 1.65 (16.7%) animals per test. Although the use of IC_{50} values to estimate
1034 starting doses for the simulated UDP deceased the number of animals used per test, it did not
1035 change the number of animals that would have died during the procedures.
1036
1037 Computer simulation modeling of ATC testing showed that, for the reference substances
1038 tested in this validation study, the prediction of starting doses using the 3T3 and NHK NRU
1039 test methods with the RC rat-only millimole regression resulted in a statistically significant
1040 ($p < 0.05$) decrease in the number of animals for ATC testing by an average of 0.51 (4.8%) to
1041 0.80 (7.3%) animals per test, depending upon the *in vitro NRU* test method and the dose-
1042 mortality slope (2.0 or 8.3) used. Animal savings improved to a mean of 0.91 (8.6%) to 1.09
1043 (10.2%) animals per test when the RC rat-only weight regression was used.
1044
1045 When test substances were grouped by GHS category, the mean animal savings for ATC
1046 testing using the RC rat-only millimole regression were statistically significant with the 3T3
1047 NRU test method at both dose-mortality slopes for substances with $5 < LD_{50} \leq 50$ mg/kg
1048 (1.15 [9.8%] to 1.17 [10.2%] animals per test), and for substances with $LD_{50} > 5000$ mg/kg
1049 (2.03 [17.1%] to 2.19 [18.3%] animals per test). Significantly more animals were needed
1050 when the 3T3-based starting doses were used, than the default starting dose for reference
1051 substances with $300 < LD_{50} \leq 2000$ mg/kg (i.e., the animal savings were negative: -0.92 [-
1052 9.5%] to -1.30 [-14.0%] animals). The mean animal savings with the NHK NRU test method
1053 and the RC rat-only millimole regression were statistically significant at both dose-mortality
1054 slopes for substances with $5 < LD_{50} \leq 50$ mg/kg (1.18 [10.2%] to 1.33 [-11.4%] animals per
1055 test), and for substances with $LD_{50} > 5000$ mg/kg (2.43 [20.5%] – 2.66 [22.2%] animals per
1056 test). When the RC rat-only weight regression was used, statistically significant savings in
1057 animal used were observed with both *in vitro* NRU test methods and dose-mortality slopes
1058 for substances with $5 < LD_{50} \leq 50$ mg/kg (1.25 [10.8%] to 1.51 [13.0%] animals per test), and
1059 substances with $LD_{50} > 5000$ mg/kg (2.94 [24.8%] to 3.25 [27.1%] animals per test). The use
1060 of IC_{50} values to estimate starting doses for the ATC refined animal use by producing

1061 approximately 0.5 to 0.6 fewer mean animal deaths per test than when the default starting
1062 dose of 300 mg/kg was used.
1063
1064 The use of the IC_{50}-based starting doses did not significantly alter the GHS category
1065 outcomes of the simulated UDP or ATC when compared with the outcomes based on the
1066 default starting dose. The concordance for GHS acute oral toxicity category for the IC_{50}-
1067 based starting dose with the default starting dose was 97 to 99% for both *in vitro* NRU
1068 methods and IC_{50}-LD_{50} regressions evaluated.
1069
1070 The magnitude of animal savings did not correlate with the accuracy of GHS categorization
1071 yielded by the NRU-predicted LD_{50} values (using the *in vitro* NRU IC_{50} values in the IC_{50}-
1072 LD_{50} regressions) or with the accuracy of GHS category outcomes since the accuracy and
1073 animals savings analyses used different standards for comparison.
1074
1075 The specific animal savings using the 3T3 and NHK NRU test methods with the RC rat-only
1076 regressions apply only to the reference substances evaluated in this validation study, and may
1077 not be broadly applicable to other substances. Spielmann et al. (1999) indicated that 76%
1078 (845/1115) of the industrial substances submitted to the Federal Institute for Health
1079 Protection of Consumers and Veterinary Medicine in Berlin, Germany, since 1982 had LD_{50}
1080 >2000 mg/kg. The extent to which these substances represent the population of substances in
1081 commerce is not known. However, if the results of the validation study are broadly
1082 applicable to substances to be tested in the future, and if such substances are relatively
1083 nontoxic, the selection of starting doses using the *in vitro* NRU test methods may save a
1084 considerable number of animals since animal savings for the validation study were highest
1085 for the least toxic substances.
1086

[This Page Intentionally Left Blank]

11.0 PRACTICAL CONSIDERATIONS 11-3

11.1 Transferability of the 3T3 and NHK NRU Test Methods 11-3
11.1.1 Facilities and Major Fixed Equipment 11-4
11.1.2 Availability of Other Necessary Equipment and Supplies 11-5
11.1.3 Problems Specific to the NHK Test Method 11-7

11.2 3T3 and NHK NRU Test Method Training Considerations 11-7
11.2.1 Required Training and Expertise 11-8
11.2.2 Training Requirements to Demonstrate Proficiency 11-9
11.2.3 Personnel Needed to Perform the *In Vitro* NRU Cytotoxicity Test Methods 11-10

11.3 Cost Considerations 11-11
11.3.1 3T3 and NHK NRU Test Methods 11-11
11.3.2 Rodent Acute Oral Toxicity Testing 11-12

11.4 Time Considerations for Performing the 3T3 and NHK NRU Tests 11-15
11.4.1 The 3T3 NRU Test Method 11-15
11.4.2 The NHK NRU Test Method 11-15
11.4.3 Prequalification of NHK Medium 11-15
11.4.4 *In Vivo* Testing 11-16
11.4.5 Using NRU Cytotoxicity Test Methods to Determine Need to Perform the Limit Test 11-16

11.5 Summary 11-17

[This Page Intentionally Left Blank]

43 **11.0 PRACTICAL CONSIDERATIONS**

44 The 3T3 and NHK NRU test methods are proposed as adjuncts, rather than replacements for,
45 *in vivo* acute oral toxicity assays. Data from these *in vitro* basal cytotoxicity test methods are
46 used with a linear regression model to predict the rat acute oral LD_{50} of the test substance,
47 which is then used to determine the starting dose for *subsequent rodent acute* oral toxicity
48 tests, as described in **Sections 10.2.2** and **10.3.2**. This section discusses practical issues
49 involved in using these two *in vitro* NRU test methods for predicting starting doses for rat
50 acute oral toxicity tests. Practical issues that need to be considered with respect to the
51 implementation of these cell culture methods include the need for, and availability of,
52 specialized equipment, personnel training and expertise requirements, cost considerations,
53 and time expenditures.

54 **11.1 Transferability of the 3T3 and NHK NRU Test Methods**

55 Transferability of a test method is defined as the ability of a method or procedure to be
56 accurately and reliably performed in different, competent laboratories (ICCVAM 2003).
57 Accuracy and reliability of these NRU test methods are discussed in **Sections 6** and **7**,
58 respectively.
59
60 Protocols for the 3T3 and NHK NRU test methods, including solubility testing, and
61 prequalification of keratinocyte growth medium, have been optimized and are available on
62 the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/methods/invitro.htm). The
63 protocols were designed with GLP-compliance in mind and can be easily implemented or
64 adapted by scientists with the appropriate technical experience.
65
66 Although the *in vitro* and *in vivo* test methods require some similar, general laboratory, skills
67 (e.g., preparation of solutions and test substance doses, record keeping), *in vitro* testing
68 requires skills specific to cell culture systems (e.g., aseptic techniques, microscopic
69 evaluation of cell cultures, propagation of cells in medium) but not to the maintenance,
70 handling, or treatment of rodents.
71
72

11.1.1 Facilities and Major Fixed Equipment

The following lists of facility requirements, equipment and supplies, and training and expertise are common to most *in vitro* mammalian cell culture laboratories. Required equipment and supplies are also described in detail in the validation study 3T3 and NHK protocols (**Appendices B** and **C**), the *Guidance Document* (ICCVAM 2001b, **Appendix D**), and Hartung et al. 2002.

11.1.1.1 *Facility Requirements*

The testing facility should be appropriate for operating a scientific laboratory (e.g., laboratory space, air handling procedures, access to utilities, shipping/receiving department [for appropriate receipt and handling of cell culture materials], etc.). Each facility should provide:

- Adequate facilities, equipment, and supplies
- Proper health and safety guidelines
- Satisfactory quality assurance procedures

Each facility should conform to all appropriate statutes (i.e., local, state, provincial, federal, national, international) concerning safety guidelines (e.g., general workplace safety guidelines, chemical handling and disposal guidelines, biohazard guidelines). Hartung et al. (2002) provides recommended safety guidelines for working with potentially infectious materials (e.g., HIV, hepatitis B, hepatitis C) and human materials (e.g., cells, tissues, fluids).

11.1.1.2 *Cell Culture Laboratory*

The testing facility should have a designated cell culture laboratory to ensure that *in vitro* cytotoxicity assays are performed under clean and proper aseptic conditions. The dedicated laboratory should be located such that through traffic is minimal to reduce possible disturbances that can lead to contamination which could compromise the cell culture assays. The room temperature of the laboratory should be regulated, monitored, and documented. Access to the laboratory and its supplies and test chemicals should be restricted to appropriate personnel.

11.1.1.3 *Major Equipment*

Each testing facility should have at a minimum the following equipment:

- Incubator (37ºC ± 1ºC, 90% ± 10% humidity, 5.0% ± 1% CO_2/air)
- Laminar flow clean bench/cabinet (standard: "biological hazard")
- Inverted phase contrast microscope
- 96-well plate spectrophotometric plate reader equipped with 540 nm ± 10 nm filter (if testing in 96-well plates)
- Autoclave
- Refrigerator
- Freezer (-70ºC)
- Cryogenic (liquid nitrogen) freezer/storage unit
- Computer

Equipment maintenance and calibration should be routinely performed and documented according to GLP guidelines and testing facility SOPs.

11.1.2 Availability of Other Necessary Equipment and Supplies

11.1.2.1 *General Equipment*

Each testing facility should have at a minimum the following equipment:

- Low speed centrifuge
- Adjustable temperature waterbath
- Pipettors
- Balance
- pH meter
- Cell counting system
- Water bath sonicator
- Magnetic stirrer
- Vortex mixer
- Antistatic bar ionizer (for reduction of static on tissue culture plates)

Equipment maintenance and calibration should be routinely performed and documented as per GLP guidelines and testing facility SOPs. The types of equipment listed in this section are available from scientific and laboratory supply companies (e.g., Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/).

133 11.1.2.2 *Cell Culture Materials and Supplies*

134 The following supplies are needed for the *in vitro* NRU test methods. Specific product and
135 private company names are provided either as an identification of actual materials/brands
136 used in the validation study or as examples. Mention of these names does not imply
137 endorsement of the product or company.

138 • Tissue culture plasticware (flasks [e.g., 25 cm^2, 75-80 cm^2], 96-well plates,
139 disposable pipettes)
140 • Laboratory glassware (e.g., flasks, bottles, graduated cylinders)
141 • Adhesive film plate sealers (e.g., Excel Scientific SealPlate™)
142 • Sterile filtration systems (e.g., vacuum filtration units with 0.22 µm and 0.45
143 µm sterile filters)
144 • Culture medium and supplements (e.g., Dulbecco's Modification of Eagle's
145 Medium [DMEM]; prequalified NHK medium)
146 • Newborn Calf serum (bovine)
147 • Balanced salt solutions (e.g., Hanks' Balanced Salt Solution [HBSS] without
148 Ca^{2+} and Mg^2, Dulbecco's Phosphate Buffered Saline [D-PBS])
149

150 Cell culture supplies are generally available through the major scientific and laboratory
151 supply companies and through specialty companies (e.g., GIBCO, SIGMA-Aldrich,
152 CAMBREX/Biowhittaker, Becton Dickinson). Compositions of culture media,
153 supplements/additives, salt solutions, NRU assay chemicals, and the volumes of each needed
154 for each test method, should be defined. All tissue culture flasks and dishes needed to assure
155 proper cell propagation should be identified.

156 11.1.2.3 *Cell Cultures*

157 3T3 Mouse Fibroblasts: BALB/c 3T3 cells, clone 31, can be obtained from
158 national/international cell culture repositories (e.g., American Type Culture Collection
159 [ATCC], Manassas, VA, product # CCL-163).
160

161 NHKs: These non-transformed keratinocyte cells from cryopreserved primary or secondary
162 cells can be obtained from national/international cell culture repositories (e.g., CAMBREX
163 Bio Science, 8830 Biggs Ford Road, Walkersville, MD), or isolated from donated tissue

using proper collection, preparation, and propagation techniques. It may be difficult, at times, to obtain adequate supplies of keratinocytes; the preparation of a pool of cells depends on the availability of tissue donors. It is recommended that testing laboratories procure of a commercially available stock pool of cells and store them indefinitely in a cryogenics freezer.

All cell stock and cultures used for testing must be certified as free of contamination by mycoplasma and bacteria.

11.1.3 Problems Specific to the NHK NRU Test Method

FAL had difficulty obtaining an adequate supply of NHK medium during the validation study. Communication between the UK distributor and the laboratory was uneven and the SMT attempted to resolve the supply issue on several occasions. The other laboratories periodically had difficulties in obtaining NHK medium and supplements that adequately supported keratinocyte growth. Although the purchased medium and supplements met the manufacturer's QA/QC standards, certain lots of the medium and supplements did not support the growth of NHK cells to the extent needed in the test protocol. To deal with these problems, an NHK medium prequalification protocol was incorporated into the study to avoid unnecessarily repeating studies because of medium and supplements that did not adequately support cell growth. These experiences illustrate the need for multiple sources of keratinocyte cell culture medium. They also suggest that the NHK results could be more variable than the 3T3 results because of the batch-to-batch differences in NHK growth medium and supplements.

11.2 3T3 and NHK NRU Test Method Training Considerations

The ECVAM Good Cell Culture Practice Task Force Report 1 (Hartung et al. 2002) encouraged the establishment of practices and principles that will reduce uncertainty in the development and application of *in vitro* test methods. Training in good cell culture practices, in conjunction with good laboratory practices, are essential for all *in vitro* cytotoxicity testing and should be employed to ensure that data produced from the 3T3 and NHK NRU test methods are reproducible, credible, and acceptable.

193 *In vitro* cytotoxicity test methods require personnel trained specifically in sterile tissue/cell
194 culture techniques and general laboratory procedures. Personnel should have mandatory
195 training in good cell culture practices, in the specialized culture procedures needed for these
196 assays, and in safety and handling practices appropriate to the types of substances that may
197 be tested in the laboratory (Hartung et al. 2002).
198
199 The facility management should establish scientific guidelines and procedures, train and
200 supervise professional and technical staff, and evaluate results and performance within their
201 discipline area relative to the testing requirements. Performance of the tests requires a
202 moderate degree of technical capability and a high degree of skill in monitoring and
203 maintaining appropriate cell growth conditions, troubleshooting the potential and real
204 problems in culture systems, and analyzing and interpreting *in vitro* cytotoxicity data. Each
205 individual engaged in the conduct of a study, or responsible for its supervision, shall have
206 education, training, and experience, or combination thereof, to enable that individual to
207 perform the assigned duties. The NRU test methods do not require that personnel be trained
208 to perform *in vivo* testing.

209 11.2.1 Required Training and Expertise

210 Personnel performing *in vitro* testing should have training in basic cell culture aspects such
211 as: sterile technique, handling culture media, feeding cultures, cell counting, subculture
212 (trypsinization), detection and elimination of contamination, cell growth and measurement of
213 growth curves, viability assays, and storage and freezing/thawing of cells. Additionally,
214 training is encouraged for special culture procedures such as primary cell and tissue cultures,
215 toxicity testing, and viability assays. Laboratory personnel should be trained in the
216 application of GLP requirements (see **Section 8.1.1**), and in the safe storage, handling, and
217 disposal of toxic substances.

218 11.2.1.1 *Specific Training and Expertise Needed*

219 Personnel performing the *in vitro* cytotoxicity test methods should be well experienced in
220 general cell culture techniques and should be able to:

221 - Work with cryogenic freezing apparatus
222 - Pipette solutions with large volume pipettors and multi-channel pipettors

- Establish cells in culture vessels under aseptic conditions and monitor growth; recognize normal and abnormal cell growth characteristics; and document observations of cell cultures throughout all aspects of the procedure
- Perform the *in vitro* assays by following the protocols to grow the cells, count, transfer, and feed the cells, treat the cells with test substances, perform application of adhesive plate sealers to culture plates for control of volatile substances, perform the NRU assay, perform optical density measurements, transfer data to electronic templates
- Operate equipment necessary for maintaining cell culture laboratories (e.g., incubators, biohazard hoods, spectrophotometric microtiter plate readers)

11.2.1.2 *General Laboratory Expertise Needed*

Personnel should also be able to understand and perform basic laboratory techniques and laboratory management:

- Prepare cell culture solutions (e.g., culture medium, NRU solutions), measure pH, know proper storage conditions, and maintain proper documentation
- Prepare test substances for application to cell cultures, follow solubility protocols to adequately prepare test chemicals in solution, recognize solubility issues (e.g., insolubility nature of chemical, precipitation), and implement procedures for dissolving the test chemicals
- Monitor and control laboratory environment (e.g., temperature, humidity, lighting, traffic), maintain equipment to support cell cultures (e.g., temperature, humidity, gas flow, calibrations)

11.2.2 Training Requirements to Demonstrate Proficiency

Laboratories establish their own criteria for proficiency but, over the course of training, laboratory personnel should be able to understand the protocol, perform the protocol with guidance from an experienced supervisor/trainer and, eventually, perform the protocol with minimal or no supervision. An experienced supervisor determines when a technician is adequately trained because there are no standardized criteria or tasks that can be used to accurately measure competence. After the technician demonstrates competence in executing all the aspects of the test protocols(s), it is appropriate to perform routine assessments of technical competence using a benchmark, coded control test substance (e.g., SLS). It is

254 essential that the laboratory staff be certified as proficient in using the test methods to test
255 unknowns.
256
257 The laboratories in the validation study were selected because of their experience in
258 performing *in vitro* cytotoxicity assays but were required to develop additional skills through
259 Phases I and II (e.g., data collection and transfer to Excel® and PRISM® templates).
260 Inexperienced laboratory personnel were trained by having them perform "training" assays
261 using SLS. In the early phases of the validation study, the laboratories continued training by
262 testing coded reference substances of various toxicities, and performing solubility testing on
263 substances of varying solubilities. These procedures helped improve proficiency among the
264 laboratory for the final phase of the validation study.

265 11.2.2.1 *Proficiency With GLP-Compliance*

266 Results from these test methods will be submitted to regulatory agencies that will, for the
267 most part, require GLPs. Laboratories should work toward attaining GLP compliance. GLP
268 compliance in each laboratory is determined by its independent QA unit. ECBC and IIVS
269 conducted this validation study in compliance with GLP (see **Section 8.1.1**). Their respective
270 QA units (as per GLPs) reviewed the various aspects of the study and issued QA statements
271 that addressed whether the test methods and the results described in the Final Report
272 accurately followed the test protocol and reflected the raw data produced during the study,
273 and provided assurance that all testing was done under according to GLP. FAL (which was
274 non-GLP-adherent) followed the GLP standards referenced in **Section 8.1.1** as guidelines for
275 conducting this study. FAL had no QA unit to judge GLP compliance.

276 11.2.3 <u>Personnel Needed to Perform the *In Vitro* NRU Test Methods</u>

277 • Study Director: the individual with the overall responsibility for the technical
278 conduct of the testing (e.g., is familiar with the test procedures, provides SOPs
279 and ensures GLP compliance, analyzes and interprets the data, determines test
280 acceptance, oversees recordkeeping procedures, and produces the test reports.
281 • Quality Assurance Officer: monitors the testing to assure conformance with
282 GLP requirements, must be independent of the Study Director.

- Laboratory Technician(s): individuals trained in sterile tissue/cell culture techniques and general laboratory procedures and who are capable of performing the test methods according to GLPs.

11.3 Cost Considerations

11.3.1 3T3 and NHK NRU Test Methods

11.3.1.1 *Equipment Costs*

Major instruments and equipment needed to implement the *in vitro* cytotoxicity test methods are described in **Section 11.1.1**. Ranges of costs for some of the equipment were obtained from on-line catalogues for two major scientific equipment and supplies companies (Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/). These prices are for equipment that will meet the minimum needs of the NRU test methods (see **Table 11-1**). These costs were researched in August 2006.

Table 11-1 Costs for Major Laboratory Equipment

Equipment	Range of Costs[1]
Class II Biological Safety Cabinet	$7,300 – $12,200
CO_2 Incubator	$5,100 – $16,400
Spectrophotometer Microplate Reader	$5,000 – $7,500
Freezer (capable of -70°C)	$8,000 – $15,300
Refrigerator	$1,300 – $9,800
Centrifuge (benchtop model)	$2,100 – $8,500
Microscope (inverse phase contrast)	$3,000 – $14,500
Coulter Counter[2, 3]	$3,000 – $9,000
Autoclave (benchtop model)[2]	$3,500 – $15,400
Cryogenic (liquid nitrogen) Storage Container	$1,000 – $3,700

[1]From on-line scientific equipment catalogues (Thomas Scientific - http://www.thomassci.com/index.jsp; Fisher Scientific - https://www.fishersci.com/). [searched August 2006]
[2]May be useful, but not required for performing the tests.
[3]Other automatic cell counters may be used.

304 11.3.1.2 *Costs for Cell Cultures and Supplies*

305 Supplies such as cell culture chemicals, the reagents used to measure NRU, and cell culture
306 plasticware are available from numerous suppliers, and are not cost prohibitive.
307
308 The 3T3 NRU test method is generally less expensive to perform than the NHK NRU test
309 method. One vial of the immortalized 3T3 cells (~$200 [ATCC]) can be propagated
310 indefinitely by passaging cells and periodically cryopreserving batches of cells. The NHK
311 NRU test method requires a fresh sample of primary cells for each test run (~$380 per vial
312 [CAMBREX]). Because primary NHK cells are passaged only once after initiating the
313 culture, there are no cells available to cryopreserve a stock batch of cells. The D-MEM
314 medium used for the 3T3 cells is less expensive, more "generic", and more readily available
315 than keratinocyte-specific NHK medium. (See **Table 11-2**).

316 11.3.1.3 *Commercial Testing*

317 The following price quotes are provided as examples of test costs and were acquired from
318 commercial laboratories through Internet contact or through personal communication. Use of
319 information from these specific laboratories does not imply endorsement of them.
320
321 A representative of MB Research Laboratories (Spinnerstown, PA,
322 http://www.mbresearch.com/) provided a quote (personal communication, 2005) for an in
323 vitro 24-hr cytotoxicity test (but not a 48-hour test period) of $1050 (USP standards[1]) or
324 $1950 (ISO standards[1]) for a set of three test chemicals. The lead laboratory for the
325 NICEATM/ECVAM study, IIVS (Gaithersburg, MD, http://www.iivs.org/) provides
326 commercial laboratory GLP-compliant testing using this study's protocols (48-hour test
327 period) at a cost of $1120 - $1850 per chemical/sample for one cell type (personal
328 communication 2005) (see **Table 11-2)**.

329 11.3.2 Rodent Acute Oral Toxicity Testing

330 As stated in **Section 11.3.1.3**, presentation of price quotes from commercial laboratories is to
331 provide examples of test costs and does not imply an endorsement of that laboratory. **Table**

[1] USP=United States Pharmacopeia; ISO=International Standards Organization. These organizations provide international standard testing requirements for products that require high quality for public use.

332 **11-3** provides some commercial prices for acute oral systemic toxicity testing. MB Research
333 Laboratories performs the UDP test at a cost of $750 for three rats and charges $250 for each
334 additional rat needed. In the best-case scenario, the UDP test needs only three rats ($750). In
335 the worst-case scenario, this test would need an additional 12 rats (15 maximum for the test);
336 the total cost of the test would be $3,750. In this costing strategy, $250 is saved for each rat
337 not used by an accurate prediction of the starting dose by the 3T3 or NHK NRU test method.
338 Because the *in vitro* cytotoxicity test costs from $350 to $1850 per chemical, there is no net
339 savings in animal costs if fewer than two to six animals are saved.
340

341 **Table 11-2 Costs for Cell Culture Materials and Commercial Laboratory *In Vitro***
342 **Cytotoxicity Testing**
343

Item	Cost (approximate)	Number of Tests Possible	Other
3T3 Cells	~$200/vial[1]	indefinite	One vial can produce an indefinite supply of cells by propagating the cells in culture and periodically freezing a pool of cells.
NHK Cells	~$380/vial[2]	~5 (96-well plates)	Since cells are passaged only once beyond cryopreservation, new vials should be thawed as needed to maintain continuous testing.
Dulbeccos' Minimum Essential Medium (D-MEM) with supplements	~$20/500mL[3]	~15 (96-well plates)	Establish cells in culture (~20 mL/vial of cells; 60 mL/3 vials), seed cells in 96-well plates (12 mL/plate; 180 mL/15 plates); prepare stock solution and eight concentration dilutions (~20 mL/chemical; 300 mL/15 plates).
NHK Medium with supplements	~$80/500 mL[2]	~15 (96-well plates)	Same as DMEM (above)
Commercial Laboratory Testing (MB Research Laboratories [GLP-compliant])	$1050/$1950 (USP/ISO) per 3 test materials[4]	1 test/material	*in vitro* NRU cytotoxicity test (24-hour test period)
Commercial Laboratory Testing (Institute for In Vitro Sciences [GLP-compliant])	$1120 (GLP) per test material (minimum of 5 materials tested simultaneously)[4]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)
Commercial Laboratory Testing (Institute for In Vitro Sciences))	$1850 (GLP) per single test material (tested individually)[4]	1 range finder, 2 definitive tests per test material	*in vitro* NRU cytotoxicity test (48-hour test period)

344 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; USP/ISO=
345 United States Pharmacopeia/International Standards Organization GLP=Good Laboratory Practices
346 [1]Catalogue price from American Type Culture Collection (ATCC) (http://www.atcc.org/)
347 [2]Catalogue price from CAMBREX (http://www.cambrex.com/Welcome.asp)
348 [3]Catalogue price from INVITROGEN (http://www.invitrogen.com/content.cfm?pageid=1)
349 [4]Personal communication (Raabe 2005)
350

The President of Product Safety Laboratories, Gary Wnorowski, (Dayton, NJ, http://www.productsafetylabs.com/), provided a cost quote of $2700 for determination of a rat LD_{50} value using the UDP test; the cost is independent of the number of rats that are needed. Each test dose is administered ~24-48 hours after the previous dose and each animal test generally does not exceed four days. The time involved in providing the LD_{50} value is approximately three months (initiation of the test to provision of the final report). Having the estimated LD_{50} value would not affect the cost of the *in vivo* test but could reduce the number of animals needed.

Bio Research Laboratories performs the rat acute oral toxicity test using a test method that determines lethality and signs of acute toxicity from a waste sample administered in a single dose, by gavage, to a limited number of rats. The bioassay determines if the test sample produces an LD_{50} either greater than or less than a regulatory threshold corresponding to a hazardous waste designation (i.e., 5000, 500, 50 mg/kg). A minimum of 10 rats is used at the tested dose for the regulatory threshold value that is relevant to the test sponsor. In this testing scenario, knowledge of the estimated LD_{50} would not reduce animal use or test costs if a single predetermined dose is tested.

Table 11-3 Commercial Prices for Conducting *In Vivo* Acute Rat Toxicity Testing

Test	GLP-Compliant	Non GLP-Compliant	Company
Acute Oral Toxicity UDP: Limit Test - 2000 mg/kg	$1200	$1000	Product Safety Laboratories
Acute Oral Toxicity UDP: Limit Test - 5000 mg/kg	$800	$650	Product Safety Laboratories
Acute Oral Toxicity UDP: LD_{50}	$2700	$2200	Product Safety Laboratories[1]
Acute Oral Rat Toxicity: single dose[2]	$950	NA	Bio Research Laboratories
Acute Oral Rat Toxicity: two doses[2]	$1500	NA	Bio Research Laboratories
Acute Oral Rat Toxicity: LD_{50}	$3000	NA	Bio Research Laboratories
Acute Oral Toxicity – UDP	$730 for the first 3 animals; $250 each additional animal	NA	MB Research Laboratories[1]

Abbreviations: UDP=Up-and-Down Procedure; GLP=Good Laboratory Practices; NA=Not available.
[1]Personal communication (Wnorowski 2005).

[2]Washington State Biological Testing Methods #80-12 For the Designation of Dangerous Waste; Part B: Acute Oral Rat Toxicity Test [http://www.ecy.wa.gov/pubs/80012.pdf]. This test method is an adaptation of the EPA Health Affects Test Guidelines OPPTS 870.110 Acute Oral Toxicity and American Society for Testing and Materials (ASTM) methods E 1163-90 (Standard test method for estimating acute oral toxicity in rats) and E 1372-90 (Standard test method for conducting a 90-day oral toxicity study in rats).

11.4 Time Considerations for Performing the 3T3 and NHK NRU Tests

11.4.1 The 3T3 NRU Test Method

Approximately one week is needed to thaw cryopreserved 3T3 cells, propagate them, and passage them at least two times before subculturing them into the 96-well test plates. After subculture into 96-well plates, the cells are incubated another 24 hours to reach the proper confluence, and then exposed to test chemical for 48 hours. The initial 3T3 NRU test (range finder or definitive test) takes approximately 10 days. However, after the cells are established in culture, they can be passaged for approximately two months before having to go back to the cryopreserved cells to start a new culture. A 3T3 NRU test can be completed in less than four consecutive days when started from an established stock culture. Multiple substances can be tested at the same time, and different tests can overlap each other; thus, many substances can be tested in a relatively short time.

11.4.2 The NHK NRU Test Method

Approximately one week is needed to thaw cryopreserved NHK cells, propagate them, and passage them into the 96-well test plates. After subculture into 96-well plates, the cells are incubated another 48-72 hours to reach the proper confluence and then exposed to test chemical for 48 hours. The entire NHK NRU test (range finder or definitive test) requires approximately 11-12 days. Cells can be seeded at different densities from one starter vial in the culture flasks so that passaging the cultures can take place on different days. Once the cells are established in culture, they are passaged once to the 96-well test plates and an NHK NRU test can usually be completed in five to six consecutive days. Multiple substances can be tested at the same time, and different tests can overlap each other; thus, many substances can be tested in a relatively short time.

11.4.3 Prequalification of NHK Medium

The protocol for the prequalification of NHK medium requires nearly identical steps, and similar time-line (i.e., 11-12 days), as required for the NHK rangefinder and definitive tests.

406 Table 11-2 provides an estimate of how many tests could be performed using one 500 mL
407 bottle of medium with supplements (~15 tests in 96-well plates).

408 11.4.4 *In Vivo* Testing

409 According to guidelines for acute oral toxicity testing, single animals or groups of animals
410 are dosed in sequence, usually at 2-4 day intervals, and observations are generally made for
411 up to 14 days (for animals that are not moribund) for the main test and limit dose test (EPA
412 2002a; OECD 2001a; OECD 2001b, OECD 2001c). The addition of 3T3 or NHK NRU
413 testing to estimate a starting dose prior to the implementation of the UDP main test or limit
414 dose test would take 10-12 days, but could save up to 14 days of observation for every
415 animal not used.

416 11.4.5 Use of NRU Cytotoxicity Test Methods to Determine Need to Perform the Limit
417 Test

418 The *in vitro* NRU test methods can provide a savings of time when used to determine if an *in*
419 *vivo* acute oral toxicity limit test can be employed as the initial test for a substance with
420 unknown *in vivo* toxicity. If the IC_{50} value from an *in vitro* NRU test could accurately predict
421 an LD_{50} that is greater than, or equal to, the limit test starting dose (i.e., 5000 mg/kg for the
422 UDP, 2000 mg/kg for the ATC), the *in vivo* test could start at the limit test dose. This
423 approach has the potential to eliminate the need to do the main test and could result in a net
424 savings of six days for the UDP test method and about one day for the ATC test method.
425 **Table 11-4** illustrates the following:

426 - Time needed to perform the 3T3 and NHK NRU test
427 - Time needed to reach the limit test starting dose when initiating the *in vivo*
428 main test using the default starting doses (UDP and ATC)

429

430 The times presented in **Table 11-4** are best-case scenarios:

431 - 3T3 cells reach ≤50% confluence in approximately 24 hours
432 - NHK cells reach >20% confluence in approximately 48 hours
433 - Animals show no evident toxicity 48 hours post-dosing, and additional
434 animals are dosed at the next higher default dose

435

436

436 **Table 11-4 Comparison of Time Needed for *In Vitro* and *In Vivo* Testing**
437

Time	3T3 NRU Test Method	NHK NRU Test Method	UDP (5000 mg/kg upper limit)	ATC (2000 mg/kg upper limit)
Day 1	Seed cells in 96-well plate Incubate for 24 ± 2 hr	Seed cells in 96-well plate Incubate for approximately 48 to 72 hr	Dose 1 animal at default dose (175 mg/kg) Observe for 48 hr	Dose 3 animals at default dose (300 mg/kg) Observe for 48 hr
Day 2	Apply test substance Incubate for 48 ± 0.5 hr	Incubate	Observe	Observe
Day 3	Incubate	Apply test substance Incubate for 48 ± 0.5 hr	No death Dose 1 animal at next default dose (550 mg/kg) Observe 48 hr	0 – 1 animal dies Dose 3 animals at default dose (300 mg/kg) Observe 48 hr
Day 4	NRU: 3 ± 0.1 hr Elute NR: 0.33 to 0.75 hr OD_{540} measurement Calculate IC_{50} Estimate LD_{50} and Starting Dose*	Incubate	Observe	Observe
Day 5		NRU: 3 ± 0.1 hr Elute NR: 0.33 to 0.75 hr OD_{540} measurement Calculate IC_{50} Estimate LD_{50} and Starting Dose*	No death Dose 1 animal at next default dose (1750 mg/kg) Observe 48 hr	0 – 1 animal dies Dose 3 animals at next default dose (2000 mg/kg) **Starting Point for the Limit Test**
Day 6			Observe	
Day 7			No death Dose 1 animal at next default dose (5000 mg/kg) **Starting Point for the Limit Test**	

438 Abbreviations: 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; UDP=Up-and-Down
439 Procedure; ATC=Acute Toxic Class method; hr=Hour; NR=Neutral red; OD_{540}=Optical density at 540 nm.
440
441
442

443 ## 11.5 Summary

444 • All equipment and supplies should be readily commercially available. During
445 the validation study, direct communication with the NHK medium supplier
446 insured that specific lots of medium were available to the laboratories. The

447 test methods are expected to be transferable to laboratories experienced with
448 mammalian cell culture methods.

449 • Much of the training and expertise needed to perform the 3T3 and NHK NRU
450 test methods are common to mammalian cell culture procedures. Additional
451 technical training would not be extensive because these test methods are
452 similar to other *in vitro* mammalian cell culture assays, and no extraordinary
453 techniques are necessary. GLP training should be provided to technicians to
454 ensure proper adherence to protocols and documentation procedures.

455 • Prices for commercial testing for one chemical are $1,120 to $1,850 (**Table**
456 **11-2**) for *in vitro* cytotoxicity testing in the 3T3 and NHK test methods,
457 respectively, to determine the IC_{50} (Raabe 2005, personal communication). In
458 contrast, the *in vivo* rat acute oral testing for LD_{50} determination could cost
459 from $750 - $3,750 (**Table 11-3**), depending on the test method used and the
460 toxicity of the test substance. Comparison of costs of *in vitro* testing to *in vivo*
461 testing is difficult because the *in vitro* NRU test methods are not replacements
462 for the animal testing, and animal testing would be performed regardless of
463 the responses of the 3T3 or NHK cells. The use of these *in vitro* NRU test
464 methods may not reduce the overall cost of the *in vivo* rat acute oral toxicity
465 test, but has the potential to reduce the number of animals needed for a study.

12.0 REFERENCES

Adegunloye BI, Sofola OA. 1997. Differential responses of rat aorta and mesenteric artery to norepinephrine and serotonin *in vitro*. Pharmacology 55:25-31.

Ames A III. 2000. CNS energy metabolism as related to function. Brain Res Rev 34:42-68.

Anon. 1993. Annexes I, II, III, and IV to Commission Directive 93/21/EEC of 27 April 1993 adapting to technical progress for the 18th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L110A, 1-86. (EU toxicity categories 25, 25-200, 200-2000, > 2000).

ASTM. 1999. American Society of Testing and Materials (ASTM) Standard E691-99, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. American National Standards Institute. Washington, D.C. Available: http://www.ansi.org/ [accessed 22 December 2005].

ATSDR. 1999a. Toxicological Profile for Cadmium. Update PB/99/166621. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 1999b. Toxicological Profile for Diethyl phthalate. PB/95/264214/AS. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 2000a. Toxicological Profile for Endosulfan. Update. PB/2000/108023. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 2000b. Toxicological Profile for Chromium. Update. PB/2000/108022. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 2001. Toxicological Profile for Di-n-butyl phthalate. Update. PB/2001/109104/AS. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 2004a. Toxicological Profile for Pyrethrins and Pyrethroids. Update. PB2004-100004. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

ATSDR. 2004b. Toxicological Profile for Selenium. Update. PB2004-100005. US Dept. of Health and Human Services, Atlanta GA. Available: http://www.atsdr.cdc.gov/toxpro2.html. [accessed 2 June 2005].

Babich H, Martin-Alguacil N, Raul C, Rosenberg DW, Borenfreund E. 1991. Response of human cell cultures to cytotoxicants requiring metabolic activation. In: Alternative Methods in Toxicology (Goldberg AM, ed.), Vol. 8. New York:Mary Ann Liebert, Inc., 263-276.

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test for classification and labelling of chemicals. Toxicol In Vitro 9**:**871-929.

Bernson V, Bondesson I, Ekwall B, Stenberg K, Walum E. 1987. A Multicentre evaluation study of in vitro cytotoxicity. Altern Lab Anim 14:144-146.

Bliss CI. 1938. The determination of the dosage-mortality curve from small numbers. Quarterly Journal of Pharmacy and Pharmacology 11:192-214.

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

Borenfreund E, Puerner J. 1984. A simple quantitative procedure using monolayer cultures for cytotoxicity assays (HTD/ NR-90). J Tissue Culture Meth 9:7-9.

Borenfreund E, Puerner JA. 1985. Toxicity determination *in vitro* by morphological alterations and neutral red absorption. Toxicol Lett. 24:119-124.

Botham PA. 2004. The validation of in vitro methods for skin irritation. Toxicol Lett 149:387-390.

Brantom PG, Bruner LH, Chamberlain M, De Silva O, Dupuis J, Earl LK, et al. 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. Toxicol In Vitro 11:141-179.

Calabrese EJ. 2005. Cancer biology and hormesis: human tumor cell lines commonly display hormetic (biphasic) dose responses. Crit Rev Toxicol 35:463-582.

Canadian Centre for Occupational Health and Safety. 2005. IPCS INTOX Database. Available: http://www.intox.org/databank/index.htm. [accessed 29 November 2005].

Casarett LJ, Klaassen CD, Doull J, eds. 2001. Casarett and Doull's Toxicology, The Basic Science of Poisons. 6th Edition. New York:McGraw-Hill.

Charles River Laboratories Catalog. 2002. Available: http://www.criver.com. [accessed 2 June 2005].

Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al. 1996a. MEIC evaluation of acute systemic toxicity. Part I. Methodology of 68 *in vitro*

toxicity assays used to test the first 30 reference chemicals. Altern Lab Anim 24 (suppl. 1):51-272.

Clemedson C, McFarlane-Abdulla E, Andersson M, Barile FA, Calleja MC, Chesné C, et al. 1996b. MEIC evaluation of acute systemic toxicity. Part II. *In vitro* results from 68 toxicity assays used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. Altern Lab Anim 24 (suppl. 1):273-311.

Clemedson C, Barile FA, Ekwall Ba, Gómez-Lechón MJ, Hall T, Imai K, et al. 1998a. MEIC evaluation of acute systemic toxicity. Part III. *In vitro* results from 16 additional methods used to test the first 30 reference chemicals and a comparative cytotoxicity analysis. Altern Lab Anim 26 (suppl. 1):93-129.

Clemedson C, Andersson M, Aoki Y, Barile FA, Bassi AM, Calleja MC, et al. 1998b. MEIC evaluation of acute systemic toxicity. Part IV. *In vitro* results from 67 toxicity assays used to test reference chemicals 31-50 and a comparative cytotoxicity analysis. Altern Lab Anim 26 (suppl. 1):131-183.

Clemedson C, Nordin-Andersson M, Bjerregaard H.F, Clausen J, Forsby A, Gustafsson H, Hansson U, Isomaa B, Jorgensen C, et al. 2002. Development of an *in vitro* test battery for the estimation of acute human systemic toxicity: an outline of the EDIT project Altern Lab Anim 30:313-321.

Clemedson, C. 2005. A-Cute-Tox project an Integrated Project under the EU 6FP with the aim to optimize and pre-validate an *in vitro* test strategy for predicting human acute toxicity. European Society of Toxicology In Vitro Newsletter. November. No. 18. p6.

Clothier R, Willshaw A, Cox H, Garle M, Bowler H, Combes R. 1999. The use of human keratinocytes in the EU/COLIPA international in vitro phototoxicity test validation study and the ECVAM/COLIPA study on UV filter chemicals. Altern Lab Anim 27:247-259.

Coeck S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, et al. 2005. Guidance on Good Cell Culture Proactice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice. Altern Lab Anim 33:261-287.

Coldwell BB, Boyd EM. 1966. The acute rectal toxicity of acetylsalicylic acid. Can J Physiol Pharmacol 44: 909-918

Conolly RB, Lutz WK. 2004. Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol Sci 77:151-157.

Cooper-Hannan R, Harbell J, Coecke S, et al. 1999. The principles of good laboratory practice: application to *in vitro* toxicology studies. Altern Lab Anim 27:539-577.

Cosmetic Ingredient Review Panel. 1983. Final report on the safety assessment of sodium borate and boric acid. J Am Coll Toxicol 2:87-125.

138
139 CPSC. 2003. Commercial Practices. Hazardous Substances and Articles; Administration And
140 Enforcement Regulations. 16 CFR 1500.42.
141
142 Creppy EE, Chiarappa P, Baudrimont I, Borracci P, Moukha S, Carratu MR. 2004.
143 Synergistic effects of fumonisin B1 and ochratoxin A: are in vitro cytotoxicity data
144 predictive of in vivo acute toxicity? Toxicology 201:115-23.
145
146 Curren R, Bruner L, Goldberg A, Walum E. 1998. 13th meeting of the Scientific Group on
147 Methodologies for the Safety Evaluation of Chemicals (SGOMSEC): Validation and acute
148 toxicity testing. Environ Health Perspect 106 (suppl. 2): 419-425.
149
150 Curren RD, Moyer GO, Wilt N, Clear ML, Sizemore AM, Mun G. 2003. Assessment of
151 protocol variables in cytotoxicity assays utilizing BALB/c 3T3 cells and normal human
152 keratinocytes [Abstract]. Toxicologist 72:157. Available:
153 http://www.iivs.org/pages/publication_view.php?doc_id=43. [accessed 23 September 2005].
154
155 Dal Negro G, Vandin L, Bonato M, Sciuscio D. 2006. Toward refinement of the colony-
156 forming unit-granulocyte/macrophage clonogenic assay: inclusion of a metabolic system.
157 Toxicol In Vitro 20:743-9.
158
159 DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Class 6,
160 Division 6.1 – Definitions. 49 CFR 173.132. Available:
161 http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
162
163 DOT. 2003. Shippers--General Requirements for Shipments and Packagings. Assignment of
164 Packing Group and Hazard Zones for Division 6.1 Materials. 49 CFR 173.133. Available:
165 http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
166
167 DOT. 1999. Code of Federal Regulations Title 49. SubChapter C. Assignment of packing
168 group and hazard zones for Division 6.1 materials.
169
170 Deichmann WB. 1969. Toxicology of Drugs and Chemicals. New York. Academic Press,
171 Inc. pg.67.
172
173 Dixon WJ, Massey FJ. 1981. Introduction to Statistical Analysis, 4th ed. Milwaukee:Quality
174 Press.
175
176 Ekwall B. 1983. Screening of toxic compounds in mammalian cell cultures. Ann New York
177 Acad Sci 407:64-77.
178
179 Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998a.
180 MEIC evaluation of acute systemic toxicity. Part V. Rodent and human toxicity data for the
181 50 reference chemicals. Altern Lab Anim 26 (suppl. 2):571-616.
182

Ekwall B, Clemedson C, Craaford B, Ekwall B, Hallander S, Walum E, Bondesson I. 1998b. MEIC evaluation of acute systemic toxicity. Part VI. The prediction of human toxicity by rodent LD50 values and results from 61 *in vitro* methods. Altern Lab Anim 26 (suppl. 2):617-658.

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

Ekwall B, Ekwall B, Sjostrom M. 2000. MEIC evaluation of acute systemic toxicity: Part VIII. Multivariate partial least squares evaluation, including the selection of a battery cell line tests with a good prediction of human acute lethal peak blood concentrations for 50 chemicals. Altern Lab Anim 28 (suppl. 1):201-234.

EPA. 1992. Reregistration Eligibility Document Citric Acid. List D, Case 4204. Office of Pesticide Programs. Washington, DC:U.S. Environmental Protection Agency. Available: http://cfpub.epa.gov/pesticides/reregistration. [accessed 2 June 2005].

EPA. 1995. Reregistration Eligibility Decision (RED) Diquat Dibromide. EPA 738-R-95-016. Office of Prevention, Pesticides and Toxic Substances. Washington, DC:U.S. Environmental Protection Agency. Available: http://cfpub.epa.gov/pesticides/reregistration.[accessed 2 June 2005].

EPA. 1996. Trichloroacetic acid (CASRN 76-03-9) file in Integrated Risk Information System. Office of Research and Development. Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/iris.[accessed 23 September 2005].

EPA 738-R-95-016. Office of Prevention, Pesticides and Toxic Substances. Washington, DC:U.S. Environmental Protection Agency. Available: http://cfpub.epa.gov/pesticides/reregistration.[accessed 2 June 2005].

EPA. 1998. Product Properties Test Guidelines OPPTS 830.7840 Water Solubility: Column Elution Method; Shake Flask Method. EPA 712–C–98–041. Washington, DC: U.S. Environmental Protection Agency.

EPA. 2000. High Production Challenge Program. Office of Pollution Prevention and Toxics (OPPT). Washington, DC:U.S. Environmental Protection Agency. Available: http://www.epa.gov/chemrtk/volchall.htm [accessed 21 December 2005].

EPA. 2000a. Toxic Substances Control Guidelines; Final Rule. Code of Federal Regulations, Title 40, Part 799. Available: http://www.epa.gov/oppt/chemtest/notice7.pdf. [accessed 2 June 2005].

EPA. 2000b. Toxicological Review of Chloral Hydrate. EPA/635/R-00/006. Washington, DC:U.S. Environmental Protection Agency. Available: www.epa.gov/iris. [accessed 2 June 2005].

EPA. 2002a. Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. EPA 712–C–02–190. Washington, DC: U.S. Environmental Protection Agency.

EPA. 2002b. Statements of Policies and Interpretations. Toxicological and Ecological Studies. 40 CFR 159.165. Available: http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].

EPA. 2002c. Health Effects Test Guidelines OPPTS 870.1000 Acute Toxicity Testing - Background. EPA 712–C–02–189. Washington, DC: U.S. Environmental Protection Agency.

EPA. 2002d. Emergency Planning and Notification. 40 CFR 355. Available: http://www.access.gpo.gov/nara/cfr/waisidx_00/40cfr355_00.html. [accessed 2 June 2005].

EPA. 2003a. Good Laboratory Practice Standards. Toxic Substances Control Act. 40 CFR 792. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/40cfr160_03.html [accessed 21 December 2005].

EPA. 2003b. Good Laboratory Practice Standards. Federal Insecticide, Fungicide, and Rodenticide Act. 40 CFR 160.

EPA. 2004. U.S. EPA HPV Challenge Program Submission. Propanoic acid, 2-hydroxy-, compd. with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CAS No. 68227-46-3]. Submitted by PPG Industries, Inc. Revised December 17, 2004. http://www.**epa**.gov/chemrtk/prop2hyd/c13863rt3.pdf [accessed 06 December 2005].

European Centre for the Validation of Alternative Methods. (ECVAM). 1997. Statement on the scientific validity of the 3T3 NRU PT test (an in vitro test for phototoxic potential). http://ecvam.jrc.it/index.htm. [accessed 15 August 2005].

EU. 2003. Report from the Commission to the Council and the European Parliament. Third Report from the commission to the Council and the European Parliament of the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. Brussels, 22.01.2003, COM (2003) 19 final.

Fautrel A, Chesné C, Guillouzo A, De Sousa G, Placidi M, Rahmani R, et al. 1993. A multicentre study of acute *in vitro* cytotoxicity in rat hepatocytes: tentative correlation between *in vitro* toxicity and *in vivo* data. Altern Lab Anim 21:281-284.

FDA. 1993. Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food. Center for Food Safety and Applied Nutrition. Washington, DC: U.S. Food and Drug Administration.

FDA. 2003. Good Laboratory Practice for Nonclinical Laboratory Studies. 21 CFR Part 58. Available: http://www.access.gpo.gov/nara/cfr/waisidx_03/21cfr58_03.html [accessed 21 December 2005].

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

Fentem JH, Archer GEB, Balls M, Botham PA, Curren RD, Earl LK, et al. 1998. The ECVAM international validation study on in vitro tests for skin corrosivity. 2. Results and evaluation by the Management Team. Toxicol In Vitro 12:483-524.

Fentem JH, Briggs D, Chesne C, Elliott GR, Harbell JW, Heylings JR, et al. 2001. A prevalidation study on in vitro tests for acute skin irritation: results and evaluation by the Management Team. Toxicol In Vitro 15:57-93.

Freshney RI. 2000. Culture of Animal Cells: A Manual of Basic Technique, 4th edition. New York:Wiley-Liss.

Garle M., Fentem JH, Fry JR. 1994. *In Vitro* cytotoxicity tests for the prediction of acute toxicity. *In Vivo*. Toxicol In Vitro 8:1303-1312.

Gennari A, van den Berghe C, Casati S, Castell J, Clemedson C, Coecke S, et al. 2004. Strategies to replace in vivo acute systemic toxicity testing. The report and recommendations of ECVAM Workshop 50. Altern Lab Anim 32:437-459.

Gettings SD, Bagley DM, Demetrulias JL, Dipasquale LC, Hintze KL, Rozen MG, et al. 1991. The CTFA evaluation alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase I) Hydro-alcoholic Formulations; (Part 2) Data analysis and biological significance. In Vitro Toxicol 4:247-288.

Gettings SD, Bagley DM, Chudkowski M, Demetrulias JL, Dipasquale LC, Galli CL et al. 1992. Development of potential alternatives to the Draize eye test - The CTFA evaluation of alternatives program - Phase II - Review of materials and methods. Altern Lab Anim 20:164-171.

Gettings SD, Dipasquale LC, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al. 1994a. The CTFA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase II) Oil/water emulsions. Food Chem Toxicol. 32:943-976.

Gettings SD, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, Curren RD, et al. 1994b. The CTFA evaluation of alternatives program: Phase III (surfactant-based formulations). World Congress on Alternatives and Animal Use in the Life Sciences, Baltimore, Maryland, USA, Nov. 14-19, 1993. In Vitro Toxicol 7:166.

Gettings SD, Lordo RA, Hintze KL, Bagley DM, Casterton PL, Chudkowski M, et al. 1996. The CFTA evaluation of alternatives program: An evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (Phase III) Surfactant-based Formulations. Food Chem Toxicol 34:79-117.

Gilman AG, Goodman LS, Rall TW, Murad F (eds.). 1985. Goodman and Gilman's The Pharmacological Basis of Therapeutics, 7th Edition. New York:Macmillan Publishing Co.

Glaxo Wellcome. 2000. Myleran,(busulfan) Product Information. Research Triangle Park, NC:Glaxo Wellcome, Inc.

GraphPad Software. 1994-2000. Prism® version 3.0a for Macintosh. GraphPad Software, San Diego, California. Available: www.graphpad.com. [accessed 21 November 2005].

Grisham JW, Smith GJ. 1984. Predictive and Mechanistic Evaluation of Toxic Responses in Mammalian Cell Culture Systems. Pharmacol Rev 36(Suppl.):151S-171S.

Gülden M, Dierickx P, Seibert H. 2006. Validation of a prediction model for estimating serum concentrations of chemicals which are equivalent to toxic concentrations in vitro. Toxicology in Vitro 20:1114-1124.

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

Hackenberg U, Bartling H. 1959. Messen und Rechnen im pharmakologischen Laboratorium mit einem speziellen Zahlensystem (WL24-System). Arch Exp Pathol Pharmakol 235:437-463.

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

Halle W. 1998. Toxizitätsprüfungen in Zellkulturen für eine Vorhersage der akuten Toxizität (LD_{50}) zur Einsparung von Tierversuchen. Life Sciences/ Lebenswissenschaften, Volume 1, 94 pp., Jülich: Forschungszentrum Jülich.

Halle W. 2003. The Registry of Cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. Altern Lab Anim 31:89-198. (*English translation of Halle 1998*)

Halle W, Spielmann H. 1992. Two procedures for the prediction of acute toxicity (LD50) from cytotoxicity data. Altern Lab Anim 20:40-49.

Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Gilman AG, eds. 1996. Goodman and Gilman's The Pharmacological Basis of Therapeutics, 9th Edition. New York:McGraw-Hill.

Harlan Company. 2002. Available: http://www.harlan.com/us/index.htm [accessed 2 June 2005]

Harbell JW, Koontz SW, Lewis RW, Lovell D, Acosta D. 1997. IRAG working group 4. Cell cytotoxicity assays. Interagency Regulatory Alternatives Group. Food Chem Toxicol 35:79-126.

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, and Lewis D. 2002. Good cell culture practice: ECVAM good cell culture practice task force report 1. Altern Lab Anim 30:407-414. Available: http://ecvam.jrc.it/publication/index5007.html [accessed 02 September 2005].

Heimann R, Rice RH. 1983. Rat esophageal and epidermal-keratinocytes - intrinsic differences in culture and derivation of continuous lines. J Cell Physiol 117:362-367.

Hintze KL, Janus J, Marenus KD, Muscatiello MJ, Pape WJW, Renskers KJ, et al. 1992. Development of potential alternatives to the Draize eye test - The CTFA evaluation of alternatives program - Phase II - Review of materials and methods. Altern Lab Anim 20:164-171.

Hunter WJ, Lingk W, Recht JP. 1979. Intercomparison study on the determination of single administration toxicity in rats. Journal of the Association of Official Analytical Chemists 62:864-873.

Hyclone®. 1996. Heat inactivation – are you wasting your time? Art to Science 15:1-5. Available: http://www.hyclone.com/library/arttoscience.htm. [accessed 27 July 2006].

ICCVAM. 2000. The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals and Products. Proposed Test Method and Background Review Document, April 14, 2000. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

ICCVAM. 2001a. Report Of The International Workshop On In Vitro Methods For Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

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

ICCVAM. 2001c. The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals. NIH Publication No. 02-4501. National Institute for

410 Environmental Health Sciences, Research Triangle Park, NC. Available:
411 http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
412
413 ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised,
414 and Alternative Test Methods. NIH Publication No. 03-4508. National Institute for
415 Environmental Health Sciences, Research Triangle Park, NC. Available:
416 http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].
417
418 ICCVAM. 2006. PEER REVIEW PANEL REPORT: The Use of *In Vitro* Basal Cytotoxicity
419 Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing.
420 http://iccvam.niehs.nih.gov/methods/invidocs/brdvalstdy.htm. [accessed August 2, 2006].
421
422 IOM. 2004. Forging a Poison Prevention and Control System. Washington: National
423 Academies Press.
424
425 INVITTOX. 1991. The FRAME Modified Neutral Red Uptake Cytotoxicity Test. Protocol
426 Number 3 (IP-3A). http://embryo.ib.amwaw.edu.pl/invittox/. [accessed 12 Sep 2005].
427
428 Karczmar A. 1998. Invited review: Anticholinesterases: dramatic aspects of their use and
429 misuse. Neurochem Int 32:401-411.
430
431 King AV, Jones PA. 2003. In-house assessment of a modified in vitro cytotoxicity assay for
432 higher throughput estimation of acute toxicity. Toxicol In Vitro. Oct-Dec;17(5-6):717-22.
433
434 Knox P, Uphill PF, Fry JR, Benford J, Balls M. 1986. The FRAME multicentre project on *in*
435 *vitro* cytotoxicity. Food Chem Toxicol 24:457-463.
436
437 Kohn J, Durham HD. 1993. S9 liver fraction is cytotoxic to neurons in dissociated culture.
438 Neurotoxicology 14:381-386.
439
440 Life Technologies, Inc. 1997. Culture of human keratinocytes in defined serum-free medium.
441 Focus 19:1-5. Available: http://www.invitrogen.co.jp/focus/191002.pdf. [accessed 27 July
442 2006].
443
444 Lipnick RL, Cotruvo JA, Hill RN, Bruce RD, Stitzel KA, Walker AP, et al. 1995.
445 Comparison of the up-and-down, conventional LD50, and fixed-dose acute toxicity
446 procedures. Food Chem Toxicol 33:223-231.
447
448 Litchfield JT, Wilcoxon F. 1949. A simplified method of evaluating dose-effect experiments.
449 J Pharmacol Exp Therap 96:99-113.
450
451 Litovitz TL, Klein-Schwartz W, White S, et al. 2000. 1999 Annual report of the american
452 association of poison control centers toxic exposure surveillance system. Am J Emerg Med
453 18:517-574.
454

455 The MathWorks. 1996-2004. MATLAB® software. The MathWorks, Inc., Natick,
456 Massachusetts.
457
458 Matsuno T, Masuda, Shimizu. 1971. Pipobroman: acute toxicity and general toxicological
459 reaction. Kiso to Rinsho Clinical Report 5:1894-1917.
460
461 MDL Information Systems 2001, 2002. Registry of Toxic Effects of Chemical Substances
462 (RTECS®) accessed online by subscription.
463
464 Mediatech, Inc. 2006. Heat Inactivation of Serum. Technical Information #S055.0303.001.
465 Available: http://cellgro.bizatomic.net/shop/customer/home.php. [accessed 27 July 2006].
466
467 Microsoft Corporation. 1998-2001. Microsoft® Excel® 2000 software for Mac®. Microsoft
468 Corporation, Redmond, Washington.
469
470 Miller LC, Tainter ML. 1944. Estimation of the LD50 and its error by means of logarithmic-
471 probit graph paper. Proc Soc Exp Biol Med 57:261-264.
472
473 Miyazaki H, Sekine T, Endou H. 2004. The multispecific organic anion transport family:
474 properties and pharmacological significance. Trends Pharmacol Sci 25:654-662.
475
476 National Toxicology Program. 2002. Annual Plan For Fiscal Year 2002. Public Health
477 Service, Department of Health and Human Services. NIH Publication No. 03-5309. National
478 Institute for Environmental Health Sciences, Research Triangle Park, NC. Available:
479 http://ntp.niehs.nih.gov/index.cfm?objectid=03611496-026C-BB0F-B2A1527014B2D4B1.
480 [accessed 2 June 2005].
481
482 National Library of Medicine (NLM). 2005. Haz-Map, Occupational Exposure to Hazardous
483 Agents, Specialized Information Services, National Institutes of Health, Department of
484 Health & Human Services. Available: http://hazmap.nlm.nih.gov/. [accessed 29 November
485 2005].
486
487 National Library of Medicine. 2000, 2001, 2002, 2005. Hazardous Substances Data Bank.
488 Accessed online via TOXNET. National Institutes of Health, Department of Health &
489 Human Services. Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB. [accessed
490 2 June 2005].
491
492 Occupational Safety and Health Act. 2003. Toxic and Hazardous Substances. Hazard
493 Communication 29 CFR 1910.1200
494 Available: http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
495
496 Occupational Safety and Health Act. 2000. Labeling Requirements for Pesticides and
497 Devices. Labeling Requirements. 40 CFR 156.10. Available:
498 http://www.gpoaccess.gov/cfr/index.html. [accessed 2 June 2005].
499
500 OECD. 1996. Guideline for Testing of Chemicals, 423, Acute Toxic Class Method. Paris

France: OECD.

OECD. 1998. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 1: OECD principles on Good Laboratory Practice. (as revised in 1997). ENV/MC/CHEM(98)17. Paris: OECD. Available: http://www.oecd.org [accessed 2 June 2004].

OECD. 1998a. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as Endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11.

OECD. 1998b. OECD Guideline for Testing of Chemicals 425: Acute oral toxicity—Up-and-Down Procedure. OECD, Paris.

OECD. 2000. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19. Paris France: OECD. Available: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]

OECD. 2001a. Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-Down Procedure. Paris France: OECD. Available: http://www.oecd.org [accessed 2 June 2004].

OECD. 2001b. OECD Series on Testing and Assessment, 33, Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures. ENV/JM/MONO(2001)6. Paris France: OECD. Available: http://www.oecd.org [accessed 12 September 2004].

OECD. 2001c. Guideline for Testing of Chemicals, 420, Acute Oral Toxicity – Fixed Dose Method. Paris France: OECD.

OECD. 2001d. Guideline For Testing of Chemicals, 423, Acute Oral Toxicity – Acute Toxic Class Method. Paris France: OECD.

OECD. 2001e. Guidance Document on Acute Oral Toxicity Testing. 24. Paris France: OECD. Available: http://www.oecd.org/ehs [accessed 9 October 2006].

OECD. 2004a. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 14: Advisory Document of the Working Group on Good Laboratory Practice: The Application of the Principles of GLP to In Vitro Studies. ENV/JM/MONO (2004)26. Paris: OECD.

OECD. 2004b. Guideline for Testing of Chemicals, 432, *In Vitro* 3T3 NRU Phototoxicity Test. Paris France: OECD. Available: http://www.oecd.org. [accessed 30 August 2005].

547 Orphan Medical, Inc. 1999. NDA 20-954 Busulfex™ (busulfan) Injection. NDA 20-954.
548 Available: www.fda.gov/cder/foi/label/1999/20954lbl.pdf. [accessed 2 June 2005].
549
550 Pesticide Action Network North America. 2005. Pesticide Action Network [PAN] Pesticides
551 Database. Available: http://www.pesticideinfo.org/Index.html. [accessed 29 November 2005]
552
553 Peloux AF, Federici C, Bichet N, Gouy D, Cano JP. 1992. Hepatocytes in primary culture: an
554 alternative to LD_{50} testing? Validation of a predictive model by multivariate analysis. Altern
555 Lab Anim 20:8-26.
556
557 Phillips JC, Gibson WB, Yam J, Alden CL, Hard GC. 1990. Survey of the QSAR and *in vitro*
558 approaches for developing non-animal methods to supersede the *in vivo* LD50 test. Food
559 Chem Toxicol 28:375-394.
560
561 Raabe HA, Moyer GO, Mun GC, Sizemore AM, Harbell JW, Merrill JC. 2004. Induction of
562 a zone of cell death in multi-well plates by refeeding [Abstract]. Toxicologist 78:50.
563 Available: http://www.iivs.org/pages/publication_view.php?doc_id=45. [accessed 28 August
564 2006].
565
566 Rasmussen ES. 1999 Cytotoxicity of MEIC chemicals nos. 11-30 in 3T3 mouse fibroblasts
567 with and without microsomal activation. In Vitr Mol Toxicol 12:125-132.
568
569 Riddell RJ, Panacer DS, Wilde SM, Clothier RH, Balls M. 1986. The importance of exposure
570 period and cell type in *in vitro* cytotoxicity tests. Altern Lab Anim 14:86-92.
571
572 Roguet R. Colovio J, Gaetani Q., Sossou KG, Rogier A. 1993. Cytotoxicity of 28 MEIC
573 chemicals to rat hepatocytes using two viability endpoints: correlation with acute toxicity
574 data in rat and man. Altern Lab Anim 21:216-224.
575
576 SAS Institute Inc. 1999. SAS/STAT User's Guide Version 8. Cary, NC:SAS Institute, Inc.
577
578 Schlede E, Mischke U, Roll R, Kayser D. 1992. A national validation study of the acute-
579 toxic-class method--an alternative to the LD50 test. Archives of Toxicology 66:455-470.
580
581 Schlede E, Mischke U, Diener W, Kayser D. 1995. The international study of the acute toxic
582 class method (oral). Archives of Toxicology 69:659-670.
583
584 Schmahl D, Osswald H. 1970. Experimental studies on the carcinogenic effects of anticancer
585 chemotherapeutics and immunosuppressive agents. Arzneimittelforschung 20:1461-1467.
586
587 Seibert H, Gulden M, Kolossa M, Schepers G. 1992. Evaluation of the relevance of selected
588 *in vitro* toxicity test systems for acute systemic toxicity. Altern Lab Anim 20:240-245
589
590 Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, et al. 1996. Acute
591 toxicity testing *in vitro* and the classification and labelling of chemicals. The report and
592 recommendations of ECVAM Workshop 16. Altern Lab Anim 24:499-510.

Sina JF, Galer DM, Sussman RG, Gautheron PD, Sargent EV, Leong B, et al. 1995. A collaborative evaluation of seven alternatives to the Draize eye irritation test using pharmaceutical intermediates. Fundam Appl Toxicol 26:20-31.

Sekine T, Cha SH, Endou H. 2000. The multispecific organic anion transporter (OAT) family. Pflugers Arch 440:337-350.

Spielmann H, Genschow E, Liebsch M, Halle W. 1999. Determination of the starting dose for acute oral toxicity (LD_{50}) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Lab Anim 27:957-966.

Spielmann H, Balls M, Dupuis J, Pape WJW, Pechovitch G, de Silva O, et al. 1998. EU/COLIPA "*In Vitro P*hototoxicity" Validation Study, Results of Phase II (blind trial), Part 1: the 3T3 NRU phototoxicity test. Toxicol In Vitro 12:305-327.

Spielmann H, Balls M, Dupuis J, Pape WJW, De Silva O, Holzhütter H.-G, et al. 1998b. A study on the phototoxic potential of UV filter chemicals from Annex VII of EU Directive 76/768/EEC in the 3T3 NRU *in vitro* phototoxicity test. Altern Lab Anim 26:679-705.

Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhütter HG, et al. 1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, the HET-CAM test and the 3T3 NRU cytotoxicity test. Altern Lab Anim 24**:**741-858.

Spielmann H, Kalweit S, Liebsch M, Wirnsberger T, Gerner I, Bertram-Neis E, et al. 1993. Validation study of alternatives to the Draize eye irritation test in Germany: Cytotoxicity testing and HET-CAM test with 136 industrial chemicals. Toxicol In Vitro 7:505-510.

Spielmann H, Gerner S, Kalweit S, Moog R, Wirnserberger T, Krauser K, et al. 1991. Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxicol In Vitro 5:539-542.

Stallard N, Whitehead A. 2004. A statistical evaluation of the fixed dose procedure. Altern Lab Anim 32 (suppl. 2):13-21.

Taconic Farms, Inc. 2002. Taconic Farms Animal Models, Sprague Dawley® Outbred Rats. Available: http://www.taconic.com/anmodels/spragued.htm. [accessed 2 June 2005].

Thompson WR. 1947. Use of moving averages and interpolation to estimate median-effective dose. Bacteriol Rev 11:115-145.

Thomson PDR®. 2004. Physicians' Desk Reference, 58th edition. Montvale, NJ: Thomson PDR.

638 Triglia D, Wegener PT, Harbell J, Wallace K, Mathesen D, Shopsis C. 1989. Interlaboratory
639 validation study of the keratinocyte neutral red bioassay from Clonetics Corporation. In: In
640 Vitro Toxicology, New Directions (Goldberg A, ed). Alternative Methods in Toxicology,
641 Volume 7. New York: Mary Ann Liebert, 357-365.
642
643 Tsao MC, Walthall BJ, Ham RG. 1982. Clonal growth of normal human epidermal-
644 keratinocytes in a defined medium. J Cell Physiol 110: 219-229.
645
646 UN. 1977. United Nations Economic and Social Council. Joint meeting of the RID safety
647 committee and the group of experts on the transportation of dangerous goods. Trans/GE 15/R
648 274, 2.
649
650 UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals
651 (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and
652 Geneva. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html
653 [accessed 12 September 2005].
654
655 Vernot EH, MacEwen JD, Haun CC, Kinkead ER. 1977. Acute toxicity and skin corrosion
656 data for some organic and inorganic compounds and aqueous solutions. Toxicol Appl
657 Pharmacol 42:417-423.
658
659 Wallum, E. 1998. Acute oral toxicity. Environ Health Perspect 106:497-503.
660
661 Weil CS. 1952. Tables for convenient calculation of median effective dose (LD50 or ED50)
662 and instructions in their use. Biometrics 8(3):249-263.
663
664 Weil CS, Wright GJ. 1967. Intra- and interlaboratory comparative evaluation of single oral
665 test. Toxicol Appl Pharmacol 11:378-388.
666
667 WHO. 1998. The WHO Recommended Classification of Pesticides by Hazard and
668 Guidelines to Classification 1998-1999. International Programme of Chemical Safety.
669 WHO/PCS/98.21. Geneva:World Health Organization.
670
671 Willshaw A, Moore LJ, Balls M. 1994. *In vitro* alternatives for the detection of photoirritant
672 chemicals - the EEC COLIPA trial. Toxicol In Vitro 8:723-725.
673
674 Zbinden G, Flury-Roversi M. 1981. Significance of the LD50-test for the toxicological
675 evaluation of chemical substances. Arch Toxicol 47:77-99.
676
677 Zuang V, Balls M, Botham PA, Coquett A, Corsini E, Curren RD, et al. 2002. Follow-up to
678 the ECVAM prevalidation study on in vitro tests for acute skin irritation. Altern Lab Anim
679 30:109-129.
680

[This Page Intentionally Left Blank]

13.0 GLOSSARY[1]

Accuracy[2]: (a) The closeness of agreement between a test method result and an accepted reference value. (b) The proportion of correct outcomes of a test method. It is a measure of test method performance and one aspect of "relevance". Accuracy is highly dependent on the prevalence of positives in the population being examined.

Acute Toxic Class (ATC) method: An acute oral systemic toxicity test method based on testing groups of animals at fixed doses in a sequential manner. The lethality outcomes are used to classify a test substance into the appropriate GHS acute oral toxicity category.

Adjusted R^2: R^2 values that are adjusted for the relative proportion of data points to explanatory variables. Adjusted $R^2 = 1 - (1 - R^2)[(n – 1)/(n - k - 1)]$ where k = number of independent variables and n = number of observations. See "coefficient of determination."

ANOVA: One-way (and two-way) analysis of variance. ANOVA compares the measurements (continuous variables) of three or more groups when the data are categorized in one way (one-way) or two ways (two-way). ANOVA assumes that the populations compared are normally distributed and that the variances for the groups to be compared are approximately equal.

Assay[2]: The experimental system used. Often used interchangeably with "test" and "test method."

Biphasic dose-response: Dose-response in which cytotoxicity increases (as dose increases), plateaus, and then increases again. See **Section 2.6.3**.

[1] The definitions in this Glossary are restricted to their uses with respect to *in vitro* cytotoxicity testing and the NRU test methods.
[2] Definition used by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM 2003).

Category prediction: The GHS hazard category that includes the predicted LD_{50} value for a test chemical.

Coded substances: Substances labeled by code rather than name so that they can be tested and evaluated without knowledge of their identity or anticipation of test results. Coded substances are used to avoid intentional or unintentional bias when evaluating laboratory or test method performance.

Coefficient of determination: In linear regression, it denotes the proportion of the variance in Y and X that is shared. Its value ranges between zero and one and it is commonly called called "R^2." For example, $R^2 = 0.45$, indicates that 45% of the variance in Y can be explained by the variation in X and that 45% of the variance in X can be explained by the variation in Y.

Coefficient of variation: A statistical representation of the precision of a test. It is expressed as a percentage and is calculated as follows:

$$\left(\frac{standard\ deviation}{mean}\right) \times 100\%$$

Concordance[2]: The proportion of all substances tested that are correctly classified as positive or negative. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "accuracy." Concordance is highly dependent on the prevalence of positives in the population being examined. In the NICEATM/ECVAM study, concordance was used to describe the proportion of test substances that were correctly classified into GHS acute oral toxicity hazard categories, or to describe the proportion of test substances for which the laboratories obtained the same classification result.

Confluence: A state in which cells in culture come into contact with other cells in the same culture to form a complete sheet of cells (monolayer). For this study, confluence is

57 determined as a percentage of cell coverage of the tissue culture vessel growth surface (e.g.,
58 cell monolayer has 80% confluency).
59
60 **Cytotoxicity:** The adverse effects resulting from interference with structures and/or
61 processes essential for cell survival, proliferation, and/or function. For most chemicals,
62 toxicity is a consequence of non-specific alterations in "basal cell functions" (i.e., via
63 mitochondria, plasma membrane integrity, etc.), which may then lead to effects on organ-
64 specific functions and/or death of the organism. These effects may involve the integrity of
65 membranes and the cytoskeleton, cellular metabolism, the synthesis and degradation or
66 release of cellular constituents or products, ion regulation, and cell division.
67
68 **Definitive test:** The main test of the cytotoxicity assay for determining the IC_{50}. The
69 concentration closest to the range finder test IC_{50} serves as the midpoint of the concentrations
70 tested in a definitive test. Compared to the range finder test, the definitive test uses a smaller
71 dilution factor for the concentrations tested.
72
73 **Discordant chemicals:** Chemicals for which the LD_{50} is not accurately predicted by the IC_{50}
74 (and the associated regression formula) or the GHS toxicity category is not accurately
75 predicted by the IC_{50} (and the associated regression formula). Also referred to as "outliers."
76
77 **EDIT:** Evaluation-guided Development of New *In vitro* Test Batteries. An international
78 project coordinated by the Scandinavian Society for Cell Culture to develop new *in vitro* tests
79 for toxicity and toxicokinetics to be incorporated into test batteries for predicting acute and
80 chronic systemic toxicity.
81
82 **Endpoint[2]:** The biological process, response, or effect assessed by a test method.
83
84 **Fixed Dose Procedure (FDP):** An acute oral systemic toxicity test method based on testing
85 groups of animals at fixed doses. Evident toxicity outcomes are used to classify a test
86 substance into the appropriate GHS acute oral toxicity category.
87

F_G: An empirical factor for the RC regression line that represents the expected precision of LD_{50} predictions from basal cytotoxicity data. The LD_{50} values of 73% of the 347 RC chemicals are localized in the dose range around the RC regression line by $\mathbf{F_G} \leq \log 5$. The factor represents the expected difference between the LD_{50} determined in animal experiments and the LD_{50} estimated from the IC_{50} on the RC regression line.

Geometric mean: The antilog of the mean of the logarithm of the values. It is less affected by extreme values than the arithmetic mean.

Globally Harmonized System (GHS): A classification system presented by the United Nations that provides (a) a harmonized criteria for classifying substances and mixtures according to their health, environmental and physical hazards, and (b) a harmonized hazard communication elements, including requirements for labeling and safety data sheets.

Good Laboratory Practices (GLP)[2]: Regulations promulgated by the U.S. Food and Drug Administration and the U.S. Environmental Protection Agency, and principles and procedures adopted by the Organization for Economic Cooperation and Development and Japanese authorities that describe record keeping and quality assurance procedures for laboratory records that will be the basis for data submissions to national regulatory agencies.

Guidance Document: *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b).

Hazard[2]: The potential for an adverse health or ecological effect. A hazard potential results only if an exposure occurs that leads to the possibility of an adverse effect being manifested.

Hill function: The IC_{50} values are determined from the concentration-response using a Hill function which is a four parameter logistic mathematical model relating the concentration of the test chemical to the response (typically following a sigmoidal shape).

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log IC50 - X)\text{HillSlope}}}$$

118 where Y= response, X is the logarithm of dose (or concentration), Bottom is the minimum
119 response, Top is the maximum response, $\log IC_{50}$ is logarithm of X at the response midway
120 between Top and Bottom, and HillSlope describes the steepness of the curve.
121
122 **Hill function (revised):** Some unusual dose-responses did not fit the Hill function well. To
123 obtain a better model fit, the Bottom parameter was estimated without constraints (the
124 previous practice was to use Bottom = 0). However, when Bottom ≠ 0, the EC_{50} reported by
125 the Hill function was not the same as the IC_{50} since the Hill function relies on EC_{50} defined
126 as the point midway between Top and Bottom. Thus, the Hill function calculation using the
127 Prism® software was rearranged to calculate the concentration corresponding to the IC_{50} as
128 follows.
129

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

130
131 X is the logarithm of concentration at 50% response, $\log EC_{50}$ is logarithm of concentration at
132 the response midway between Top and Bottom, Top is the maximum response, Bottom is the
133 minimum response, Y = 50 (i.e., 50% response), and HillSlope describes the steepness of the
134 curve.
135
136 **Hormesis:** a dose-response characterized by a compound's ability to cause a stimulatory
137 effect in low doses or an inhibitory effect in high doses.
138
139 **IC_{50}:** test chemical concentration producing 50% inhibition of the endpoint measured (i.e.,
140 cell viability).
141
142 **Interlaboratory reproducibility[2]:** A measure of whether different qualified laboratories
143 using the same protocol and test substances can produce qualitatively and quantitatively
144 similar results. Interlaboratory reproducibility is determined during the prevalidation and

145 validation processes and indicates the extent to which a test method can be transferred
146 successfully among laboratories.
147
148 **Intralaboratory repeatability[2]:** The closeness of agreement between test results obtained
149 within a single laboratory when the procedure is performed on the same substance under
150 identical conditions within a given time period.
151
152 **Intralaboratory reproducibility[2]:** The first stage of validation; a determination of whether
153 qualified people within the same laboratory can successfully replicate results using a specific
154 test protocol at different times.
155
156 ***In vitro:*** In glass. Refers to assays that are carried out in an artificial system (e.g., in a test
157 tube or petri dish) and typically use single-cell organisms, cultured cells, cell-free extracts, or
158 purified cellular components.
159
160 ***In vivo:*** In the living organism. Refers to assays performed in multicellular organisms.
161
162 **K_{ow}**: Octanol:water partition coefficient.
163
164 **LC_{50}:** Acute lethal serum or blood concentrations.
165
166 **LD_{50}**: The calculated value of the oral dose that produces lethality in 50% of test animals
167 (rats and mice). The LD_{50} values serve as reference values for the *in vitro* tests.
168
169 **LD_{50} (initial):** Acute oral rat and mouse LD_{50} values used during the chemical selection
170 process. For RC chemicals, LD_{50} values were those used in the RC regression, which were
171 largely from the 1983/84 RTECS®. For chemicals that were not included in the RC, the
172 initial LD_{50} values came from HSDB or 2002 RTECS®.
173
174 **LD_{50} (reference):** Acute oral rodent LD_{50} values from rats and mice were located through
175 literature searches and references from major toxicity databases such as RTECS®. Studies

176 were reviewed to identify the most appropriate LD_{50} values for each chemical. Values
177 obtained using feral animals, preanesthetized animals, or animals less than 4 weeks of age
178 were not used. Values reported as inequalities were not used. Reference LD_{50} values were
179 determined by calculating the geometric mean of the acceptable LD_{50} values. Data were used
180 in generation of the 3T3 and NHK NRU regressions.
181
182 **Maximum:minimum value**: Ratio of minimum acceptable LD_{50} to maximum acceptable
183 LD_{50}.
184
185 **MEIC:** Multicentre Evaluation of *In Vitro* Cytotoxicity. An international effort established
186 by the Scandinavian Society for Cell Toxicology and initiated in 1983 to evaluate the
187 relationship and relevance of *in vitro* cytotoxicity for predicting the acute toxicity of
188 chemicals in humans.
189
190 **Millimolar regressions:** Linear regressions with IC_{50} values in mmol/L and LD_{50} values in
191 mmol/kg.
192
193 **Negative control:** An untreated sample containing all components of a test system, except
194 the test substance solvent, which is replaced with a known non-reactive material, such as
195 water. This sample is processed with test substance-treated samples and other control
196 samples to determine whether the solvent interacts with the test system.
197
198 **Neutral red (NR):** A weakly cationic water-soluble dye that stains living cells by readily
199 diffusing through the plasma membrane and concentrating in lysosomes where it
200 electrostatically binds to the anionic lysosomal matrix.
201
202 **Neutral red uptake (NRU):** Concentration of neutral red dye in the lysosomes of living
203 cells. Altering the cell surface or the lysosomal membrane by a toxicological agent causes
204 lysosomal fragility and other adverse changes that gradually become irreversible. The NRU
205 test method makes it possible to distinguish between viable, damaged, or dead cells because

206 these changes result in decreased uptake and binding of NR measurable by optical density
207 absorption readings in a spectrophotometer.
208
209 **NHK: N**ormal **H**uman epidermal **K**eratinocytes (from neonatal foreskin).
210
211 **Optical density (OD):** The absorption (i.e., OD measurement) of the resulting colored
212 solution (colorimetric endpoint) in the NRU assay measured at 540 nm ± 10 nm in a
213 spectrophotometric microtiter plate reader using blanks as a reference
214
215 **Outlier:** For any measurement, an extreme value in the NICEATM/ECVAM study was
216 referred to as an "outlier" if it passes a statistical test for outliers at the 99% level. With
217 respect to chemicals, it refers to chemicals that do not fit (using the specified criteria) an
218 IC_{50}-LD_{50} linear regression model. It may also refer to chemicals for which the predicted
219 GHS toxicity category does not match the reference *in vivo* GHS toxicity category.
220
221 **Performance[2]:** The accuracy and reliability characteristics of a test method (see "accuracy",
222 "reliability").
223
224 **pH:** A measure of the acidity or alkalinity of a solution. pH 7.0 is neutral; higher pHs are
225 alkaline, lower pHs are acidic.
226
227 **Plate reader:** A spectrophotometric device for measuring light intensity as a function of
228 color/wavelength (i.e., optical density/absorption at 540 nm ± 10 nm for NRU) in 96-well
229 microtiter tissue culture plates.
230
231 **Positive control:** A sample containing all components of a test system and treated with a
232 substance known to induce a positive response, which is processed with the test substance-
233 treated and other control samples to demonstrate the sensitivity of each experiment and to
234 allow for an assessment of variability in the conduct of the assay over time.
235

236 **Predictivity**[2]: Proportion of *in vivo* category matches for all substances with *in vitro*
237 predictions for a particular category. Predictivity is an indicator of test accuracy.
238
239 **Protocol**[2]**:** The precise, step-by-step description of a test, including the listing of all
240 necessary reagents, criteria and procedures for the evaluation of the test data.
241
242 **Quality assurance (QA)**[2]**:** A management process by which adherence to laboratory testing
243 standards, requirements, and record keeping procedures is assessed independently by
244 individuals other than those performing the testing.
245
246 **Range finder:** Initial test performed to determine starting doses for the main (definitive) test.
247 The NRU assays test eight concentrations of the test chemical or the positive control (PC) by
248 diluting the stock solution in log dilutions to cover a large concentration range.
249
250 **RC regression:** log (LD_{50}) = 0.435 x log (IC_{50}) + 0.625; for estimating an LD_{50} value in
251 mmol/kg (body weight) from an IC_{50} value (in mM).
252
253 **Reduction alternative**[2]**:** A new or modified test method that reduces the number of animals
254 required.
255
256 **Reference substances:** Substances selected for use during the research, development,
257 prevalidation, and validation of a proposed test method because their response in the *in vivo*
258 reference test method or the species of interest is known (see "reference test"). Reference
259 substances should represent the classes of chemicals for which the proposed test method is
260 expected to be used and cover the range of expected responses (negative, weak to strong
261 positive).
262
263 **Reference test method**[2]**:** The accepted *in vivo* test method used for regulatory purposes to
264 evaluate the potential of a test substance to be hazardous to the species of interest.
265

Refinement alternative[2]: A new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being.

Registry of Cytotoxicity (RC): Database that consists of *in vivo* acute oral toxicity data (i.e., LD_{50} values) from rats and mice and *in vitro* cytotoxicity data (i.e., IC_{50} values) from multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998). A regression model constructed from these data was proposed by ZEBET, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral systemic toxicity tests

Relevance[2]: The extent to which a test method correctly predicts or measures the biological effect of interest in humans or another species of interest. Relevance incorporates consideration of the "accuracy" or "concordance" of a test method.

Reliability[2]: A measure of the degree to which a test method can be performed reproducibly within and among laboratories over time. It is assessed by calculating intra- and inter-laboratory reproducibility and intralaboratory repeatability.

Replacement alternative[2]: A new or modified test method that replaces animals with nonanimal systems or one animal species with a phylogenetically lower one (e.g., a mammal with an invertebrate).

Reproducibility[2]: The consistency of individual test results obtained in a single laboratory (intralaboratory reproducibility) or in different laboratories (interlaboratory reproducibility) using the same protocol and test substances (see intra- and inter-laboratory reproducibility).

RTECS®: Registry of Toxic Effects for Chemical Substances**.** Compendium of data extracted from the open scientific literature. The database includes toxicity data (e.g., acute toxicity) and specific numeric toxicity values (e.g., LD_{50}). Compiled by the U.S. National Institute for Occupational Safety and Health (NIOSH) and now licensed to MDL Information Systems, Inc.

297 **Sensitivity[2]:** The proportion of all positive substances that are classified correctly as positive
298 in a test method. It is a measure of test method accuracy.
299
300 **Simulation modeling:** Computer simulation modeling of the acute systemic toxicity assays
301 to determine animal use. The simulation process uses a simulated population of animals for
302 testing, a reference endpoint (i.e., "true" LD_{50} value), and its assumed log-normal
303 distribution. Mortality is assumed to have a mean equal to the log of the true LD_{50}. The SD,
304 which reflects the variability of the simulated population, is the inverse of the slope of the
305 dose-mortality curve. Due to a lack of information for the real dose-mortality curve, the
306 simulations assumed slopes of 0.5, 0.8, 2, 4, and 8.3.
307
308 **Solubility:** The amount of a test substance that can be dissolved (or thoroughly mixed with)
309 culture medium or solvent. The solubility protocol was based on a U.S. EPA guideline (EPA
310 1998) that involves testing for solubility in a particular solvent, beginning at a relatively high
311 concentration and proceeding to successively lower concentrations by adding more solvent as
312 necessary for dissolution. Testing stops when, upon visual observation, the procedure
313 produces a clear solution with no cloudiness or precipitate.
314
315 **Solvent control:** An untreated sample containing all components of a test system, including
316 the solvent that is processed with the test substance-treated and other control samples to
317 establish the baseline response for the samples treated with the test substance dissolved in the
318 same solvent. When tested with a concurrent negative control, this sample also demonstrates
319 whether the solvent interacts with the test system.
320
321 **Specificity[2]:** The proportion of all negative substances that are classified correctly as
322 negative in a test method. It is a measure of test method accuracy.
323
324 **Spirit of GLP:** Guidance provided in the Statement of Work specifically for the non GLP-
325 compliant laboratory that participated in the validation study. Based on the GLP standards
326 referenced in the ECVAM Workshop 37 Report (Cooper-Hannan 1999) and the OECD
327 Principles of GLP (OECD 1998). "*Laboratories that are non GLP-compliant shall adhere to*

GLP principles and other method parameters. Documentation and accountability shall be equal to GLP requirements. Laboratories must make assurances that they are equal in performance criteria and that there is parity amongst the laboratories."

TESS: Toxic Exposure Surveillance System. A comprehensive poisoning surveillance database maintained by the American Association of Poison Control Centers (AAPCC).

Test[2]: The experimental system used; used interchangeably with "test method" and "assay".

Test method[2]: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay". See also "validated test method" and "reference test".

Test method component: Structural, functional, and procedural elements of a test method that are used to develop the test method protocol. These components include unique characteristics of the test method, critical procedural details, and quality control measures.

3T3: BALB/c 3T3 clone A31 mouse fibroblasts developed in 1968 from disaggregated 14- to 17-day-old BALB/c mouse embryos (American Type Culture Collection [ATCC]; # CCL-163).

Tiered testing: A testing strategy where all existing information on a test substance is reviewed, in a specified order, before *in vivo* testing.

Toxicity underpredicted: Actual LD_{50} value of a test substance is lower than the predicted LD_{50} value.

Toxicity overpredicted: Actual LD_{50} value of a test substance is higher than the predicted LD_{50} value.

359 **Transferability[2]:** The ability of a test method or procedure to be accurately and reliably
360 performed in different, competent laboratories.
361
362 **Up-and-Down Procedure (UDP):** An acute oral systemic toxicity test method used to
363 minimize the number of animals required to estimate the acute oral toxicity of a chemical,
364 estimate the LD_{50} and confidence intervals (CI), and observe signs of toxicity. Single animals
365 are tested sequentially. Subsequent doses are based on the outcome of the previous animal.
366
367 **Validated test method[2]:** An accepted test method for which validation studies have been
368 completed to determine the accuracy and reliability of this method for a specific proposed
369 use.
370
371 **Validation[2]:** The process by which the reliability and accuracy of a procedure are established
372 for a specific purpose.
373
374 **Vehicle control (VC):** The VC consists of appropriate cell culture medium for the cells in
375 the test (i.e., DMEM for 3T3 cells and keratinocyte growth medium for the NHK cells). For
376 chemicals dissolved in DMSO, the VC consists of medium with the same amount of solvent
377 as that used in the test chemical concentrations that are applied to the 96-well test plate. The
378 final DMSO concentration is ≤ 0.5 % (v/v) in the VCs.
379
380 **Volatility:** Ability of a test chemical to evaporate. A general indicator of volatility issues in
381 the NRU test methods is the percent difference in the mean OD values for the two VC
382 columns on the test plate. If the difference is greater than 15%, then chemical volatility can
383 be suspected, especially if the VC adjacent to the highest test concentration had a
384 significantly reduced OD value. Volatility may be an issue for compounds with a specific
385 gravity of less than 1.
386
387 **Weight of evidence (process):** The strengths and weaknesses of a collection of information
388 are used as the basis for a conclusion that may not be evident from the individual data.

Weight regressions: Linear regressions with IC_{50} values in µg/mL and LD_{50} values in mg/kg.

ZEBET: The German National Center for the Documentation and Evaluation of Alternative Methods to Animal Experiments.