Table 2 Main findings of 4-ethylphenol (4EP) at the end of the dosing in the newborn and the young rat main studies

	Newborn rat study (mg/kg/day)				Young rat study (mg/kg/day)			
	0	30	100	300	0	100	300	1000
Male								
No. animals examined	12	12	12	12	14	7	7	14
Clinical toxic signs†	1‡	0	0	10	0	0	0	11
Death	0	0	0	0	0	0	0	0
Delayed righting reflex	0	0	0	4*				
No. animals examined	6	6	6	6	7	7	7	7
ALT (IU/L)	27 ± 7	21 ± 5	23 ± 2	25 ± 4	24 ± 3	24 ± 1	28 ± 3	41 ± 9**
Total cholesterol (mg/dL)	82 ± 13	83 ± 14	84 ± 8	91 ± 5	66 ± 6	58 ± 8	63 ± 9	68 ± 9
Relative liver weight (g/100 g BW)	3.37 ± 0.14	3.39 ± 0.22	3.40 ± 0.13	3.68 ± 0.16**	3.13 ± 0.18	3.28 ± 0.18	3.46 ± 0.16**	3.58 ± 0.17**
Relative kidney weight (g/100 g BW)	1.18 ± 0.05	1.17 ± 0.08	1.17 ± 0.06	1.22 ± 0.07	0.80 ± 0.05	0.79 ± 0.05	0.79 ± 0.05	0.89 ± 0.03**
Forestomach, hyperplasia	0	0	0	0	0	0	1	7
Female								
No. animals examined	12	12	12	12	14	7	7	14
Clinical toxic signs†	0	0	0	12	0	0	0	9
Death	0	0	0	2§	0	0	0	0
Delayed righting reflex	0	0	1	1				
No. animals examined	6	6	6	5	7	7	7	7
ALT (IU/L)	19 ± 3	20 ± 3	20 ± 2	19 ± 1	22 ± 8	21 ± 2	20 ± 2	27 ± 4
Total cholesterol (mg/dL)	80 ± 11	84 ± 11	85 ± 12	85 ± 23	61 ± 13	69 ± 10	65 ± 5	82 ± 14**
Relative liver weight (g/100 g BW)	3.25 ± 0.12	3.26 ± 0.05	3.37 ± 0.11	3.63 ± 0.23**	3.07 ± 0.17	2.99 ± 0.15	3.12 ± 0.12	3.47 ± 0.21**
Relative kidney weight (g/100 g BW)	1.21 ± 0.11	1.17 ± 0.05	1.20 ± 0.05	1.26 ± 0.07	0.82 ± 0.04	0.84 ± 0.06	0.83 ± 0.05	0.88 ± 0.05
Forestomach, hyperplasia	0	0	0	0	0	0	0	6

Values are given as the mean ± SD. *$P < 0.05$ and **$P < 0.01$ indicate significantly different from control group. BW, body weight.

†Hypoactivity, hypothermia, tremor, straub tail, deep respiration or emaciation for newborn rats and salivation, staggering gait, prone/lateral position or soiled perigenital fur for young rats; ‡straub tail casually occurred on PND 9; §each female died on day 10 and 12 of dosing.

28-Day study of 4EP in young rats

In the dose-finding study, 4/5 males and all 5 females at 2000 mg/kg/day died after the first dosing and the remaining 1/5 males was killed because of moribundity on day 3 of dosing. At 1000 mg/kg/day, 1/5 females showed soiled perineal fur on days 5–7 of dosing and then died on day 8 of dosing. The body weight of females was significantly lower on day 2 of dosing at 1000 mg/kg/day. Significantly high values of ALT and total cholesterol at 1000 mg/kg/day and significantly high value of ALT at 500 mg/kg/day were detected in males. Significantly low value of alkaline phosphatase and significantly high value of potassium at 1000 mg/kg/day were detected in females. In the necropsy findings for rats died during the dosing period, acute changes, such as red coloration of the lung, forestomach and kidney, thinning of the mucosa in the glandular stomach, discoloration of the liver and spleen, blood pooling in the urinary bladder and abdominal dropsy were observed at 2000 mg/kg/day and reddish spots of the glandular stomach and atrophy of the thymus and spleen were detected at 1000 mg/kg/day. For the surviving rats, thickening of the mucosa in the forestomach was observed in 2/5 males and 3/4 females at 1000 mg/kg/day at the end of the dosing period. At 1000 mg/kg/day, significantly high values of the relative liver weight in males and females and a significantly low value of relative spleen weight in females were observed. At 500 mg/kg/day, a significantly low value of relative spleen weight in females was observed.

In the main study (Table 2 and Fig. 5), clinical signs, such as salivation, staggering gait, a lateral position and soiled perigenital fur, were observed in 11/14 males and 9/14 females at 1000 mg/kg/day. At this dose, salivation for males and females was observed within 30 min after dosing daily from day 6 to the end of the dosing period. Staggering gait and a lateral position were occasionally observed in males and females for 1 h from a few minutes after dosing, and soiled perigenital fur was occasionally observed for males and females. Significantly low body weights from days 7–28 of dosing in males and from days 14–28 in females were also observed. In urinalysis, a significantly high volume of urine was observed in females at 1000 mg/kg/day at the end of the dosing period. In the blood biochemistry, significantly high values of ALT in males and total cholesterol in females at 1000 mg/kg/day were observed. In the necropsy findings, thinning of the mucosa in the glandular stomach in 5/7 males and 6/7 females and reddish spots in the glandular stomach in 1/7 females were observed at 1000 mg/kg/day at the end of the dosing period. Significantly high values of relative liver weight at 300 and 1000 mg/kg/day in males and at 1000 mg/kg/day in females were observed at the end of the dosing period. Significantly high value of relative kidney weight at 1000 mg/kg/day in males was observed at the end of the dosing period. Erosion, hyperplasia of squamous cells, degeneration of squamous cells and/or edema of the submucosa in the forestomach was observed in all 7 males at 1000 mg/kg/day. Hyperplasia of squamous cells in the forestomach was observed in 1/7 males at 300 mg/kg/day. Hyperplasia of squamous cells in the esophagus, degeneration of squamous cells, edema of the submucosa, granulation of the submucosa, hyperplasia of squamous cells and/or ulcer in the forestomach were observed in 6/7 females at 1000 mg/kg/day. There were no effects of 4EP treatment at the end of the recovery period except for the lowered body weight of males at 1000 mg/kg/day.

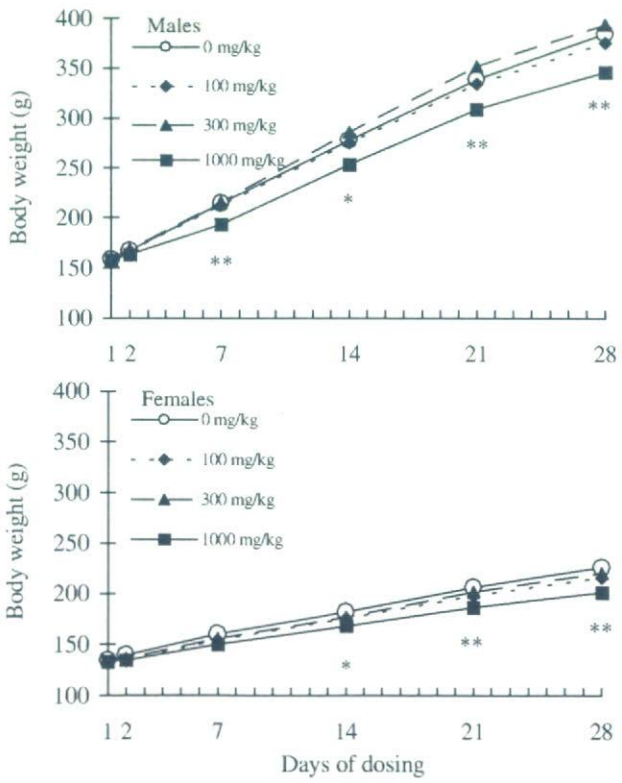

Fig. 5 Body weight curves in 28-day study of 4-ethylphenol (4EP) in young rats.

DISCUSSION

In the present paper, we determined the toxicity of 3EP and 4EP in newborn rats and reevaluated the toxicity of these chemicals in young rats, then compared the susceptibility of newborn rats in terms of NOAEL and UETL with that of young rats.

As for the administration of 3EP, NOAEL in the newborn rat study was concluded to be 100 mg/kg/day based on the lowered body weight at 300 mg/kg/day, although an increase in relative liver weight in females with no histopathological change and no changes in parameters of blood biochemistry related to liver damage was observed at 100 mg/kg/day in the main study. NOAEL in the young rat study was concluded to be 300 mg/kg/day based on the clinical toxic signs (staggering gait, prone/lateral position, tremor and soiled perigenital fur), changes in the liver (high values of weight and ALT or total cholesterol) and lesions in the forestomach at 1000 mg/kg/day. As clear toxicity did not appear in the newborn rat study even at the highest dose, we were not able to estimate UETL for 3EP.

As for the administration of 4EP, NOAEL in the newborn rat study was concluded to be 30 mg/kg/day based on the delay in the development of the righting reflex at 100 mg/kg/day. At 300 mg/kg/day, most animals showed clinical toxic signs and some females died in both the main and dose-finding studies. NOAEL in the young rat study was concluded to be 100 mg/kg/day, based on the lesions in the forestomach at 300 mg/kg/day. At 1000 mg/kg/day, clinical toxic signs were observed in all animals with the lesions in the forestomach. At this dose, no animal died in the main study but 1/5 females died in the dose-finding study (data not shown). When the dose of 1000 mg/kg/day for young rats was presumed as a UETL, which was the minimum lethal dose expecting the possibility of one female death, equivalent UETL for newborn rats was considered to be in the range of 200–250 mg/kg/day because 2/12

and 2/5 females died at 300 mg/kg/day in the main and dose-finding newborn studies, respectively.

In the newborn rat studies, slightly lowered body weight was observed after 3EP treatment, and deaths, hypoactivity, Straub tail, deep respiration and a delayed righting reflex were clearly observed after 4EP treatment. In the young rat studies, salivation, staggering gait, changes in the liver, including high values of liver weight and ALT or total cholesterol and lesions in the forestomach were clearly observed after 3EP and 4EP treatments. As for NOAEL, the susceptibility of newborn rats to 3EP and 4EP was approximately 3 times higher than that of young rats. The reason that newborn rats had higher susceptibility than young rats could be that newborn rats have immature metabolic activity, thus oxidation and conjugation of 3EP or 4EP in their livers would occur less, and toxic effects of the parent chemicals would continue longer.

The change of the mucosa and lesions of the submucosa and squamous cells in the forestomach caused by the corrosiveness of 3EP and 4EP were observed in young rats, but not in newborn rats. Generally, the phenols have similar toxicological effects and phenol is a protoplasmic poison and extremely corrosive (Bloom & Brandt 2001; Manahan 2003). 3EP and 4EP are irritating to the eyes, skin, mucous membranes and upper respiratory tract (Lenga 1985). Histopathological findings were not observed in the newborn rat study at any dose. The fact could be expected from the assumption that the membrane of the gastrointestinal tract of newborn rats would be more quickly renewed than that of young rats because of a higher turnover rate of the gastric membrane in developing newborn rats (Majumdar & Johnson 1982).

Methylphenol is an analog chemical of ethylphenol. Methylphenols or cresols, including three isomers, were reviewed as to their toxicity, and they have strong skin irritation and induce symptoms of poisoning (ASTDR 1992; WHO 1995; Stouten 1998). These reviews show that 4-methylphenol is more toxic than 3-methylphenol on the repeated-dose toxicity. In the present study, severer lesions in the forestomach were found after administration of 4EP than with 3EP in young rats. 4EP was also more toxic than 3EP in the newborn rat study. Deaths occurred after administration of 4EP.

Based on NOAEL, the susceptibility of newborn rats to 3EP and 4EP appeared to be almost 3 times higher than that of the young rats, being consistent with our previous results for four chemicals, 4-nitrophenol, 2,4-dinitrophenol, 3-aminophenol and 3-methylphenol, which showed 2–4 times differences in the toxic response between newborn and young rats. As for 3EP, unequivocal toxicity was not observed in the newborn rat study. As for 4EP, UETL in the young rat study was 4–5 times higher than that in the newborn rat study. In conclusion, newborn rats were 3–5 times more susceptible to 3EP and 4EP than young rats.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Office of Chemical Safety, Pharmaceutical and Medical Safety Bureau, Ministry of Health, Labour and Welfare, Japan.

REFERENCES

Agency for Toxic Substances and Disease Registry (ASTDR) (1992) *Toxicological Profile for Cresols*. US Public Health Service. ASTDR, Atlanta.

Bloom JC, Brandt JT (2001) Toxic responses of the blood. In: Klaassen CD (ed.). *Casarett and Doull's Toxicology: the Basic Science of Poisons*, 6th edn. McGraw-Hill, New York, pp. 389–417.

Chemical Products' Handbook (2004) *Chemical Products of 14504 '14504 no Kagakushohin'*;. The Chemical Daily, Tokyo (in Japanese).

Fukuda N, Ito Y, Yamaguchi M *et al.* (2004) Unexpected nephrotoxicity induced by tetrabromobisphenol A in newborn rats. *Toxicol Lett* **150**: 145–155.

Hirata-Koizumi M, Kusuoka O, Nishimura N *et al.* (2005) Susceptibility of newborn rats to hepatotoxiciy of 1,3-dibromopropane and 1,1,2,2-tetrabromoethane, compared with young rats. *J Toxicol Sci* **30**: 29–42.

Horikawa Y, Yamashita M, Morino K, Koyama S, Wada M, Maki S (1998) Industrialization of the process for cyancoupler intermediate production. *Res Dev Rep Sumitomo Chem* **2**: 44–48.

Koizumi M, Yamamoto Y, Ito Y *et al.* (2001) Comparative study of toxicity of 4-nitrophenol and 2,4-dinitrophenol in newborn and young rats. *J Toxicol Sci* **26**: 299–311.

Koizumi M, Nishimura N, Enami T *et al.* (2002) Comparative toxicity study of 3-aminophenol in newborn and young rats. *J Toxicol Sci* **27**: 411–421.

Koizumi M, Noda A, Ito Y *et al.* (2003) Higher susceptibility of newborn than young rats to 3-methylphenol. *J Toxicol Sci* **28**: 59–70.

Lenga RE (ed.). (1985) *The Sigma-Aldrich library of chemical safety data*. Sigma-Aldrich Corp, Milwaukee.

Majumdar APN, Johnson LR (1982) Gastric mucosal cell proliferation during development in rats and effects of pentagastrin. *Am J Physiol* **242**: G135–G139.

Manahan SE (2003) *Toxicological Chemistry and Biochemistry*, 3rd edn. Lewis publishers, Florida.

Ministry of Health, Labour and Welfare, Japan (MHLW) (2001a) Twenty-eight-day repeated dose oral toxicity test of 3-ethylphenol in rats. *Toxicity Testing Reports of Environmental Chemicals* 8: 750–758. MHLW, Japan.

Ministry of Health, Labour and Welfare, Japan (MHLW) (2001b) Twenty-eight-day repeated dose oral toxicity test of 4-ethylphenol in rats. *Toxicity Testing Reports of Environmental Chemicals* 8: 555–566. MHLW, Japan.

Ministry of Health and Welfare Japan (MHW) (1988) *Standard Concerning Testing Facility Provided in Article 4 of Order Prescribing Test Items Relating to New Chemical Substances and Toxicity Research of Designated Chemical Substances*. Planning and Coordination Bureau, Environment Agency, no. 39, Environmental Health Bureau, Ministry of Health and Welfare, no. 229, Basic Industries Bureau, Ministry of International Trade and Industry, no. 85, March 31, 1984, and amendments, November 18, 1988. MHW, Japan.

Organisation for Economic Cooperation and Development (OECD) (2004a) *The 2004 OECD List of High Production Volume Chemicals*. OECD, Paris.

Organisation for Economic Cooperation and Development (OECD) (2004b) *Draft Guidance Document on Reproductive Toxicity Testing and Assessment*. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 43, November 10, 2004 (First Version). OECD, Paris.

Schwenk M, Gundert-Remy U, Heinemeyer G *et al.* (2002) Children as a sensitive subgroup and their role in regulatory toxicology: DGPT workshop report. *Arch Toxicol* **77**: 2–6.

Stouten H (1998) Cresols (o-, m-, p-). DECOS and SCG basis for an occupational standard. *Arbete Och Hälsa* **27**: 1–44.

Takahashi M, Ogata H, Izumi H *et al.* (2004) Comparative toxicity study of 2,4,6-trinitrophenol (picric acid) in newborn and young rats. *Congenit Anom Kyoto* **44**: 204–214.

Thompson DC, Perera K, London R (1995) Quinone methide formation from para isomers of methylphenol (cresol), ethylphenol, and isopropylphenol: Relationship to toxicity. *Chem Res Toxicol* **8**: 55–60.

US Environmental Protection Agency (US EPA) (1998) *The EPA Children's Environmental Health Yearbook*. US EPA, Washington DC.

US Environmental Protection Agency (US EPA) (2002) *Child-Specific Exposure Factors Handbook*. US EPA, Washington DC.

World Health Organization (WHO) (1995) *Cresols, Environmental Health Criteria 168*. International Programme on Chemical Safety. WHO, Geneva.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Mutation Research 633 (2007) 67–79
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Current issues

Relevance and follow-up of positive results in *in vitro* genetic toxicity assays: An ILSI-HESI initiative☆

Véronique Thybaud[a], Marilyn Aardema[b], Daniel Casciano[c], Vicki Dellarco[d], Michelle R. Embry[e,*], B. Bhaskar Gollapudi[f], Makoto Hayashi[g], Michael P. Holsapple[e], David Jacobson-Kram[h], Peter Kasper[i], James T. MacGregor[j], Robert Rees[k]

[a] *Drug Safety Evaluation, sanofi-aventis, 94400 Vitry sur Seine, France*
[b] *Procter & Gamble Co., Miami Valley Innovation Center, 11810 East Miami River Road, Cincinnati, OH 45239-8707, USA*
[c] *Dan Casciano & Associates, 47 Marcella Dr., Little Rock, AR 72223, USA*
[d] *Office of Pesticide Programs, U.S. Environmental Protection Agency, 1200 Pennsylvania Ave., N.W., Washington, DC 20460, USA*
[e] *ILSI Health and Environmental Sciences Institute, One Thomas Circle, NW, Ninth Floor, Washington, DC 20005-5802, USA*
[f] *Toxicology and Environmental Research and Consulting, The Dow Chemical Company, Building 1803, Washington Street, Midland, MI 48642, USA*
[g] *Division of Genetics and Mutagenesis, National Institute of Health Sciences, 1-81-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*
[h] *Office of New Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 20993, USA*
[i] *Federal Institute for Drugs and Medical Devices (BfArM), Kurt-Georg-Kiesinger Allee 3, D-53175 Bonn, Germany*
[j] *Toxicology Consulting Services, 201 Nomini Drive, Arnold, MD 21012, USA*
[k] *Genetic Toxicology, GlaxoSmithKline, Park Road, Ware Herts SG12 0DP, UK*

Received 15 May 2007; accepted 17 May 2007
Available online 24 May 2007

Abstract

In vitro genotoxicity assays are often used to screen and predict whether chemicals might represent mutagenic and carcinogenic risks for humans. Recent discussions have focused on the high rate of positive results in *in vitro* tests, especially in those assays performed in mammalian cells that are not confirmed *in vivo*. Currently, there is no general consensus in the scientific community on the interpretation of the significance of positive results from the *in vitro* genotoxicity assays. To address this issue, the Health and Environmental Sciences Institute (HESI), held an international workshop in June 2006 to discuss the relevance and follow-up of positive results in *in vitro* genetic toxicity assays. The goals of the meeting were to examine ways to advance the scientific basis for the interpretation of positive findings in *in vitro* assays, to facilitate the development of follow-up testing strategies and to

☆ This document represents the consensus of the participants' views expressed as individual scientists and does not necessarily represent the policies and procedures of their respective institutions.

* Corresponding author. Tel.: +1 202 659 3306; fax: +1 202 659 3617.
E-mail address: membry@hesiglobal.org (M.R. Embry).

1383-5718/$ – see front matter
doi:10.1016/j.mrgentox.2007.05.010

define criteria for determining the relevance to human health. The workshop identified specific needs in two general categories, *i.e.*, improved testing and improved data interpretation and risk assessment. Recommendations to improve testing included: (1) re-examine the maximum level of cytotoxicity currently required for *in vitro* tests; (2) re-examine the upper limit concentration for *in vitro* mammalian studies; (3) develop improved testing strategies using current *in vitro* assays; (4) define criteria to guide selection of the appropriate follow-up *in vivo* studies; (5) develop new and more predictive *in vitro* and *in vivo* tests. Recommendations for improving interpretation and assessment included: (1) examine the suitability of applying the threshold of toxicological concern concepts to genotoxicity data; (2) develop a structured weight of evidence approach for assessing genotoxic/carcinogenic hazard; and (3) re-examine *in vitro* and *in vivo* correlations qualitatively and quantitatively. Conclusions from the workshop highlighted a willingness of scientists from various sectors to change and improve the current paradigm and move from a hazard identification approach to a "realistic" risk-based approach that incorporates information on mechanism of action, kinetics, and human exposure..

Keywords: Genotoxicity; *In vitro* assays; Carcinogenesis; Workshop report

1. Introduction

Human exposure to DNA-damaging agents is an important health issue because gene and chromosomal mutations can potentially lead to adverse health consequences, including cancer, reproductive impairment, developmental anomalies, or genetic diseases. Current regulatory practice is to use a battery of genetic toxicity tests to determine if a chemical has the potential to cause mutations or chromosomal damage. Tests conducted *in vitro* in bacteria and mammalian cells play an important role in this battery [1–13]. During the past 15 years, accumulated evidence has shown a high rate of positive results in the *in vitro* tests, especially in those assays performed in mammalian cells [14]. Importantly, a large number of the mammalian *in vitro* positive findings have not been confirmed in *in vivo* genotoxicity and/or carcinogenicity studies, and this raises the question of their specificity and relevance in human risk assessment [15].

Positive *in vitro* results generally lead to costly and time consuming additional testing, including mechanistic studies and *in vivo* genetic toxicity testing in rodent models. In the context of regulating pharmaceuticals, pesticides or industrial chemicals, such positive *in vitro* results could potentially lead to prohibiting the use and/or development of compounds of negligible concern for adverse human effects. It is increasingly accepted that positive results should not be considered in isolation, and that a weight of evidence approach considering all pertinent data should be the preferred approach [4,16–18]. As part of this weight of evidence approach, information on the mode of action, kinetics, and the extent of human exposure is useful for risk assessment.

The low dose portion of the dose–response curve is generally assumed to be linear with no threshold for compounds known to interact with DNA directly (*e.g.*, alkylating and intercalating agents). Other compounds can induce DNA damage as a secondary effect and act through non-DNA reactive mechanisms (*e.g.*, inhibition of topoisomerase, mitotic spindle disruption, inhibition of protein and DNA synthesis, imbalance of nucleotide pools). In this latter case, it is accepted that a threshold dose level exists below which no genotoxic effect is expected to occur. Despite recently accumulated information about possible indirect mechanisms of action [19–22], there is still a need to improve and enhance our understanding of these mechanisms and to provide clear recommendations on the approaches to be used to identify the mode of action and include these data in risk assessment.

It is agreed that *in vitro* models are imperfect models of *in vivo* biology. Nonetheless, business and regulatory decisions are often made on the basis of qualitative outcomes in these assays. In addition, extreme experimental conditions currently recommended in regulatory guidelines for the *in vitro* genotoxicity models are seen as a potential source of artifacts and irrelevant findings (*e.g.*, high level of cytotoxicity, precipitating concentrations, concentrations up to 5000 μg/ml or 10 mM that are very unlikely to be attained *in vivo*). Hence, there is a need for a better understanding of the limitations of the currently used *in vitro* models in order to more adequately interpret the *in vitro* findings, and to identify the key criteria for the development of better and more relevant predictive models for *in vivo* biology.

The need to re-consider the evaluation of *in vitro* positive findings and their impact on risk assessment has been recently highlighted by regulatory authorities [16,17,23–25], and organizations such as the International Workshops on Genotoxicity Testing (IWGT) and European Centre for the Validation of Alternative Methods (ECVAM) [18,26].

Taking all these questions/points into consideration, the Health and Environmental Sciences Institute (HESI), the global branch of the International Life Sciences Institute (ILSI) recently identified "the relevance and

follow-up of positive results in *in vitro* genetic toxicity" as an emerging issue. A HESI subcommittee was formed to address the following key objectives:

- improve the scientific basis of the interpretation of results from *in vitro* genetic toxicology tests for purposes of accurate human risk assessment,
- develop follow-up strategies for determining the relevance of *in vitro* test results to human health, and
- provide a framework for the integration of the *in vitro* testing results into a risk-based assessment of the effects of chemical exposures to human health.

In order to identify the actions to be initiated, HESI organized an international multi-sector workshop in Washington, DC, on June 21 and 22, 2006, which was attended by 45 experts in the fields of genetic toxicology, carcinogenesis and risk assessment. Participants included representatives from the United States Department of Agriculture (USDA), United States Environmental Protection Agency (USEPA), United States Food and Drug Administration (USFDA), Health Canada, Japan National Institute of Health Sciences (NIHS), and the European Food Safety Authority (EFSA), and over 15 companies from various industries involved in the development of products including industrial chemicals, agricultural chemicals, pharmaceuticals, and cosmetics. Several of the participants were also involved in other initiatives such as IWGT or ECVAM, and this workshop facilitated coordination between the different initiatives. Further information about the workshop participants and their affiliations can be found at http://www.hesiglobal.org/Committees/EmergingIssues/ToxTesting.

The program of this workshop consisted of a series of plenary lectures followed by break-out group sessions to address three main topics. These were:

- *Break-out group #1*: how to establish relevance of *in vitro* findings to humans using mechanistic and *in vivo* data.
- *Break-out group #2*: how to factor in quantitative consideration of the impact of dose–response.
- *Break-out group #3*: how to improve our testing for genetic toxicity.

This publication summarizes the three break-out group discussions and deliberations, and provides recommendations by the workshop participants to support the development of more reliable approaches to genetic toxicity risk assessment and risk management. No attempt was made at harmonizing the format of break-out group reports, but Table 1 captures the key recommendations and some of the commonalities shared between the three break-out groups. The conclusion statements will be used as a starting point for the next steps of the collaborative work. The different topics identified will be examined further in depth in the near future, in order to evaluate if new technical and scientific approaches can be used to address the identified issues and questions.

2. Summaries of the break-out group discussions

2.1. Break-out group #1: how to establish relevance of in vitro findings to humans using mechanistic and in vivo data

2.1.1. Break-out group #1: background

The positive results obtained in *in vitro* genotoxicity tests are often not confirmed in *in vivo* genotoxicity and carcinogenicity tests. To predict carcinogenesis from *in vitro* findings, there is a significant need to develop a weight of evidence approach that considers human exposure information and incorporates an understanding of the mechanism of action, metabolism and tissue distribution *in vivo*. The participants of this break-out group evaluated the relevance of *in vitro* data to humans by focusing on the concordance with *in vivo* genotoxicity and carcinogenicity data. This evaluation was done for different endpoints (*e.g.*, gene mutations, chromosome damage, primary DNA damage) with consideration of mechanistic information. The discussion also included the level of information needed to define a threshold and a potential safety margin in humans.

2.1.2. Break-out group #1: report

Genetic toxicology testing is an integral and essential part of the safety evaluation of chemicals (pharmaceuticals, pesticides, industrial chemicals, and consumer products). The primary focus of this testing is to assess the inherent potential of a substance to compromise the integrity of the genetic material. The types of heritable genetic events that are relevant to human risk assessment include gene mutations, structural chromosomal changes, and aneuploidy. Hence, all three end points should be included in human health assessments.

Currently, a battery of short-term tests is initially used to identify the genotoxic potential of a test material. There was a general consensus among the members of the break-out group that this initial battery of tests could be prescriptive. Towards this end, a battery, comprised of three or four tests (*e.g.*, a bacterial reverse mutation test, a test for chromosomal aberrations and/or mutations in mammalian cell cultures, and an *in vivo* test for

Table 1
IVGT break-out group key recommendations

Recommendation	Group #1—how to establish relevance of *in vitro* findings to humans using mechanistic and *in vivo* data	Group #2—how to factor in a quantitative consideration of the impact of dose–response	Group #3—how to improve our testing for genetic toxicity
Improving testing			
• Re-examine maximum level of cytotoxicity currently required for *in vitro* tests to determine scientific validity and evaluate appropriate measures for cytotoxicity	X		X
• Re-examine whether the current 10 mM upper limit concentration for *in vitro* mammalian studies is justified	X—conduct retrospective analysis using animal and human PK data and consider compatibility with cellular metabolic efficiency and enzyme saturation to identify an appropriate top concentration		X—evaluate the validity and applicability of current requirements which were originally based on early analysis of small databases
• Improve testing strategies using current *in vitro* assays so as to more reliably assess genotoxic hazard and predict carcinogenesis		X—understand limitations of *in vitro* tests	X—evaluate whether extended *in vitro* exposures (24 h) are contributing false positives, determine if human lymphocytes are more predictive compared to cell lines, and evaluate most appropriate metabolism system to use
• Define criteria to guide selection of the appropriate follow-up *in vivo* study(ies)	X	X	
• Develop new and more predictive *in vitro* and *in vivo* tests that could ultimately be used in addition or as a replacement of current models			X—p53 DNA repair proficient cells, metabolically relevant cells, systems with multi-endpoint analyses
Improving data interpretation and risk assessment			
• Examine the suitability of applying the threshold of toxicological concern (TTC) concepts to genotoxicity data	X—Examine the suitability of deriving benchmark doses, NOAELs, LOAELs from genotoxicity data	X—evaluate different mechanistic classes to identify thresholds or acceptable margins of exposure; evaluate available data to examine the scientific support for low-dose linearity	
• Develop a structured weight of evidence approach with robust qualitative and quantitative criteria for assessing genotoxic/carcinogenic hazard	X—consider structural alerts, assay strength/weakness, consistency and reproducibility of findings, *etc.*	X—utilize *in vivo* and *in vitro* dose–response data and human exposure information to characterize and bin levels of concern	
• Re-examine *in vitro* and *in vivo* correlations and assess the ability of current genotoxicity assays to predict carcinogenicity outcomes using thorough analysis and robust criteria	X—mine databases and conduct retrospective analyses to determine the value of *in vitro* and *in vivo* tests in predicting the outcomes of animal cancer studies	X—expand retrospective analysis to examine dose–response (*i.e.*, vivo potencies with *in vitro* concentration effects) by chemical class and type of damage	X—retrospective analysis should be based on a compiled dataset that includes ONLY current acceptance and interpretation criteria

X: Recommendation made by break-out group.

cytogenetic damage in bone marrow of rodents), as prescribed by a number of expert/regulatory bodies around the globe (*e.g.*, ICH, EEC, U.S.EPA, JMHLW, *etc.*), are considered to be still valid. For certain molecules, one or more of the above tests may not be relevant or useful (*e.g.*, bacterial reverse mutation test for peptides). Positive findings (*i.e.*, identification of genotoxicity) in one or more of the initial tests may require further investigation and usually trigger additional testing. However, the selection of additional test(s) at this stage cannot be prescriptive and should be handled on a case-by-case basis.

There was also a general consensus that the protocols and data interpretation strategies currently used in the conduct of the initial battery of tests needed improvement. For example, the rationale for the selection of the top concentration including the required levels of cytotoxicity currently prescribed by various regulatory guidelines for *in vitro* tests may need further examination. In this context, the group identified a need to undertake a retrospective analysis of any available internal exposure data from animal toxicokinetic and, where available, human pharmacokinetic studies to help establish a general guidance to limit the highest concentrations that need to be evaluated in *in vitro* genotoxicity studies utilizing mammalian cell cultures. Using information from such an analysis, one might be able to identify an appropriate top concentration to be used in *in vitro* assays (instead of the current 10 mM limit) that could generally be agreed not to be excessively above the typical pharmacologically active range for most drugs, above the K_ms for most relevant enzymes including those involved in metabolic activation/detoxification, and above the typical blood and tissue levels expected at the most extreme human exposures that would occur in actual usage situations. While this approach might be suitable in most cases, there could be instances where higher concentrations may need to be evaluated with certain agents which may have potentially extreme human exposures.

The group acknowledged that toxicokinetic and toxicodynamic considerations dictate that findings from a well conducted *in vivo* genetic toxicity test that evaluates relevant endpoints and target tissues should carry more emphasis or weight than conflicting results from corresponding *in vitro* assays. At this time, however, there are no validated *in vivo* protocols amenable to assess all genetic events of human relevance (*i.e.*, mutations, chromosomal aberrations, and aneuploidy) in multiple tissues. Because of this limitation an integrative approach of different *in vitro* genetic toxicology studies will continue to play an important role in safety assessment programs. Furthermore, although analyses performed to date suggest that the results from *in vivo* tests correlate better than *in vitro* tests in predicting the outcomes of animal carcinogenicity studies, an exhaustive analysis of all available databases has not yet been performed that would allow one to make a definitive conclusion on this issue. Therefore, this group identified such an analysis as a worthwhile future activity.

Genetic toxicology is an integral part of the field of toxicology and as such the general weight of evidence principles of data interpretation widely accepted for other toxicities should be equally applicable to genotoxicity (*i.e.*, considering all pertinent information when available including, metabolism, kinetics, mechanism, dose–response and human exposure, placing emphasis on reliable *in vivo* results over *in vitro* findings, and acknowledging data limitations). The central dogma in toxicology is that it is the dose that determines the risk of toxicity. Accordingly, an experimentally derived no-adverse-effect-level (NOAEL) or mathematically modeled "bench mark dose (BMD)" in conjunction with a set of uncertainty factors usually forms the basis to establish a human exposure level to toxicants without an expectation of an adverse outcome. The group had a cursory discussion on the applicability of such an approach to mutagenicity data, irrespective of the mode of action of an agent (*i.e.*, even for DNA reactive mutagens which are currently excluded from this approach), for setting acceptable exposure levels. The following uncertainties were discussed in the use of experimental data to derive allowable human exposure levels: (1) extrapolation from *in vitro* to *in vivo* situations, (2) extrapolation from non-human species to humans, (3) existence of susceptible subpopulations among humans, (4) severity of the effect studied, and (5) deficiency in the database used to derive the NOAEL or BMD. The group concluded that further discussion is needed on the suitability of this approach as well as the use of factors (*e.g.*, 3× or 10×) to account for each of the identified uncertainties to be used in deriving the permissible exposure levels.

2.1.3. Break-out group #1: conclusions and recommendations

Based on the deliberations described above, break-out group #1 proposed the following conclusions and recommendations:

- Critically examine the currently required maximum level of cytotoxicity in *in vitro* mammalian assays.
- Re-evaluate the current 10 mM upper limit concentration for *in vitro* mammalian studies using a retrospective analysis, taking into account the following:

- animal and human pharmacokinetic data;
- metabolic efficiency;
- enzyme saturation;
- typical blood and tissue levels at the most extreme human exposure situations.

- Apply general weight of evidence principles of data interpretation accepted for other types of toxicity to genotoxicity data, considering:
  - metabolism;
  - kinetics;
  - mechanism;
  - dose–response and human exposure;
  - placing emphasis on reliable *in vivo* results over *in vitro* findings;
  - acknowledging data limitations.
- Critically examine the suitability of applying the concepts of benchmark dose, NOAELs, LOAELs, and uncertainty factors to genotoxicity data.
- Conduct a retrospective in-depth review of the available genotoxicity databases to better understand the respective contribution of *in vitro* and *in vivo* assays to the prediction of carcinogenic potential.

2.2. *Break-out group #2: how to factor in a quantitative consideration of the impact of dose–response*

2.2.1. *Break-out group #2: background*

The participants of this break-out group focused on how to use knowledge of the *in vivo* factors that determine genotoxic responses (including exposure, pharmacokinetics, metabolism, and mechanism) to interpret responses in *in vitro* and *in vivo* laboratory genetic toxicology tests, and to improve estimation of the risk of genetic damage and/or adverse health outcomes in humans. The discussion was focused on situations in which decisions must be made in the absence of carcinogenicity data, such as (a) the stage of pharmaceutical development at which *in vitro* and limited *in vivo* genetic toxicology information, but no carcinogenicity data, are available or (b) screening of industrial chemicals, and included (c) the near future of cosmetic and health-care product regulation in which decisions may need to be made primarily on the basis of *in vitro* data or at most with some limited *in vivo* data.

2.2.2. *Break-out group #2: report*

The general question addressed by this group was: "What information on exposure and genotoxicity such as potency, nature of genetic lesion, shape of dose–response curve and mode of action is needed to define acceptable exposure levels or levels of no concern for exposed humans". The specific questions addressed included:

- Is there a quantitative relationship between potency *in vitro* and potency *in vivo* for induction of the types of damage of interest (*e.g.*, adducts, strand breaks, nucleotide alterations, mutations, chromosomal aberrations, *etc.*), for (1) agents that do not require metabolic activation and (2) agents that do require metabolic activation?
- Can a combination of *in vitro* potency data, with or without *in vivo* potency data, and human exposure data provide an index of risk that supports regulatory decision-making in the absence of carcinogenicity data?
- By using such an index, can a level of risk be defined that is considered inconsequential or acceptable for a given human exposure?

Additionally, the general default assumptions about the shape of the dose–response curves for "genotoxic" *versus* "non-genotoxic" agents were discussed, including the assumptions that (1) agents that react with or "directly" damage, DNA should be assumed to have linear dose–response relationships as a conservative default, whereas (2) DNA non-reactive agents, *i.e.*, genotoxicants that act through a primary target or mechanism other than direct reaction with DNA, are considered likely to have non-linear dose–response relationships with a definable "threshold" below which *in vivo* risk of damage can be considered negligible. In particular, the extent of the scientific data available that supports these presumptions was questioned. It was felt that some generalizations can be made with regard to type of damage, mechanism of action, and/or class of agents, but that a more rigorous evaluation of the situations and conditions involved was needed. For certain classes of chemicals, such as specific DNA synthesis inhibitors or agents that interact very specifically with known non-DNA targets, it was agreed that a threshold below which significant DNA damage would not occur could be defined. Systematic approaches to evaluate available data that supports improved categorization of chemical classes and supports appropriate assumptions about expected dose–response relationships are recommended (see below).

Some data were presented suggesting that there are practical thresholds even for DNA-reactive genotoxic agents [27,28] and showing also that agents considered non-genotoxic, such as sucrose, can produce significant effects even *in vivo* if sufficient exposure is achieved [29]. The limited data presented suggests that a more comprehensive survey and analysis of results in the liter-

ature and those available from HESI member companies is warranted in order to determine whether practical thresholds can be defined for DNA-reactive genotoxic agents. Based on this analysis, an informed decision can be made as to whether it is necessary to move from the current practice of decision-making on the basis of qualitative or semi-quantitative characterization of agents to a more quantitative assessment of genotoxic risk under defined exposure conditions *in vivo*. As noted above, the group felt that sufficient data were already available to document that some classes of non-DNA reactive genotoxicants, such as most aneugens (based on known modes of action/dose–response data), agents that cause nucleotide pool imbalance or glutathione depletion, and DNA synthesis inhibitors, have a non-linear dose–response curve and that safe thresholds or margins of exposure can be defined for such agents. The group recommended a systematic compilation and analysis of data for both DNA reactive and non-reactive mutagens that explores the dose–response and modes of action to more thoroughly examine the default presumption of low dose linearity. A logical mechanism for achieving this would be via an expert committee charged with producing a "white paper" and subsequent journal publication. It was noted that it would likely also be necessary to build a consensus on acceptable methods for describing the shape of the dose–response curve and for evaluating the mode of action.

Considerable attention was placed on the question of whether information on the extent of human exposure (magnitude, duration, and route) can be used to define levels of concern about genotoxic damage. The "level of concern" and "threshold of toxicological concern" concepts (LOC, TTC) used for assessing environmental risks, direct and indirect food additives and pharmaceutical impurities [2,30–33] were cited as examples of how this is already being done in certain cases. The possibility of extending these concepts by combining human exposure information with information about *in vitro* dose–response relationships and/or *in vivo* animal genotoxicity information was discussed and it was concluded that this could be a profitable area of focus within the current HESI project.

The group recommended that *in vivo* potency information (both genotoxicity and carcinogenicity information), information about *in vitro* concentration–effect relationships in relation to effects in *in vivo* models, and likely concentrations achieved in anticipated human exposure situations should be evaluated as a basis for future recommendations. It was recommended that this analysis include a determination of whether available information allows development of semi-quantitative categories (bins) of concern (*e.g.*, low, intermediate, high) for some classes of chemicals based on:

- human exposure data;
- *in vivo* potency test data (*e.g.*, tumor data, genetic toxicity data);
- *in vitro* concentration far exceeding achievable *in vivo* exposure (*e.g.*, blood/tissue concentrations, DNA adducts);
- mode of action and metabolism/pharmacokinetics.

It was suggested that a weight of evidence approach that considers structural alerts, weaknesses and strengths of each assay, and consistency and reproducibility of the findings would be needed. Along with human exposure information, the biological plausibility for the response to occur in humans should be considered. Because positive findings in *in vitro* assays, particularly in mammalian cell systems, can be problematic, correlative *in vivo* data are preferred to evaluate the potential for human risk. However, the limitations of *in vivo* assays (*e.g.*, ability to measure relevant events in potential target cell populations, sensitivity of certain endpoints when exposure is short-term, metabolic and pharmacokinetic differences among species) need to be considered in developing an appropriate weight of evidence approach. The above recommendations for activities to be undertaken are directed at providing evidence-based approaches to these issues.

The group noted that any effort undertaken should be coordinated with the existing effort of the International Workshops on Genotoxicity Testing (IWGT) directed at improved genetic toxicology testing strategies, and especially the IWGT working group on appropriate follow-up testing when *in vitro* positive genotoxic responses occur [18].

2.2.3. Break-out group #2: conclusions and recommendations

Based on the discussions described above, break-out group #2 proposed the following conclusions and recommendations:

- Although *in vitro* assays are useful, recent analyses of expanded datasets have illustrated the limitations of these tests. Improved approaches are needed that allow the results of these *in vitro* assays to be better used in assessing genotoxic hazard.
- An evaluation of *in vivo* and *in vitro* genetic toxicology data including dose–response, by chemical class and type of damage, is needed to determine the feasibil-

ity of developing a tiered or quantitative classification system for genotoxic hazard.
  - The evaluation should include examination of the relationship between *in vitro* and *in vivo* responses, for different mechanistic classes of genotoxicants, analyzed separately by whether the agent is directly active or requires metabolic activation for genotoxic activity.
  - The evaluation should include correlation of tissue exposure *in vivo* with genetic damage *in vivo* (including tumor response) and *in vitro*, to support development of (semi) quantitative estimates of levels of concern.
  - An analysis is recommended to determine if we can develop different bins of concern (*e.g.*, low, intermediate, high) for some classes of chemicals based on human exposure data; *in vivo* potency test data (*e.g.*, tumor data, genetic toxicity data); and *in vitro* concentration in relation to achievable *in vivo* exposure (*e.g.*, blood/tissue concentrations, DNA adducts).
- An evaluation of the literature and available data bases (pesticides, drugs, NTP, *etc.*) is needed to determine the scientific support for low dose linearity *versus* practical thresholds for different classes of genetic toxicants.
- Whenever possible, *in vivo* dose–response and human exposure information should be used in a weight of the evidence approach to evaluate the potential for human risk.
  - Because concerns were raised over limitations of currently available *in vivo* methods, a review should be undertaken of available information to define these limitations so that the combination of *in vitro* and *in vivo* information can be used more effectively.

2.3. Break-out group #3: how to improve our testing for genetic toxicity

2.3.1. Break-out group #3: background

The participants in this break-out group started with the premise that we can not throw out the 'tried and tested' approaches without having something with which to replace them. It was recognized that the 'Ames test' has a very robust database and would be difficult to throw out, and that the most problematic tests are currently the *in vitro* chromosome damage tests, as they demonstrate the higher rates of positives. Additionally, it has been suggested that the Ames results generally correspond with structure-activity models based on electrophilicity [34]. This break-out group focused their attention on the need to develop *in vitro* models that are more predictive models for *in vivo* biology, and that reduce artifacts.

2.3.2. Break-out group #3: report

All *in vitro* systems are at best imperfect models for the biological effects seen *in vivo*. This generalization holds true for *in vitro* genotoxicity tests used as hazard identification tools in the prediction of carcinogenicity, especially in view of our current understanding that epigenetic events play a key role in carcinogenicity. One of the challenges in using *in vitro* genotoxicity assays as predictors of carcinogens was highlighted in a recent analysis by Kirkland et al. [15,35] of over 700 chemicals that have rodent carcinogenicity data, which found that 75–95% of non-carcinogens were positive in one or more of the standard *in vitro* genotoxicity assays. In this analysis, the false positive rate (defined as positive in mutagenicity assay but negative in a rodent cancer bioassay) was highest in mammalian cell tests such as the chromosomal aberration assay in Chinese hamster cells or the tk gene mutation assay in L5178Y mouse lymphoma cells. As a consequence of such positive *in vitro* genotoxicity data, numerous animal studies and mechanistic research projects are conducted in order to determine whether effects seen *in vitro* are biologically relevant *in vivo*. These studies are costly, time consuming, utilize many animals, and do not always give a definitive answer.

The findings reported by Kirkland et al. [15,35] and in several other earlier analyses [36–38] have recently been confirmed in an analysis by Matthews et al. [39,40] of a larger database of FDA and EPA chemicals. These recent analyses confirm earlier analyses on smaller data sets in the late 1980s to early 1990s. It is recognized that efforts to correlate the genetic toxicology assays with the cancer bioassay data are complicated by the fact that not all the genetic toxicology and cancer data have been evaluated according to current standards of acceptability and interpretation. Despite this, there was consensus that there is great value in developing new tests and/or approaches for predicting *in vivo* genotoxins and potential carcinogenic chemicals. This topic was the focus of the break-out group.

In vitro genotoxicity assays are used for a variety of purposes, from the rapid screening of potential drugs or other chemicals of interest to the detailed mode of action analyses for carcinogenicity risk assessment [16,23]. The participants focused their discussions on the use of genetic toxicity tests for predicting whether a chemical has the potential to cause carcinogenicity via a mutagenic mechanism, *i.e.*, hazard identification. This use of genetic toxicity tests to determine whether the mode of

action (MOA) of a known carcinogen is via a mutagenic mechanism was not addressed by this workgroup.

Over the years, it has been become apparent that the *in vitro* genotoxicity tests, particularly the mammalian cell assays, detect some non-DNA reactive agents (*i.e.*, the primary target of the chemical or its metabolite(s) is not DNA, for example, topoisomerase inhibitors) in addition to DNA-reactive agents. In fact, there has been an effort over the years to expand the spectrum of genetic events detected in each assay, for instance by extending the length and types of chemical treatment. Furthermore, there has been pressure to increase the numbers and types of assays in various genotoxicity testing batteries in order to detect the full spectrum of genetic events and/or as many rodent carcinogens as possible. The majority of the working group felt that this proliferation of testing was contributing to the generation of "false positive results" with respect to predicting whether a chemical will be a carcinogen. As such, we discussed the necessity of refocusing genetic toxicology tests on the detection of DNA reactive carcinogens. Other members expressed the view that genotoxicity assays detect genetic damage and would be expected to respond to all insults that damage DNA (regardless as to whether the damage is caused "directly" or "indirectly"). Both short-term and long-term solutions to address these issues were discussed.

2.3.2.1. Possible short-term solutions. There is a need to identify potential sources of false positive results (for predicting carcinogenicity) obtained with the current *in vitro* genotoxicity assays. This issue relates to determining what assay conditions cause biologically irrelevant positive responses—artifacts of the *in vitro* conditions. This issue was also viewed as essential for developing any new tests or longer term approaches. To this end, we addressed the question "What kind of research or efforts can be used to improve current *in vitro* tests?" The following possible activities were discussed:

- Re-examine whether a top concentration of 10 mM is justified. The original guidance to use a top concentration of 10 mM in the mammalian cell assays when there is no toxicity is based on early analyses of small databases which showed that there was a need to test up to10 mM to detect some mutagens. Because such mutagens may be detected in the bacterial gene mutation (Ames) assay, there was agreement that there may be inadequate justification for routine use of 10 mM in mammalian cell assays. Another factor in setting a top concentration of 10 mM was to avoid osmolality effects in these assays (*i.e.*, effects due to osmotic conditions that cannot be achieved *in vivo*). Changes in osmolality are controlled for in assays conducted by today's standards.
- Re-examine the maximum level of cytotoxicity needed and the appropriate measures of cytotoxicity. By virtue of being *in vitro* tests, high, non-physiological concentrations of test chemicals can be added to *in vitro* genotoxicity assays. Similarly to the above, this group felt there was a need to determine whether detection of *in vivo* mutagens and/or DNA reactive carcinogens required routine testing up to the cytotoxic levels used in current protocols.
- Determine whether both long exposures as well as short exposures in the mammalian cell assays are required to detect *in vivo* mutagens and/or DNA reactive carcinogens, particularly those not detected by the bacterial reverse mutation assay.
- Determine if induced rat liver S9 is the most appropriate metabolic activation for *in vivo* mutagens and/or DNA reactive carcinogens. Investigate other metabolic activation systems.
- Determine if cytogenetic assays in human lymphocytes are better predictors of human hazard and more relevant to human risk assessment than currently used mammalian cell lines. Anecdotal data as well as recent publications [41] have been discussed at various meetings to suggest fewer "irrelevant" positive results occur in the chromosome aberration assay and/or micronucleus assay when conducted in human lymphocytes than in other mammalian cell lines. There is a need to determine whether this possibility can be confirmed since this could lead to a simple solution to the problem of false positives for predicting carcinogenicity. It was, however, unclear whether there was sufficient data available with human cells for this analysis.
- After the meeting there was a proposal to conduct a thorough analysis of the existing genotoxicity and cancer databases to create a dataset that includes only data that meets current acceptance and interpretation criteria. Once such an analysis has been completed, the information can be used to more accurately access the ability of the current genotoxicity assays to predict whether a chemical will be a carcinogen. This effort would also provide a sound foundation for addressing and perhaps modifying some of the assay parameters (top dose, required cytotoxicity level, *etc.*).

2.3.2.2. Approaches for the possible short-term solutions. To initiate these activities, the following approaches were discussed:

- Form an expert panel to identify a list of definitive *in vivo* genotoxins and/or DNA reactive mutagenic carcinogens which we expect *in vitro* genotoxicity tests to detect, and then search these chemicals to answer the above questions.
- Form an expert panel to analyze the role of metabolism in the mutagenicity and/or carcinogenicity of *in vivo* genotoxins and/or of DNA reactive mutagenic carcinogens.
- Initiate a collaborative experimental study to analyze different measures of cytotoxicity to determine if appropriate measures are being used.
- Search existing databases to determine whether fewer false positive results occur in human lymphocyte cytogenetics assays than in other mammalian cell lines.
- Collect HESI member company data to determine whether there are fewer "false positive" results in human lymphocyte cytogenetics assays. This data collection exercise should address the following points:
  - Include data to address whether there is increased variability of human lymphocytes relative to other commonly used cell types (data collection from contract and testing labs that use HPBL).
  - Focus on chemicals negative for bacterial gene mutation.
  - Include data comparing rat lymphocytes to other commonly employed cell lines to address the possibility that primary lymphocytes yield more relevant results.
- Initiate a collaborative experimental effort to compare cytogenetic results between different cell types. It was recognized that this effort will take the largest amount of resources, but would be the most definitive way to address the question since analyses of databases are complicated by the quality of the studies that were not designed for this purpose.

2.3.2.3. Possible mid-term solutions. Based on the analyses of current databases, the bacterial reverse mutation assay has been shown to have the highest specificity for prediction of rodent carcinogenicity of the currently used *in vitro* genotoxicity assays [15,35]. Based on this, the group discussed whether the *in vitro* mammalian cell assays could be replaced by tests or approaches that compliment the bacterial reverse mutation assay, *i.e.*, that detect *in vivo* genotoxins and/or DNA reactive mutagenic carcinogens that are negative in the bacterial assay [38]. Because changing the standard genotoxicity testing battery would require changes in regulations, this was viewed as a possible mid-term solution.

As a first step, we discussed the question "Can we accept a battery of the bacterial reverse mutation assay and an *in vivo* MN or another assay for routine testing acknowledging that some chemical classes may require alternative testing?" Types of chemicals that are potential hazards that are known to be negative in the bacterial assay include: metals, steroids/hormones, topoisomerase inhibitors, nucleoside analogs, mammalian receptor-specific chemicals and chemicals whose primary activity is the induction of large deletions and other chromosomal damage. It is the detection of this latter class of chemicals (chromosomal mutagens) that led to the establishment of the current battery.

2.3.2.4. Approaches for possible mid-term solutions. One suggested approach to address this question is to conduct a database analysis of bacterial reverse mutation and *in vivo* MN tests (or other assays) to see if these detect relevant *in vivo* genotoxins or/and DNA reactive mutagenic carcinogens. Classes not detected by this battery could be identified and appropriate testing recommendations determined. This would involve the following steps: search existing databases like the database used by Kirkland et al. [15,35]; the database used by Matthews et al. ([40,41], EPA GENE-Tox [42], and others; and obtain HESI member Company data. This is best accomplished using databases that have been thoroughly evaluated to include only data that meets all current criteria for acceptability and interpretation.

2.3.2.5. Possible long-term solutions. While the above approaches have the potential to reduce problems with current *in vitro* genotoxicity tests and their interpretation in the near term, which was the primary focus of this workshop, there was also some discussion about the need for the development of new generation tests that could be used in the future—tests that could be specifically designed to address the features that the current tests lack. Features of new tests that were discussed included use of mammalian cells/cell lines to insure appropriate mammalian cell targets are present, use of p53 and DNA repair proficient cells that are metabolically relevant, and development of assays that would allow multi-endpoints analyses.

There was also discussion of the use of only *in vivo* genotoxicity tests in the future and/or of the need for the development of a new generation of *in vivo* tests that measure the full spectrum of mutagenic events. For optimal utility, such new systems should provide for rapid mutant detection and not require the *in vitro* growth of cells to enumerate mutants. While deemed scientifically appropriate, it was noted that the use of *in vivo* tests alone would not, under some current regulatory guidelines, be acceptable for some testing purposes, including indus-

trial chemicals, cosmetics, *etc.*, but their use could be valuable for some applications.

2.3.2.6. Approach for the possible long-term solutions. We discussed the utility of holding a workshop to discuss new generation *in vitro* and *in vivo* tests. In a workshop recently sponsored by ECVAM, some of the newer *in vitro* assays were discussed [26]. A workshop that discussed both new *in vitro* and new *in vivo* tests would be valuable.

2.3.3. Break-out group #3: conclusions and recommendations

This workgroup approached its discussion with the goal of capturing a wide variety of opinions and generating a number of options for improving the identification of chemicals that are carcinogens prior to the completion of any cancer bioassays. While there were diverse opinions concerning the utility of the current tests and approaches, there was general agreement that new tests and approaches are needed. The workgroup also agreed that, to make significant progress on this issue in a reasonable length of time, a variety of parallel activities would be required. As such, we encourage partnering of the various interested stakeholders in these initiatives.

3. Overall workshop conclusions

Table 1 summarizes the key recommendations of the workshop, and identifies some of the commonalities shared between the three break-out groups.

There was general agreement among workshop participants that the rate of *in vitro* positive findings not confirmed *in vivo* is too high to justify using qualitative outcomes as the sole basis of regulatory decision-making and that there is a critical need for an improved evaluation process and for better predictive models. The active participation of the workshop attendees in the discussions during the break-out groups highlighted a willingness to change and improve the current paradigm and to move from a hazard identification approach to a risk based approach that considers both toxicity and human exposure information. A general consensus was reached that the following points should be considered in the near future:

- Genotoxicity data should be considered along with other pertinent information, including extent of human exposure and dose–response relationships, in line with other toxicology end points. A weight of evidence approach should be widely applied that considers genotoxic exposure (*e.g.*, reproducibility, presence of cytotoxicity, corroborative data between studies evaluating the same end point), the relative potencies of these responses (by chemical class and type of damage), as well as the route, magnitude and duration of human exposure. When available, the weight of evidence approach should also integrate information on mode of action (*e.g.*, presence of DNA adducts/strand-breaks), metabolism and tissue concentrations *in vivo*, and tumor-related response such as relevant non-neoplastic and preneoplastic lesions. Moreover, whether the genotoxicity observed with a given chemical is a key event in the multistep process of carcinogenesis and the role of other key events (*e.g.*, regenerative proliferation, mitogenic stimulation of preneoplastic foci) should be further evaluated in case of tumor findings.
- Protocols need to be improved to reduce and possibly avoid the generation of artifacts and the unnecessary and extensive use of animal studies and resources, to minimize extreme high dosing conditions that would never be achieved *in vivo*, and to incorporate dosing conditions that are more realistic to human exposure situations to enable better extrapolation of the results. A collaborative effort was suggested to compare the results obtained with different cell types (*e.g.*, primary human lymphocytes *versus* cell lines), to evaluate the limits of different cytotoxicity measurements *in vitro*, to re-consider the rationale for the selection of the top concentration levels (*e.g.*, level of cytotoxicity, precipitates, and 10 mM limit), to review data obtained after short- and long-term exposures, to reconsider the metabolic activation in the *in vitro* systems, and in the case of *in vivo* tests to develop the possibility of evaluating multiple genotoxic end points from the same treated animals. This could be accomplished by examining existing databases, private and public, and by determining if certain assays could be eliminated or substituted. It is likely that some experimental work would be needed to obtain the information needed.
- The appropriateness of non-linear low dose–response extrapolations for both DNA reactive and non-reactive carcinogens should be further evaluated. A white paper should be prepared to examine the scientific validity or lack thereof of the low dose linear extrapolation for genotoxicity/carcinogenicity. Moreover, guidance should be given to clarify the acceptable approaches to define dose–response relationships, and to establish the existence of non-linearity. The development of uncertainty factors for establishing the "thresholds" or (semi) quantitative estimates of levels of concern was suggested.

The workshop participants stressed the importance of developing a risk-based paradigm for evaluating genotoxicity data that incorporates dose–response and human exposure information. Specific needs were identified in two general categories, *i.e.*, improving testing, and improving data interpretation and risk assessment. Recommendations to improve testing included (1) re-examine and evaluate the maximum level of cytotoxicity currently required for *in vitro* tests; (2) re-examine the current 10 mM upper limit concentration for *in vitro* mammalian studies; (3) develop improved testing strategies using current *in vitro* assays to more reliably assess genotoxic hazard and predict carcinogenesis; (4) define criteria to guide selection of the appropriate follow-up *in vivo* studies; (5) develop new and more predictive *in vitro* and *in vivo* tests, that could ultimately be used in addition or in replacement of the current models. Recommendations for improving data interpretation and risk assessment included: (1) examine the suitability of integrating threshold concepts in the assessment of genotoxicity data; (2) develop a structured weight of evidence approach for assessing genotoxic/carcinogenic hazard; and (3) re-examine *in vitro* and *in vivo* correlations. Additionally, the participants identified the critical need for support and coordination of an international collaborative effort to address these issues. The HESI subcommittee will facilitate this coordination, address the recommendations of this workshop, and identify specific research projects that will facilitate the development of a framework for the integration of *in vitro* testing results into a risk-based assessment of the effects of chemical exposure on human health.

References

[1] COM (Committee on Mutagenicity of Chemicals in Food, Consumer Products, and the Environment), COM Guidance on a Strategy for Testing of Chemicals for Mutagenicity, United Kingdom, December 2000 (http://www.advisorybodies.doh.gov.uk/com/).

[2] K.L. Dearfield, M.C. Cimino, N.E. McCarroll, I. Mauer, L.R. Valcovic, Genotoxicity risk assessment: a proposed classification strategy, Mutat. Res. 521 (2002) 121–135.

[3] U.S. FDA (Food and Drug Administration), Toxicological Principles for the Safety of Food Ingredients. IV.C.1. Short-Term Tests for Genetic Toxicity, Redbook, 2000.

[4] U.S. FDA (Food and Drug Administration), Guidance for Industry Recommended Approaches to Integration of Genetic Toxicology Study Results. Draft Guidance. Genotoxicity, Center for Drug Evaluation and Research, November 2004.

[5] Health Protection Board (Canada), The Assessment of Mutagenicity. Health Protection Branch Mutagenicity Guidelines, Environ. Mol. Mutagen. 21 (1992) 15–37.

[6] ICH (International Cooperation on Harmonization), ICH Harmonized Tripartite Guideline. S2A. Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals, July 1995 (http://www.ich.org).

[7] ICH (International Cooperation on Harmonization), ICH Harmonized Tripartite Guideline. S2B. Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals, July 1997 (http://www.ich.org).

[8] Council Directive Concerning the Placing of Plant Protection Products on the Market (91/414/CEE), 15 July 2001.

[9] P.G.N. Kramers, A.G.A.C. Knaap, C.A. van der Heijden, R.D.F.M. Taalman, G.R. Mohn, Role of genotoxicity assays in the regulation of chemicals in The Netherlands: considerations and experiences, Mutagenesis 6 (1991) 487–493.

[10] OECD (Organisation for Economic Co-operation and Development), OECD Guidelines for Testing of Chemicals. Introduction to the OECD Guidelines on Genetic Toxicology Testing and Guidance on the Selection and Application of Assays, OECD, Paris, France, 1986.

[11] T. Sofuni, Japanese guidelines for mutagenicity testing, Environ. Mol. Mutagen. 21 (1993) 2–7.

[12] TGD, Technical Guidance Document, second ed., European Chemicals Bureau, Joint Research Centre, Italy, December 2003 (http://ecb.jrc.it/new-chemicals/documents).

[13] M.C. Cimino, Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes, Environ. Mol. Mutagen. 47 (2006) 362–390.

[14] R.D. Snyder, J.W. Green, A review of the genotoxicity of marketed pharmaceuticals, Mutat. Res. 488 (2001) 151–169.

[15] D. Kirkland, M. Aardema, L. Henderson, L. Muller, Evaluation of the ability of a battery of three genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. 1. Sensitivity, specificity and relative predictivity, Mutat. Res. 584 (2005) 1–256.

[16] U.S. Environmental Protection Agency (EPA), Guidelines for Carcinogen Risk Assessment, EPA/630/P-03/001F, March 2005 (http://www.epa.gov/cancerguidelines).

[17] K.L. Dearfield, M.M. Moore, Use of genetic toxicology information for risk assessment, Environ. Mol. Mutagen. 46 (2005) 236–245.

[18] V. Thybaud, M. Aardema, J. Clements, K. Dearfield, S. Galloway, M. Hayashi, D. Jacobson-Kram, D. Kirkland, J.T. MacGregor, D. Marzin, W. Ohyama, M. Schuler, H. Suzuki, E. Zeiger, Strategy for genotoxicity testing: Hazard identification and risk assessment in relation to *in vitro* testing, Mutat. Res. 627 (2007) 41–58.

[19] D.J. Kirkland, L. Müller, Interpretation of the biological relevance of genotoxicity test results: the importance of thresholds, Mutat. Res. 464 (2000) 137–147.

[20] L. Müller, P. Kasper, Human biological relevance and the use of threshold-arguments in regulatory genotoxicity assessment: experience with pharmaceuticals, Mutat. Res. 464 (2000) 19–34.

[21] S. Madle, W. Von der Hude, L. Broschinski, G.R. Jänig, Threshold effects in genetic toxicology: perspective of chemicals regulation in Germany, Mutat. Res. 464 (2000) 117–121.

[22] M. Kirsch-Volders, A. Vanhauwaert, U. Eichenlaub-Ritter, I. Decordier, Indirect mechanism of genotoxicity, Toxicol. Lett. 140–141 (2003) 63–74.

[23] U.S. Environmental Protection Agency (EPA), Supplemental Guidance for Assessing Cancer Susceptibility from Early-life Exposure to Carcinogens, EPA/630/R-03/003F, March 2005 (http://www.epa.gov/cancerguidelines).

[24] D. Jacobson-Kram, A. Jacobs, Use of genotoxicity data to support clinical trial or positive genetox findings on a candidate pharmaceutical or impurity ... Now what? Int. J. Toxicol. 24 (2005) 129–134.

[25] FDA (Food and Drug Administration) CDER (Center for Drug Evaluation and Research), Guidance for Industry and Review Staff: Recommended Approaches to Integration of Genetic Toxicology Study Results, January 2006.
[26] D. Kirkland, S. Pfuhler, D. Tweats, M. Aardema, R. Corvi, F. Darroudi, A. Elhajouji, H. Glatt, P. Hastwell, M. Hayashi, P. Kasper, S. Kirchner, A. Lynch, D. Marzin, D. Maurici, J.-R. Meunier, L. Müller, G. Nohynek, J. Parry, E. Parry, V. Thybaud, R. Tice, J. van Benthem, P. Vanparys, P. White, How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop, Mutat. Res. 628 (2007) 31–55.
[27] N. Asano, D.K. Torous, C.R. Tometsko, S.D. Dertinger, T. Morita, M. Hayashi, Practical threshold for micronucleated reticulocyte induction observed for low doses of mitomycin C, Ara-C and colchicines, Mutagenesis 21 (2006) 15–20.
[28] T. Sofuni, T. Nohmi, T. Ohta, M. Hayashi, Genotoxicity: Is a threshold concept applicable to evaluate the mutagenic activity of DNA-targeting substances? Environ. Mutagen Res. 27 (2005) 61–73.
[29] L.O. Dragsted, B. Daneshvar, U. Vogel, H.N. Autrup, H. Wallin, L. Risom, P. Møller, A.M. Mølck, M. Hansen, H.F. Poulsen, S. Loft, A sucrose-rich diet induces mutations in the rat colon, Cancer Res. 62 (2002) 4339–4345.
[30] R. Kroes, A.G. Renwick, M. Cheeseman, J. Kleiner, I. Mangelsdorf, A. Piersma, B. Schilter, F. van Schothorst, J.G. Vos, G. Wurtzen, Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet, Food Chem. Toxicol. 42 (2004) 65–83.
[31] M.A. Cheeseman, Thresholds as a unifying theme in regulatory toxicology, Food Addit. Contam. 22 (2005) 900–906.
[32] L. Müller, R.J. Mauthe, C.M. Riley, M.M. Andino, D. De Antonis, C. Beels, J. DeGeorge, A.G.M. De Knaep, D. Ellison, J.A. Fagerland, R. Frank, B. Fritschel, S. Galloway, E. Harpur, C.D.N. Humfrey, A.S. Jacks, N. Jagota, J. Mackinnon, G. Mohan, D.K. Ness, M.R. O'Donovan, M.D. Smith, G. Vudathala, L. Yotti, A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity, Regul. Toxicol. Pharmacol. 44 (2006) 198–211.
[33] U.S. FDA (Food and Drug Administration). Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food, Chapter III. Concern Levels and Recommended Toxicity Tests, Center for Food Safety and Applied Nutrition, Office of Food Additive Safety, 1993.
[34] D. McGregor, Comments on the proposed use of Salmonella and structural alerts for the predictions of rodent carcinogens, Mutat. Res. 222 (1989) 299–306.
[35] D. Kirkland, M. Aardema, L. Müller, M. Hayashi, Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: II. Further analysis of mammalian cell results, relative predictivity and tumour profiles, Mutat. Res. 608 (2006) 29–42.
[36] J. Ashby, R.W. Tennant, E. Zeiger, S. Stasiewicz, Classification according to chemical structure, mutagenicity to Salmonella and level of carcinogenicity of a further 42 chemicals tested for carcinogenicity by the U.S. National Toxicology Program, Mutat. Res. 223 (1989) 73–103.
[37] J. McCann, B. Ames, Detection of carcinogens as mutagens in the Salmonella/microsome test: assay of 300 chemicals: discussion, Proc. Natl. Acad. Sci. U.S.A. 73 (1976) 950–954.
[38] R.W. Tennant, B.H. Margolin, M.D. Shelby, E. Zeiger, J.K. Haseman, J. Spalding, W. Caspary, M. Resnick, S. Stasiewicz, B. Anderson, et al., Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicigy assays, Science 236 (1987) 933–941.
[39] E.J. Matthews, N.L. Kruhlak, M.C. Cimino, R.D. Benz, J.F. Contrera, An analysis of genotoxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints, Regul. Toxicol. Pharmacol. 44 (2006) 83–96.
[40] E.J. Matthews, N.L. Kruhlak, M.C. Cimino, R.D. Benz, J.F. Contrera, An analysis of genotoxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using *in silico* methods, Regul. Toxicol. Pharmacol. 44 (2006) 97–110.
[41] C. Hilliard, R. Hill, M. Armstrong, C. Fleckenstein, J. Crowley, E. Freeland, D. Duffy, S.M. Galloway, Chromosome aberrations in Chinese hamster and human cells: a comparison using compounds with various genotoxicity profiles, Mutat. Res. 616 (2007) 103–118.
[42] NLM (National Library of Medicine) National Institutes of Health, National Library of Medicine. U.S. Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances—GENE-TOX Agent Registry. Online computer database—TOXNET file. US Dept. Health and Human Services, Bethesda, MD, 2003 (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX).

Available online at www.sciencedirect.com
ScienceDirect
Mutation Research 627 (2007) 41–58
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Strategy for genotoxicity testing: Hazard identification and risk assessment in relation to *in vitro* testing

V. Thybaud [a,*], M. Aardema [b], J. Clements [c], K. Dearfield [d], S. Galloway [e], M. Hayashi [f], D. Jacobson-Kram [g], D. Kirkland [c], J.T. MacGregor [h], D. Marzin [i], W. Ohyama [j], M. Schuler [k], H. Suzuki [l], E. Zeiger [m]

[a] *Drug Safety Evaluation, Sanofi-aventis, 94400 Vitry sur Seine, France*
[b] *The Procter & Gamble Co., P.O. Box 538707, Cincinnati Ohio 45253, USA*
[c] *Covance Laboratories Ltd., Otley Road, Harrogate HG3 1PY, UK*
[d] *Food Safety and Inspection Service, Office of Public Health Science, U.S. Department of Agriculture, Washington, DC 20250, USA*
[e] *Merck Research Laboratories, West Point, PA 19486, USA*
[f] *Division of Genetics and Mutagenesis, National Institute of Health Sciences, Tokyo, Japan*
[g] *Office of New Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 20993, USA*
[h] *Toxicology Consulting Services, Arnold, MD 21012, USA*
[i] *Institut Pasteur de Lille, BP 245, 59019 Lille Cedex, France*
[j] *Yakult Central Institute for Microbiological Research, 1796 Yaho, Kunitach-shi, Tokyo 186-8650, Japan*
[k] *Pfizer Global Research & Development, Groton, CT 06340, USA*
[l] *Ina Research Inc., 2148-188 Nishiminowa, Ina-shi, Nagano 399-4501, Japan*
[m] *Errol Zeiger Consulting, Chapel Hill, NC 27514, USA*

Received 28 February 2006; received in revised form 31 July 2006; accepted 14 August 2006
Available online 27 November 2006

Abstract

This report summarizes the proceedings of the September 9–10, 2005 meeting of the Expert Working Group on Hazard Identification and Risk Assessment in Relation to *In Vitro* Testing, part of an initiative on genetic toxicology. The objective of the Working Group was to develop recommendations for interpretation of results from tests commonly included in regulatory genetic toxicology test batteries, and to propose an appropriate strategy for follow-up testing when positive *in vitro* results were obtained in these assays. The Group noted the high frequency of positive *in vitro* findings in the genotoxicity test batteries with agents found not to be carcinogenic and thought not to pose a carcinogenic health hazard to humans. The Group agreed that a set of consensus principles for appropriate interpretation and follow-up testing when initial *in vitro* tests are positive was needed. Current differences in emphasis and policy among different regulatory agencies were recognized as a basis of this need. Using a consensus process among a balanced group of recognized international authorities from industry, government, and academia, it was agreed that a strategy based on these principles should include guidance on: (1) interpretation of initial results in the "core" test battery; (2) criteria for determining when follow-up testing is needed; (3) criteria for selecting appropriate follow-up tests; (4) definition of when the evidence is sufficient to define the mode of action and the relevance to human exposure; and (5) definition of approaches to evaluate the degree of health risk under conditions of exposure of the species of concern (generally the human).

A framework for addressing these issues was discussed, and a general "decision tree" was developed that included criteria for assessing the need for further testing, selecting appropriate follow-up tests, and determining a sufficient weight of evidence to attribute a level of risk and stop testing. The discussion included case studies based on actual test results that illustrated common

* Corresponding author. Tel.: +33 1 58 93 35 38; fax: +33 1 58 93 38 44.
E-mail address: Veronique.Thybaud@sanofi-aventis.com (V. Thybaud).

doi:10.1016/j.mrgentox.2006.10.003

situations encountered, and consensus opinions were developed based on group analysis of these cases. The Working Group defined circumstances in which the pattern and magnitude of positive results was such that there was very low or no concern (*e.g.*, non-reproducible or marginal responses), and no further testing would be needed. This included a discussion of the importance of the use of historical control data. The criteria for determining when follow-up testing is needed included factors, such as evidence of reproducibility, level of cytotoxicity at which an increased DNA damage or mutation frequency is observed, relationship of results to the historical control range of values, and total weight of evidence across assays. When the initial battery is negative, further testing might be required based on information from the published literature, structure activity considerations, or the potential for significant human metabolites not generated in the test systems. Additional testing might also be needed retrospectively when increase in tumors or evidence of pre-neoplastic change is seen.

When follow-up testing is needed, it should be based on knowledge about the mode of action, based on reports in the literature or learned from the nature of the responses observed in the initial tests. The initial findings, and available information about the biochemical and pharmacological nature of the agent, are generally sufficient to conclude that the responses observed are consistent with certain molecular mechanisms and inconsistent with others. Follow-up tests should be sensitive to the types of genetic damage known to be capable of inducing the response observed initially. It was recognized that genotoxic events might arise from processes other than direct reactivity with DNA, that these mechanisms may have a non-linear, or threshold, dose-response relationship, and that in such cases it may be possible to determine an exposure level below which there is negligible concern about an effect due to human exposures. When a test result is clearly positive, consideration of relevance to human health includes whether other assays for the same endpoint support the results observed, whether the mode or mechanism of action is relevant to the human, and – most importantly – whether the effect observed is likely to occur *in vivo* at concentrations expected as a result of human exposure. Although general principles were agreed upon, time did not permit the development of recommendations for the selection of specific tests beyond those commonly employed in initial test batteries.

Keywords: Genotoxicity; Hazard identification; Risk assessment; Testing strategy

1. Introduction

Genetic toxicity testing is routinely performed to identify potential genotoxic carcinogens and germ cell mutagens. With regard to the identification of genotoxic carcinogens, all the minimal batteries of genetic toxicology tests recommended by regulatory agencies include at least two or three test procedures, generally an Ames test, a mammalian cell chromosome damage test, and in some cases a mammalian cell mutation assay [1–15]. Depending on the responses in the tests, the types of substances tested, and on their intended uses (*e.g.*, pharmaceuticals, pesticides, chemicals, cosmetics, *etc.*), one or more *in vivo* rodent tests (*e.g.*, bone marrow micronucleus; liver UDS) also have to be conducted [1–15].

The standard batteries of tests are selected to address two types of genetic damage of concern, *i.e.*, gene mutations and chromosome damage. Some more recent guidelines [1,15] have suggested inclusion of the *in vitro* micronucleus test to detect chromosome loss. Additional tests may be needed to clarify the substances' activity, or to determine if the activity seen in the initial testing is relevant. Such additional testing may include investigation of aneuploidy, chromosome non-disjunction, DNA interaction, and/or primary DNA damage [1,7].

The International Workshop on Genotoxicity Testing (IWGT) was formed in 1993 to bring together internationally recognized experts to examine genetic testing methods and strategies through meetings and workshops. Three workshops have been held so far, in 1993, 1999 and 2002. At the 2002 workshop, the IWGT Steering Committee initiated a discussion of testing strategies, but only limited topics were discussed and agreed upon at that time [16]. One topic that was not finalized was the selection of follow-up testing approaches following *in vitro* positive or equivocal results in the test battery. To address this issue, the IWGT Working Group reconvened during a fourth IWGT workshop in San Francisco, California, on September 9–10, 2005. The primary objective of this Working Group was to define the most appropriate follow-up testing strategy in case of positive results, and not to recommend revisions to the current batteries of tests.

During the past 30 years, genetic toxicology testing has been mainly used for hazard identification. Nevertheless, it is recognized that the discipline of regulatory genetic toxicology testing should consider moving from hazard identification towards an integrated risk assessment. At this IWGT meeting, the Working Group decided to focus on strategies for assessing the risk of cancer, although the importance of other health consequences of genetic damage to somatic and germ cells was recognized.

This publication describes those areas where a consensus was achieved among the members of the Working Group, and identifies areas that were discussed but not

resolved, or were not addressed because of time constraints or lack of available data. A general approach to determining the need for follow-up testing was defined. Case study examples are given that illustrate the extent of information (weight of evidence) needed to reach decisions about the extent of risk (risk assessment) based on the available data about the mode of action from testing results. Further, the steps needed to make recommendations about specific tests to be used in follow-up testing strategies were discussed.

2. Weight of evidence and mode of action considerations

Current regulatory practice often involves decision-making based on the results of batteries of tests designated by applicable regulatory agencies. These test batteries rely heavily on *in vitro* tests. Although there is general agreement that *in vitro* tests are useful for identifying potential genotoxic carcinogens and mutagens, the high incidence of positive findings in these *in vitro* assays [10] with agents that appear not to pose a carcinogenic health risk under certain conditions of exposure implies that reliable cancer health risk determinations cannot be made on the basis of *in vitro* findings alone. Recognition of the high rate of positive findings in *in vitro* assays has created a need for consensus agreement about how these results should be interpreted, and how appropriate follow-up testing should be structured in order to define the risks to humans.

The need to place findings into the context of their relevance to the health risks associated with specific exposures is, of course, not new. It has been recognized since regulatory mutagenicity testing was instituted in the mid-1970's that assessment of *in vivo* risk was an essential component of mutagenicity testing. A department-wide committee of the Department of Health Education and Welfare in the U.S. (the approximate equivalent of the current Department of Health and Human Services) was convened in the mid-1970's to recommend appropriate approaches to regulatory mutagenicity testing in the United States. The stated objective of this committee was "...to aid officials of regulatory agencies who have the responsibility for deciding: (1) advisability of promulgating test requirements for mutagenicity at the present time under any of their legislative authorities; (2) the appropriateness of mutagenicity tests for a wide range of product use and exposure categories; and (3) the reliability and interpretation of data from mutagenicity tests developed on substances of commerce within their regulatory purview in spite of the absence of formal testing requirements". This report [17], *Approaches to Determining the Mutagenic Properties of Chemicals: Risk to Future Generations*, stated:

> "It is not sufficient merely to identify substances which may pose a genetic hazard to the human population. Many such compounds will have a significant benefit factor and hence cannot reasonably be eliminated from use. Therefore, it is necessary to obtain quantitative data from relevant animal model systems from which extrapolation to humans can be made to predict virtually safe or tolerable levels of exposure".

Subsequently, the appropriate steps for conducting risk assessments and risk characterization of mutagens have been addressed [18] and the International Commission for Protection Against Environmental Carcinogens and Mutagens (ICPEMC) has delineated and published a detailed approach and recommendations [19,20]. The ICPEMC recommendations follow closely the general principles of risk assessment established by the landmark National Academy of Sciences report on risk assessment in the U.S. Federal government [21]. The evolution of strategies for assessment of mutagenic risk has been reviewed by MacGregor et al. [22].

Although the principles of risk assessment from exposures to genotoxic agents have been delineated, application of these principles varies within different regulatory agencies. Most place a major emphasis on mutagenicity data as a part of the weight of evidence for cancer risk assessment. The U.S. Environmental Protection Agency (EPA), in particular, established a procedure that incorporates information on mode of action as the focus of the risk assessment approach taken in the EPA cancer risk assessment guidelines [23]. The guidelines define the term "*mode of action*" as a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation. Increasingly, one of the key events and processes being considered is whether the carcinogenic agent is mutagenic/genotoxic or not, and if so, whether that genotoxic activity contributes to the induction of cancer (approach described in Dearfield and Moore [24]). Further, a genotoxic mode of action is coupled with more stringent regulatory control. For example, as described in a supplement to its cancer guidelines, if a mutagenic mode of action is determined for the induction of cancer, EPA will apply age-dependent adjustment factors to the cancer slope factor under certain conditions to assure further protection from early life exposures to the chemical [25]. As another example, genotoxic agents are generally restricted from use in healthy volunteers in clinical trials of new therapeutic agents [26,27].

The impact of genotoxicity data on regulatory decisions, interpretation of positive findings with respect to human risk, and the degree of quantitative risk assessment applied to genotoxicity data also varies among agencies. For example, most agencies acknowledge that mutagenic damage to germ cells constitutes a risk to subsequent generations [3,7,13,15,28], but it is quite rare to conduct a formal risk assessment of either germinal or somatic cell mutagenesis. EPA has a formal procedure for assessing germ cell risk [28], but has only conducted formal germ cell risk assessments in a few cases (ethylene oxide [29]; acrylamide [30]). For industrial chemicals [15] and for plant protection products [8] in Europe, an assessment of germ cell risk is required if a substance demonstrates genotoxic activity in somatic cells *in vivo*. The UK Committee on Mutagenicity guidance also includes assessment of germ cell risk as a separate exercise [1]. Although no formal guidance exists, experience indicates that European agencies assessing the safety of new medicines also view germ cell risk as separate from cancer risk. The U.S. Food and Drug Administration (FDA), in contrast, generally assumes that control of cancer risk also controls other health risks associated with genotoxic activity, and does not request quantitative assessments of germ cell risk. The approaches used by different agencies have recently been reviewed by Cimino [31].

In view of the differences in emphasis and policy among different regulatory agencies, and due in large part to the extent of positive findings in the genotoxicity tests with agents found not to be carcinogenic and thought not to pose a carcinogenic health hazard to humans under conditions of anticipated exposure [10,32–34], there is a general agreement among scientists and regulators in the field that a more detailed set of consensus principles for appropriate interpretation and follow-up testing when screening tests are positive would be useful. The focus of several important bodies on this issue is testimony to this general consensus; these include this IWGT effort, a recently initiated collaborative project of the International Life Sciences Institute (ILSI) on the relevance and follow-up of positive results in the genetic toxicology testing (http://www.hesiglobal.org/Committees/EmergingIssues/toxtesting/), and the issuance by the FDA of a new guidance that emphasizes a weight of evidence approach to assessing the relevance of genotoxicity test results [26,27]. In order to ensure recognition and widespread adoption of such principles, it is important that they be achieved *via* a consensus process among recognized international authorities in the field.

3. Negative results that may require follow-up testing

In most cases, when a chemical is found negative in the initial regulatory battery of tests, and appropriate conditions have been used, follow-up testing is not required. However, there are some situations in which additional testing may be necessary even when an initial regulatory battery of tests is negative. Such cases are also discussed in other reports in this volume (*e.g.*, Ku et al.; Tweats et al.). One important consideration is the relative metabolism in the laboratory model versus the human. Metabolism studies may show that humans generate a metabolite from the chemical under scrutiny that is not seen in the animal or cellular laboratory models (including rat liver S9 used in the tests). In this case, the chemical would not have been properly evaluated for human risk. Typically, if the metabolite were present at significant levels in human, additional testing with the metabolite itself (or systems that produced it) would be necessary to fully assess the potential of the chemical to induce genotoxic effects in humans [35].

In some cases, results from studies *in vivo* may suggest a need for additional genotoxicity testing. For example, positive or equivocal results in rodent carcinogenicity assays, epidemiology evidence in humans, or as suggested by some Working Group participants observation of pre-neoplastic lesions in toxicity studies, may trigger requests for additional genotoxicity testing. This situation is also discussed elsewhere in this volume (Kasper et al.). Such testing could include evaluation for the presence in the target organs of DNA adducts [36] and other DNA primary damage (*e.g.*, with assays for strand breaks), or indicators of genetic damage, such as micronucleated erythrocytes in the test animals at the end of the sub-chronic toxicity study. Newer tools, such as transgenic animals and genomics technologies may be useful in this regard.

Chemicals with structural alerts for mutagenicity but with negative results in an initial regulatory battery would usually not require additional testing, provided that the initial battery is sensitive to the type of effect indicated by the alert. The Working Group agreed that a structural alert can raise a concern, but study data are usually the final arbiter of hazard. However, if a chemical is in a structural class known to give positive results in specific genotoxicity tests or under specific experimental conditions that were not employed, then additional testing that includes these specific tests or conditions should be conducted.