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 9-7
9.1.3 Other Evaluations of 3T3 or NHK NRU Methods 9-10

9.2 Independent Scientific Reviews 9-13
9.2.1 *In Vitro* Acute Toxicity Testing for the Classification and Labeling of Chemicals..... 9-13
9.2.2 Use of *In Vitro* Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing 9-14
9.2.3 Validation of 3T3 NRU Assay for Phototoxicity 9-17

9.3 Studies Using *In Vitro* Cytotoxicity Methods with Established Performance Standards..... 9-18
9.3.1 *Guidance Document* (ICCVAM 2001b)..... 9-18
9.3.2 King and Jones (2003) 9-19
9.3.3 A-Cute-Tox Project: Optimization and Pre-Validation of an *In Vitro* Test Strategy for Predicting Human Acute Toxicity (Clemedson 2005)..... 9-19

9.4 Summary..... 9-20

[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 these 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.**

9.1.1.1 *Peloux et al. (1992)*

The authors used several different *in vitro* cytotoxicity methods with primary rat hepatocytes to determine the correlation with rat/mouse intraperitoneal (i.p.) or intravenous (i.v.) LD_{50} values for the 25 substances tested. The *in vitro* cytotoxicity methods, which used 20-hour test substance exposure durations, assessed the following endpoints: NRU, total protein content, LDH release, MTT reduction. MTT is metabolized by mitochondrial succinate dehydrogenase of viable cells to yield a purple formazan reaction product. The IC_{50} values

obtained using the four endpoints were highly correlated (r = 0.973 to 0.999) to each other. When performing the IC_{50}-LD_{50} regressions, Peloux et al. (1992) used the lowest reported published LD_{50} value for acute rat or mouse studies that administered the test substances using the i.p. or i.v. routes. The IC_{50} values obtained using NRU as the endpoint had the highest correlation coefficient, r = 0.877, to the rat/mouse i.p./i.v. LD_{50} values. The total protein assay yielded r = 0.872, the MTT reduction assay yielded r = 0.808, and the LDH release assay yielded r = 0.789.

Peloux et al. (1992) followed the recommendations of Fry et al. (1988, 1990) and used parenteral LD_{50} values rather than oral LD_{50} values for comparison with *in vitro* values. Fry et al. (1988, 1990) recommended the use of the i.p./i.v. LD_{50} values for comparisons because they proposed that cells *in vivo* receive a more direct test substance exposure via these routes than through the oral route. They had posited that *in vitro* cell cultures would mirror this (direct) toxicity because they also receive direct exposure to test substances via the cell culture medium. The authors also noted that the oral route of exposure presents confounding variables such as, 1) only a fraction of a test substance would be available in the systemic circulation due to limited absorption or pre-systemic metabolism, and 2), the level of the substance in the systemic circulation decreases due to elimination mechanisms (e.g., metabolism, excretion). Fry et al. (1990) had reported a correlation of only r = 0.49 for *in vivo/in vitro* comparisons of oral LD_{50} and IC_{50} values (from a total protein assay) and a correlation of r = 0.68 for i.p. LD_{50} and ID_{50} values[1].

9.1.1.2 *Fautrel et al. (1993)*

Six laboratories tested the cytotoxicity of 31 substances in primary rat hepatocyte cultures using a 24-hour exposure followed by measurement of NRU. The investigators performed linear regression analyses for the prediction of rat i.v., i.p., and oral LD_{50} values from the NRU IC_{50} values. The regressions for the various *in vivo* administration routes did not use the same substances because LD_{50} values were not available for all of the tested substances in all of the routes. Oral, i.v., and i.p. LD_{50} values were available for 27, 24, and 18 substances, respectively, and IC_{50} values were obtained for 15, 14, and 11 of these substances, respectively. The regression for the i.v. data was statistically significant (r = 0.88, n = 11), but the i.p. (r = 0.48, n = 14) and oral regressions (r = 0.17, n = 15) were not. The finding that the i.v. LD_{50} values corresponded more closely with the *in vitro* cytotoxicity data than did the oral LD_{50} was thought to be the result of having fewer pharmacokinetic variables (i.e., absorption, distribution, etc.) to consider following i.v. administration.

9.1.1.3 *Roguet et al. (1993)*

Roguet et al. (1993) tested the cytotoxicity of 28 MEIC substances in primary rat hepatocytes exposed for 21 hours, followed by the measurement of NRU. A correlation of the NRU IC_{50} values to oral LD_{50} values obtained from the unpublished data of B. Ekwall et al. (personal communication) yielded a statistically significant linear correlation ($p < 0.001$) with r = 0.80 when the *in vivo* and *in vitro* data were in molar units. [NOTE: The LD_{50} values subsequently published by Ekwall et al. (1998) were from the 1997 edition of RTECS®.] The authors reported that the toxicities of thioridazine, malathion, and *copper sulfate* were overestimated, and the toxicity of *potassium cyanide* was underestimated by the correlation, but their criteria for over- and under- estimation were not provided.

[1] ID_{50}: index of cytotoxicity; concentrations (µg/mL) producing a 50% reduction in protein value.

The *in vivo* toxicity of *potassium cyanide* was also underpredicted in the NICEATM/ECVAM validation study. **Table 6-3** shows that *potassium cyanide* was an outlier for which toxicity was underpredicted when using the IC_{50} values from both the 3T3 and NHK NRU test methods in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log IC_{50} mM + 0.625). The GHS category predictions using both NRU test methods and the RC rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621), and the RC rat-only weight regression (i.e., log LD_{50} =0.372 log IC_{50} + 2.024), were also higher (i.e., less toxic) than the *in vivo* category (see **Appendix L2**).

9.1.1.4 *Rasmussen (1999)*

Twenty MEIC substances were tested for cytotoxicity using NRU release from 3T3 cells following 24-hr exposure, with and without the addition of a Aroclor-induced rat liver microsomal preparation (S9 mix). Similar to the present validation study, Rasmussen (1999) observed that *xylene* was non-toxic to the cells, even though it was dissolved in ethanol instead of DMSO. In the presence of S9, the cytotoxicities of malathion, 2,4-dichlorophenoxyacetic acid, *propranolol*, thioridazine, *lithium* sulfate, *copper sulfate*, and *thallium sulfate*, were significantly decreased ($p < 0.05$), while the cytotoxicities of *1,1,1-trichloroethane*, *phenol*, *nicotine*, and *paraquat* were significantly increased ($p < 0.05$).

Because the NICEATM/ECVAM validation study used cells with little or no xenobiotic metabolizing capability, it could be expected that these systems would overpredict the toxicity of substances that would be inactivated by the addition of a metabolizing system, or to underpredict the toxicity of substances that are metabolized to more toxic substances. None of the four substances in common for which toxicity was decreased by the addition of S9 were overpredicted in the NICEATM/ECVAM study. However, the toxicities of two of the four substances in common for which toxicity was increased by the addition of S9, were underpredicted in the NICEATM/ECVAM study. **Table 6-3** shows that *nicotine* was an outlier whose toxicity was underpredicted when using the 3T3 and NHK IC_{50} values in the RC millimole regression (log LD_{50} mmol/kg = 0.435 log IC_{50} mM + 0.625). *Paraquat* was an outlier whose toxicity was underpredicted when using the NHK IC_{50} value in the RC millimole regression. The GHS category predictions for both substances using both NRU test methods with the RC rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621) and the RC rat-only weight regression (log LD_{50} mg/kg = 0.357 log IC_{50} μg/mL + 2.194) were also higher than the *in vivo* category (see **Appendix L2**).

Although both the IC_{20} and IC_{50} values were determined in the Rasmussen (1999) study, only the IC_{20} values were used for correlations with the rat acute oral LD_{50} values from RTECS®. The units of the LD_{50} values were not reported, but the correlations were assumed to be in molar units because the IC_{20} and IC_{50} values were reported in μM units. Significant correlations ($p < 0.001$) between IC_{20} and LD_{50} values were obtained with and without rat liver microsomes. The correlation of IC_{20} with LD_{50} was slightly higher with the S9 mix (r = 0.72 vs. 0.68 for oral LD_{50} values, and 0.82 vs. 0.78 for i.p. LD_{50} values).

Although the presence of S9 increased the cytotoxicity of some substances to the 3T3 cells, it decreased the toxicity of others, and yielded only a small improvement in the correlation to *in vivo* data. Rasmussen (1999) concluded that the toxicity of the S9 mix (0.32 mg protein/mL), itself, was insignificant because it reduced cell survival by less than 10% compared with cells

without S9. However, others have shown that S9 microsomal mixes could produce significant cytotoxic effects. Kohn (1993) showed that an S9 mix containing 0.07 mg protein/mL was cytotoxic to all types of murine neurons in culture when the cells were exposed for four days or longer. Non-neuronal cells tolerated higher concentration exposures of S9, but exhibited cytoplasmic inclusions when exposed to S9 at 0.35 mg protein/mL. Dal Negro et al. (2006) reported 100% cell death of human monocyte-derived U-937 cells when the S9 fraction (1 mg protein/mL) and co-factors were applied to the cells for a 72-hour incubation. Both of these studies used longer exposure durations, and/or higher protein concentrations, than the Rasmussen (1999) study.

9.1.1.5 *Creppy et al. (2004)*

Creppy et al. (2004) used a 48-hour NRU assay to determine the cytotoxicity of ochratoxin A (OTA) and fumonisin B1 (FB1) on cultured C6 glioma (rat brain), Caco-2 (human intestinal), and Vero (green monkey kidney) cells. The IC_{50} determined in the NRU assay was used in the RC millimole regression to predict rodent acute oral LD_{50} values. The predicted LD_{50} for OTA using the C6 glioma cells was similar to mouse LD_{50} values generated from four *in vivo* mouse studies, but the LD_{50} values predicted by the other cell lines were about 50 times greater. The authors found the relative insensitivity of the Vero cells surprising because OTA is a kidney toxin. There were no available *in vivo* rodent oral LD_{50} values with which to compare the predicted LD_{50} of FB1, which ranged from 671 to 924 mg/kg for the three cell types tested.

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}.

Table 9-3 EU[1] Classes of Acute Oral Toxicity

Category	LD_{50} (mg/kg)
1	$LD_{50} \leq 25$
2	$25 < LD_{50} \leq 200$
3	$200 < LD_{50} \leq 2000$
Unclassified	$LD_{50} > 2000$

Abbreviations: EU=European Union
[1]Anon (1993)

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 in category 1 for $LD_{50} \leq 25$ mg/kg. Thus, the classification of digoxin using the 4-category system required six animals.

Using such simulations of ATC testing, Halle et al. (1997) estimated that 2139 animals would be used to test the 347 substances:

- Three hundred twenty-eight would require testing with two doses using three test animals each.
- Nineteen would require testing with three doses using three animals each.

Halle et al. (1997) cited Schlede et al. (1995) in reporting that the average number of animals required to classify substances using the ATC method was 9.11 animals per test. Using this average, ATC testing of the 347 RC substances would require 3161 animals. Thus, Halle et al. (1997) estimated that there would be a 32% reduction ([3161-2139]/3161) in the number of test animals used when the LD_{50} prediction from the RC millimole regression was used with the 1996 version of the ATC method, in lieu of the standardanimal classification procedure (Halle et al. 1997).

The simulated average animal savings for the ATC in the NICEATM/ECVAM validation study at dose-response slopes of 2.0 and 8.3 was 4.8% to 10.2% (0.51 to 1.09 animals) for the 3T3 (67 reference substances) and NHK (68 reference substances) NRU test methods (see **Section 10.3.3.2**), depending on the regression evaluated. This is considerably lower than the average savings of 32% estimated by Halle et al. (1997). However, there are a number of differences between the evaluation performed by Halle et al. (1997) and the 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 evaluated were outliers (28/70 [40%] in the 3T3 NRU test method, and 31/71 [44%] in the NHK NRU test method) (see **Table 6-3**). 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*

Spielmann et al. (1999) recommended 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 experimental LD_{50} reported by Lipnick et al. (1995). Spielmann et al. (1999) concluded that the RC millimole regression provided an adequate prediction of LD_{50}, and that *in vitro* cytotoxicity data could be used to predict starting doses for the UDP. The authors recommended using the IC_{50}, with the RC millimole regression, to calculate a starting dose (i.e., an estimated LD_{50}) for the UDP, FDP, or ATC method whenever an IC_{50} was available.

If no IC_{50} was available, Spielmann et al. (1997) recommended determining cytotoxicity using a standard cell line and specific cytotoxic endpoint (e.g., NRU, total protein, MTT reduction). They recommended testing 10 to 20 RC substances to demonstrate that the *in vitro* cytotoxicity test methods provide results that are consistent with the RC millimole regression. The resulting IC_{50} values would then be used to calculate a new regression (using the LD_{50} values reported in the RC), which would be compared to the RC millimole regression. If the new regression fit into the acceptance interval ($\pm \log 5$ of the fitted regression line) of the RC millimole regression, the RC millimole regression would be used to predict starting doses for the UDP. If the new regression is parallel to the RC millimole regression, but outside the $\pm \log 5$ acceptance interval, then the new regression would be used for the prediction of the starting dose.

Spielmann et al. (1999) contended that the RC millimole regression provides a sufficient prediction of LD_{50} values from IC_{50} values for substances that do not require metabolic

activation and are not very toxic (i.e., LD_{50} > 200 mg/kg). The authors acknowledged that the fit of substances with LD_{50} <200 mg/kg to the RC millimole regression is not good, and attributed the poor fit of these substances to the need for metabolic activation to a more toxic substance. They suggested that the prediction of starting doses using cytotoxicity data can be used with the UDP and ATC methods, but not with the FDP because dosing is not sequential (which contradicted a claim made earlier in the paper that the approach could be used with the FDP). They did not estimate the number of animals that might be saved with this approach, but did recommend that the approach be validated experimentally using several established cell lines with a limited number of representative substances from the RC.

9.1.2.3 *EPA (2004): U.S. EPA HPV Challenge Program Submission*

In response to the EPA HPV Chemical Challenge Program, PPG Industries, Inc., the manufacturer of Propanoic acid, 2-hydroxy-, compound with 3-[2-(dimethylamino)ethyl] 1-(2-ethylhexyl) (4-methyl-1,3-phenylene)bis[carbamate] (1:1) [CASRN 68227-46-3], and the sponsor of this compound, submitted data to the EPA. This is an isolated intermediate used to produce a resin component of paint products. PPG provided the following types of data in their submission to the EPA: physical-chemical, environmental fate and pathway, ecotoxicity, and toxicology. The acute mammalian toxicology data were generated using both *in vitro* and *in vivo* methods.

An *in vitro* NRU cytotoxicity test was conducted with 3T3 cells to estimate a starting dose for the *in vivo* acute UDP oral toxicity test (OECD 2001a) (see **Appendix M1** for the OECD UDP test guideline). The use of this *in vitro* NRU test method was intended to minimize the number of animals used for *in vivo* testing. The estimated LD_{50} of the compound as determined by the NRU assay was 489 mg/kg. Therefore, the starting dose for the UDP study was set at 175 mg/kg, which is the first default dose below the estimated LD_{50} value; this is also the default starting dose for the UDP, and is used when no information on which to base a starting dose is available. A total of fifteen female rats received the compound at 175, 550, or 2000 mg/kg. Five of nine rats treated at 2000 mg/kg died prematurely on Days 2 and 3, and by Day 15, 2/4 surviving animals at this dose had lost up to 25% of their Day 1 body weights. The LD_{50} was estimated to be 2000 mg/kg, with a 95% confidence interval of 1123-5700 mg/kg. Thus, the *in vitro* NRU test method overpredicted the toxicity of the compound by estimating an LD_{50} value that was lower than that determined in the UDP test. The report authors reported that a greater than predicted number of animals was used for the UDP testing because the estimated LD_{50}, 489 mg/kg and, consequently, the starting dose, was much lower than the *in vivo* LD_{50} of 2000 mg/kg. However, because the UDP started with the default starting dose of 175 mg/kg, the claim that more animals were used is incorrect, because animal use with the default starting dose is the baseline against which other animal use should be compared.

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%
- Sensitivity (percentage of true irritants correctly detected) = 80%
- Predictive values (probability that an unknown agent will be properly classified)
  - Positive predictive value = 90%
  - Negative predictive value = 87%

The authors reported that there was excellent correlation among the laboratories, and good correlation between the *in vitro* IC_{50} values and *in vivo* Draize scores (Spearman Rank correlation coefficients between *in vivo* and *in vitro* data for the laboratories ranged from 0.67-0.76). The authors also concluded that the NRU test could not replace the Draize test, but may be an effective screening tool for use in a battery of *in vitro* alternatives

Sina et al. (1995)

Sina et al (1995) evaluated the NHK NRU test method along with six other *in vitro* methods to determine whether they could be used as complimentary tests in a battery approach to estimate ocular irritation. The NRU data correlated poorly with Draize ocular scores for the 33 pharmaceutical intermediates tested. The Spearman correlation coefficient for the IC_{50} and maximum average Draize score (MAS) was -0.10, and the Pearson correlation coefficient was -0.04.

[2] Clonetics® Corporation sponsored this study. It was not clear in the publication if Clonetics® Corporation participated as one of the testing laboratories.

Brantom et al. (1997)
This study examined the potential of 10 alternative methods to predict the eye irritation potential of cosmetic ingredients. Four laboratories tested 55 coded substances (23 single ingredients and 32 formulations) using the 3T3 NRU test method, and used the resulting IC_{50} values to predict modified maximum average scores (MMAS) for the Draize test.

An endpoint was generated for each test by interpolation from a plot of percent cell survival versus test substance concentration. A prediction model was developed from data of 30 single ingredients (29 surfactants and one substance not classified by the authors) to equate the IC_{50} value to an MMAS.

The interlaboratory CV for the IC_{50} values was 37.3 ± 29.8% (7.5 ± 6.8, log transformed). Most of the mean IC_{50} values from a single laboratory differed by plus or minus an order of magnitude from the means of all the laboratories for each substance, which the authors interpreted as "no significant outliers". Correlations of NRU-predicted MMAS scores with *in vivo* MMAS scores yielded Pearson's r values ranging from 0.25 to 0.32 for the four laboratories.

Although the authors concluded the interlaboratory reproducibility was good, the IC_{50} values did not predict the MMAS. The r values for the *in vitro*/*in vivo* correlations were low (0.246 to 0.316) and the tests all underpredicted irritants and overpredicted non-irritants. Four substances were outside of the 95% confidence intervals and the authors concluded that the 3T3 NRU test method had wide applicability to test the remaining 51 coded substances according to the limitations in the prediction model, but that it was not effective as a stand-alone replacement for the Draize test across the entire irritation scale. The authors did not identify the test substances.

Harbell et al. (1997)
This publication reported the results of the evaluation of 12 *in vitro* cytotoxicity assays to predict ocular irritation. Data were voluntarily submitted to the U.S. Interagency Regulatory Alternatives Group (IRAG), composed of members from CPSC, EPA, and FDA. The NHK NRU test method was one of the tests evaluated by six laboratories testing surfactants and surfactant-containing formulations (the 3T3 NRU test method was not tested). Two laboratories submitted results for the same test substances, but the other four submitted data for various sets of substances and formulations.

The correlation of results from the two laboratories that independently tested the same substances was r=0.99. Correlations between the IC_{50} data and *in vivo* maximum average Draize score (MAS) ranged from -0.92 to -0.54. The IRAG concluded that the assays were suitable as a screening and adjunct assay to assess eye irritation over the range of toxicities found in personal care and household products, and recommended that its use be limited to water-soluble materials. Although the method was also evaluated for surfactants, IRAG recommended that the evaluation continue for its performance in predicting eye irritation for various product classes (e.g., fabric softeners, shampoos). In addition, the substance's physical form should be considered because the *in vitro* toxicity of a solution of the test substance will not necessarily predict toxicity of the parent, solid substance *in vivo*.

9.1.3.2 *Predicting Human Lethal Blood Concentrations (LC)*

Seibert et al. (1992)

This single laboratory study was designed to evaluate various aspects of cellular toxicity in four *in vitro* test systems for their relevance and reliability with respect to acute systemic toxicity, in particular, human LC. The 3T3 NRU test method was one of four methods evaluated with 10 MEIC substances.

The authors stated that final conclusions on the relevance of the *in vitro* systems for *in vivo* data could not be determined because the variations in LC were unknown so that limits for over or underprediction of human *in vivo* toxicity using experimental models could not be defined. In addition, the ability of *in vitro* toxicity to predict *in vivo* toxicity may depend on toxicokinetic factors that were not considered in the *in vitro* systems.

9.2 Independent Scientific Reviews

This section summarizes independent scientific reviews of the use of *in vitro* cytotoxicity methods for the prediction of rodent acute oral toxicity, and for the reduction of animal use in acute toxicity testing. The conclusions of these reviews are compared to the conclusions of the current study. Also discussed is the 3T3 NRU phototoxicity method, because it is similar to the 3T3 NRU test method used in the current validation study and has been validated by ECVAM and is the subject of OECD Test Guideline 432 (OECD 2004).

9.2.1 *In Vitro* Acute Toxicity Testing for the Classification and Labelling of Chemicals

9.2.1.1 *Seibert et al. (1996): ECVAM Workshop 16*

ECVAM sponsored a workshop in 1994 to review the current status of various *in vitro* methods and to determine their potential uses for reducing, refining, and/or replacing the use of laboratory animals for acute systemic toxicity testing. The workshop participants reviewed various types of toxicity, *in vitro* cytotoxicity testing schemes and strategies, inclusion of biokinetic parameters, biotransformation, biodistribution *in vitro* and *in vivo*, and a proposed acute toxicity testing scheme for the classification of substances.

The workshop participants agreed that some studies showed good correlations between *in vitro* cytotoxicity data and LD_{50} values. They also acknowledged that *in vitro* basal cytotoxicity tests could not address all the different of mechanisms of acute systemic toxicity. Additional approaches to replacing animals would have to incorporate the three main types of cellular level toxic effects that can lead to in acute systemic toxicity (i.e., basal cytotoxicity, selective toxicity, and cell-specific function toxicity). The participants determined that it is also important that any alternative method take into account the active concentration and meaningful dose of a test substance in an *in vitro* cell culture system. Quantitative comparisons of test substance concentrations must be made to evaluate the effects of the test substances regarding the three types of cytotoxicity.

The biokinetics of a test substance (determined by its absorption, distribution, metabolism, and elimination) must be considered when making predictions of *in vivo* toxicity using *in vitro* toxicity data. Various methods can be used to convert *in vitro* effective concentrations of a test substance to equivalent body doses. Test substance factors, such as physicochemical characteristics (e.g., pKa, lipophilicity, volatility), estimates of protein binding, and *in vitro*

characteristics (e.g., cell concentration, cell protein concentration, ratio of cell/medium volumes, medium albumin concentration), are needed for such conversions.

An *in vitro* tiered testing scheme was proposed by the workshop participants for using *in vitro* methods to determine the acute oral toxicity of a substance:

- Stage 1: Basal cytotoxicity test
- Stage 2: Hepatocyte-specific cytotoxicity test to assess the role of biotransformation in producing toxicity
- Stage 3: Test system that evaluates non-hepatocyte-specific selective cytotoxicity (i.e., effects on cell-specific functions)

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 the NICEATM/ECVAM study (see **Section 6.4)**, the *in vitro* basal cytotoxicity tests are not suitable as replacements for rodent acute oral toxicity tests and could only be used as an adjunct test, and not a stand-alone test, for classifying substances for acute oral toxicity. However, *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 significantly different 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 7validation 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*

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 for the UDP (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 evaluated. 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 two 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 regression. The true *in vivo* LD_{50} for phenylthiourea is 3 mg/kg. Workshop 2000 participants used a best-case scenario when they assumed that *in vitro* cytotoxicity precisely predicted the true LD_{50}.

9.2.3 Validation of the 3T3 NRU Assay for Phototoxicity

An NRU assay using 3T3 cells was validated by ECVAM, and accepted for regulatory use, to detect the phototoxic potential of test substances. The 3T3 NRU test for phototoxicity requires a 60-minute exposure to the test substance, a 50-minute exposure to ultraviolet (UVA, 315-400 nm) light, followed by removal of test substance and incubation for another 24 hours in fresh medium (Spielmann et al. 1998). NR medium is then added, and NRU is measured after a 3-hour incubation. Phototoxic potential is assessed by comparing the differences in cytotoxicity between test plates containing the test substance that have not been exposed to UVA and comparable test plates exposed to UVA.

Two different models, employing the Photoinhibition Factor (PIF) and the Mean Photo Effect (MPE), were validated for the prediction of *in vivo* phototoxic potential. The accuracy of the models for classifying the phototoxic potential of the 30 substances tested in nine laboratories was 88% for the PIF, and 92% for the MPE, when compared with *in vivo* classifications. Interlaboratory variability for classification (i.e., phototoxic vs. non-phototoxic) was assessed using a bootstrapping approach. For each substance, the classification based on a single experiment was compared to the classification based on the mean PIF or mean MPE. The interlaboratory variability for classification was 0 to 18.8% using PIF and 0 to 20% using MPE.

The ECVAM Scientific Advisory Committee confirmed the scientific validity of the method in 1997 (ECVAM 1997) and its regulatory acceptance was noted in Annex V of Council Directive 67/548/EEC part B.41 on phototoxicity, in 2000. An OECD Test Guideline, 432, was finalized in 2004 (OECD 2004). The 3T3 NRU phototoxicity test is used in a tiered testing approach to determine the phototoxic potential of test substances.

The performance of the 3T3 NRU phototoxicity assay could not be compared with the 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 epidermal 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 NRU test 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 **Sections 2.6** and **2.7** 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 *Guidance Document* Figures 3 and 4, p.13 [ICCVAM 2001b]).

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 calls 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.
- Explore innovative tools and cellular systems to identify new toxicity end-points and strategies to better anticipate animal and human toxicity.
- To design a simple, robust and reliable *in vitro* test strategy associated with the prediction model for acute toxicity that is amenable to high-throughput testing.

The project has been divided into the following workpackages that will be implemented by various configurations of research partners:

- WP1: Generation of a "high quality" *in vivo* database (through literature searches and historical data) and establishment of a depository list of reference substances

- WP2: Generation of a "high quality" *in vitro* database (including data from the NICEATM/ECVAM study, EDIT studies, and MEIC studies)
- WP3: Iterative amendment of the testing strategy
- WP4: New end-points and new cell systems
- WP5: Alerts and correctors in toxicity screening (I): Role of absorption, distribution, and excretion
- WP6: Alerts and correctors in toxicity screening (II): Role of metabolism
- WP7: Alerts and correctors in toxicity screening (III): Role of target organ toxicity (i.e., neuro-, nephro-, hepato-toxicity)
- WP8: Technical optimisation of the amended test strategy
- WP9: Pre-validation of the test strategy

A-Cute-Tox aims to extend the NICEATM/ECVAM and MEIC/EDIT approaches toward a full replacement test strategy by improving the prediction of acute oral toxicity using *in vitro* methods, and then validating the testing procedure.

9.4 Summary

- *In vitro* NRU cytotoxicity test methods using various cell types have been evaluated for their correlation with rodent lethality endpoints (e.g., rat/mouse 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 LC values, *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%.