Table 4 Major toxicity findings for (hydroxyphenyl)methyl phenol in the newborn and young rat main studies

	Newborn study (mg/kg)			Young study (mg/kg)			
	0	100	200†	0	40	200	1000
Male							
Dead or moribund	0/12	0/12	3/6	0/12	0/12	0/12	0/12
Final body weight	/	–	↓	/	–	–	↓
Total cholesterol	/	–	↑	/	–	–	↓
Relative liver weight	/	–	–	/	–	–	↑
Stomach, hyperplasia	0/6	0/6	no data	0/6	0/6	0/6	6/6
Liver, centrilobular hypertrophy	0/6	0/6	no data	0/6	0/6	2/6	4/6
Female							
Dead or moribund	0/12	0/12	3/6	0/12	0/12	0/12	1/12
Final body weight	/	–	(↓)	/	–	–	(↓)
Total cholesterol	/	–	–	/	↓	↓	↓
Relative liver weight	/	–	–	/	–	↑	↑
Stomach, hyperplasia	0/6	0/6	no data	0/6	0/6	0/6	6/6
Liver, centrilobular hypertrophy	0/6	0/6	no data	0/6	0/6	0/6	4/6

Only critical data are shown in this table. † indicates a dose from the dose-finding study. Numbers are for animals with the feature in the total examined. Slashes and bars mean no statistical significance as compared with controls. ↑ indicates significant increase $P < 0.05$. ↓indicates significant decrease at $P < 0.05$. () indicates that statistical significance was not obtained. Final body weights of surviving newborn males at 200 mg/kg in the dose-finding study were reduced by 30% (14% for females, not significant), respectively. Final body weights of young male rats at 1000 mg/kg in the main study were decreased by 11.8% (5.7% for females, not significant). Increase of relative liver weights was 13% in females at 200 mg/kg, and 16 and 27% in males and females at 1000 mg/kg in the young main study.

animals. There were no chemical-related changes with other examinations, including developmental parameters. In the young study, one female became moribund and the final body weights of males were decreased at 1000 mg/kg. All animals of both sexes at this dose showed squamous hyperplasia of the forestomach or limiting ridge with ulceration, and two-thirds of the animals featured centrilobular hypertrophy of hepatocytes with decrease of total cholesterol (29–51% drop) and increase of relative liver weight. At 200 mg/kg, low incidences of centrilobular hypertrophy in the livers of males and slight increase of liver weights in females with low total cholesterol (45% drop) were found. No toxicity was apparent at 40 mg/kg in the main study. No toxicity was also found at 100 mg/kg in the dose-finding study, but a histopathological examination was not conducted. There were no abnormalities on hematological examination and urinalysis at any dose.

The pNOAEL is considered to be 100 mg/kg/day for newborn rats and 40 mg/kg/day may be appropriate for young rats because of the limited information at 100 mg/kg in the dose-finding study. Although toxicity at 1000 mg/kg for young rats was evident, the dose inducing the same effects in newborn rats was clearly less than 200 mg/kg, because half of the animals died at this dose. We speculate that the dose range for one death in 12 newborn rats would be within 140–160 mg/kg. It is clear that the dose–response curve is much steeper for newborn than young rats. Based on our consideration, pUETLs of 140–160 and 1000 mg/kg/day may be equivalent for newborn and young rats, respectively.

Trityl chloride (Table 5)

The newborn investigation was conducted at doses of 0, 20, 60, 200, and 600 mg/kg for dose-finding and 0, 12, 60, and 300 mg/kg for the main study. The young investigation was conducted at doses of 0, 30, 100, 300, and 1000 mg/kg for dose-finding and 0, 12, 60, and 300 mg/kg for the main study.

Common effects were observed in livers of newborn and young rats. In the newborn study, increase of relative liver weights were shown at 60 mg/kg and more in both sexes and centrilobular hypertrophy of hepatocytes was noted in 300 mg/kg females. In the dose-finding newborn study, one female died and increase of relative liver weights of both sexes at 600 mg/kg was more evident with low body weights (11.3% drop for males, 13.8% for females). There were no chemical-related changes with other examinations, including developmental parameters. In the young study, both sexes at 60 mg/kg showed a high incidence of centrilobular hypertrophy of hypetocytes with limited increases of relative liver weights (10–14%). At 300 mg/kg, soft feces and mucosal thickening of cecum in most animals were observed in addition to more extensive hepatic changes. Although relative kidney weights were increased at 300 mg/kg in males and 60 and 300 mg/kg in females, there were no renal histopathological findings. Hematological and blood chemical examinations revealed several slight to moderate changes (56% as the maximum) in fibrinogen, ALT, total cholesterol and glucose, as well as prolongation of prothrombin and activated thromboplastin times, at 300 mg/kg.

pNOAELs of 60 and 12 mg/kg/day for newborn and young rats appear appropriate because of the lack of information at higher doses in the dose-finding study, which showed no toxicity but without histopathological examination. The dose of 300 mg/kg in the young main study was a clear toxic level, but intensity was much stronger than that at 300 mg/kg in the newborn main study, while less that at 600 mg/kg in the dose-finding study. Based on these data, the toxicity with 300 mg/kg for young rats is considered to be within the range with 400–500 mg/kg for newborn rats.

Table 5 Major toxicity findings for trityl chloride in the newborn and young rat main studies

	Newborn study (mg/kg)				Young study (mg/kg)			
	0	60	300	600†	0	12	60	300
Male								
Death	0/12	0/12	0/12	0/6	0/12	0/6	0/12	0/12
Final body weight	/	–	–	↓	/	–	–	↓
ALT, Total cholesterol	/	–	–	–	/	–	–	↑
Relative liver weight	/	↑	↑	↑	–	–	↑	↑
Relative kidney weight	/	–	–	–	–	–	–	↑
Cecum, thickening	0/6	0/6	0/6	no data	0/6	0/6	0/6	5/6
Liver, centrilobular hypertrophy	0/6	0/6	0/6	no data	0/6	0/6	3/6	6/6
Female								
Death	0/12	0/12	0/12	1/6	0/12	0/6	0/12	0/12
Final body weight	/	–	–	↓	/	–	–	–
ALT, Total cholesterol	/	–	–	–	/	–	–	-,↑
Relative liver weight	/	↑	↑	↑	–	–	↑	↑
Relative kidney weight	/	–	–	–	–	–	↑	↑
Cecum, thickening	0/6	0/6	0/6	no data	0/6	0/6	2/6	5/6
Liver, centrilobular hypertrophy	0/6	0/6	4/6	no data	0/6	0/6	5/6	6/6

Only critical data are shown in this table. †indicates a dose from the dose-finding study. Numbers are for animals with the feature in the total examined. Slashes and bars mean no statistical significance as compared to controls. ↑ indicates significant increase $P < 0.05$. ↓indicates significant decrease at $P < 0.05$. Relative liver weights were increased by 11% for males and 8% for females at 60 mg/kg, and 29% for both sexes at 300 mg/kg in the newborn main study and by 44% for males and 46% for females at 600 mg/kg in the newborn dose-finding study. Body weight depression in males (13%) and an increase of relative liver weights (32% for males, 40% for females) were observed at 300 mg/kg in the young main study.

Therefore, pUETLs of 400–500 and 300 mg/kg/day are proposed as appropriate for newborn and young rats, respectively.

1,3,5-Trihydroxybenzene (Table 6)

The newborn investigation was conducted at doses of 0, 100, 500, and 1000 mg/kg for dose-finding and at 0, 20, 100, and 500 mg/kg for the main study. The young investigation was conducted at doses of 0, 100, 250, 500, and 1000 mg/kg for dose-finding and at 0, 30, 100, 300, and 1000 mg/kg for the main study.

Common changes were observed in the thyroids and liver. The only toxic change in newborn main study was hypertrophy of thyroid follicular cells with increase in relative thyroid weights in both sexes at 500 mg/kg. Increased relative liver weights in females were not accompanied by any histopathological changes. Although decrease of adrenal weight and histopathological alterations such as vacuolization and pigmentation were noted at the end of the dosing and recovery-maintenance periods, these were always slight and not dose-dependent. There were no chemical-related changes with other examinations, including developmental parameters, in newborn rats. In the young study, similar effects on the thyroids and liver were found at 1000 mg/kg, but the incidence of thyroid histopathological changes was slightly less than in newborn animals at 500 mg/kg.

pNOAELs of 100 and 300 mg/kg/day for newborn and young rats can be considered appropriate because of the lack of data with dose settings between 100 to 500 mg/kg in the newborn, and no histopathological examination at 500 mg/kg in the young dose-finding study. The degree of toxicity at 1000 mg/kg for young rats was almost equal to that at 500 mg/kg for newborn rats. Therefore, pUETLs of 500 and 1000 mg/kg/day are proposed as equivalents for newborn and young rats, respectively.

DISCUSSION

More than 100 000 industrial chemicals are now in use around the world and sufficient toxicity information is available for only a small proportion. The Japanese government started the Existing Chemical Safety Program to obtain minimal toxicity data sets from 28-day toxicity studies using young rats for high production volume chemicals lacking toxicity information. For the present six targeted chemicals, we found toxicity information for only two chemicals by literature search. Daniel *et al.* (1993) reported no toxic effects of 2-chlorophenol on oral administration to male and female Sprague Dawley rats at up to 257 mg/kg for 10 days or 150 mg/kg for 90 days. Our results were consistent with their data, as we found no toxicity at 500 mg/kg in young dose-finding study (14 days administration) and at 200 mg/kg in the young study (28 days), while further providing information on CNS effects at higher doses. As for (hydroxyphenyl)methyl phenol, consisting of bisphenol D, E, and F isomers, bisphenol F has been reported to have estrogenic potential evidenced by several *in vitro* and *in vivo* experiments (Hashimoto *et al.* 2001; Yamasaki *et al.* 2002; Stroheker *et al.* 2003). However, we could not establish any such activity in this study. Our results are reasonable because oral administration of bisphenol F increased relative uterus weights only at more than 100 mg/kg, but not 50 mg/kg given during PNDs 22–25 (Stroheker *et al.* 2003), while our highest dose of (hydroxyphenyl)methyl phenol was equivalent to 30 mg/kg of bisphenol F.

Table 6 Major toxicity findings for 1,3,5-trihydroxybenzene in the newborn and young rat main studies

	Newborn study (mg/kg)			Young study (mg/kg)		
	0	100	500	0	300	1000
Male						
Relative organ weight						
Liver	/	–	–	/	–	↑
Thyroids	/	–	↑	/	–	(↑)
Histopathology						
Liver	0/6	0/6	0/6	0/6	0/6	0/6
Thyroids, hypertrophy	0/6	0/6	4/6	0/6	0/6	2/6
Female						
Relative organ weight						
Liver	/	–	↑	/	–	↑
Thyroids	/	–	(↑)	/	–	(↑)
Histopathology						
Liver	0/6	0/6	0/6	0/6	0/6	0/6
Thyroids, hypertrophy	0/6	0/6	5/6	0/6	0/6	4/6

Only critical data are shown in this table. Slashes and bars mean no statistical significance as compared with controls. ↑ indicates significant increase $P < 0.05$ (except in parentheses where statistical significance was not attained). Numbers are for animals with the feature in the total examined. Increase of relative organ weights at 500 mg/kg in the newborn main study was observed for thyroids (39% for males, 24% for females) and liver (9% for females). Increase of relative organ weights at 1000 mg/kg in the young main study was observed for thyroids (14% for males, 19% for females) and liver (23% for males and 9% for females).

Table 7 Comparative susceptibility of newborn and young rats to the six chemicals

	Newborn study		Young study		pNOAEL	pUETL
	pNOAEL	pUETL	pNOAEL	pUETL	Young/Newborn	Young/Newborn
	mg/kg/day		mg/kg/day			
2-Chlorophenol	40	200–250	200	1000	5.0	4.0–5.0
4-Chlorophenol	100	300	100	500	1.0	1.7
p-(α,α-Dimethylbenzyl) phenol	30	300	100	700–800	3.3	2.3–2.7
(Hydroxyphenyl) methyl phenol	100	140–160	40	1000	0.4	6.3–7.1
Trityl chloride	60	400–500	12	300	0.2	0.6–0.8
1,3,5-*Trihydroxybenzene*	100	500	300	1000	3.0	2.0

Although there has been no reports for p-(α,α-dimethylbenzyl) phenol, it causes endocrine disruption and possible antiestrogenic activity, when administered to newborn female rats in this study. Therefore, further studies on this chemical should be conducted to elucidate the mechanisms, because the present investigation did not indicate any effects on sexual differentiation such as preputial separation, vaginal opening and the estrous cycle.

For our focus on the comparative sensitivity of newborn and young rats to chemicals, two toxicity endpoints, pNOAEL and pUETL, were newly defined as appropriate, considering the entire data sets from both main and dose-finding studies. We believe that this alternative assessment approach allowed us to make more realistic comparisons between newborn and young rats under the same experimental conditions as far as possible.

The ratios of pNOAELs for chemicals between newborn and young rats may provide an additional UF value in risk assessment according to susceptibility of newborn rats, because regulatory limit values for chemicals to protect public health of humans, including infants, are derived from the division of NOAEL by UFs. The data in Table 7 indicate newborn rats to be 1–5 times more susceptible to four of the tested chemicals, 2- and 4-chlorophenols, p-(α,α-dimethylbenzyl) phenol and 1,3,5-trihydroxybenzene, than young rats in terms of the pNOAELs, similar to the results of previous analyzes of five phenolic chemicals, 4-nitro-, 2,4-dinitro-, 2,4,6-trinitro-, 3-methyl- and 3-amino-phenols (Koizumi *et al.* 2001, 2002, 2003; Takahashi *et al.* 2004). Immaturity in the detoxification potential of phase 1 and phase 2 enzymes in newborn animals may be the major cause of higher toxicity in newborn rats (Rich & Boobis 1997; Gow *et al.* 2001), because these chemical classes are probably direct toxicants. In the case of (hydroxyphenyl)methyl phenol, the pNOAEL (100 mg/kg/day) for newborn rats was 2.5 times higher than that (40 mg/kg/day) for young rats, but it can be speculated that values are in practice rather similar because the toxicity for young rats at the high dose, 200 mg/kg, was only slight (Table 4). As for trityl chloride, newborn rats were obviously less susceptible (0.2 for the pNOAEL ratio). Similar results were

also reported from our previous analysis for bromoalkanes (Hirata-Koizumi *et al.* 2005) and may be explained by mechanisms of action and metabolic characteristics of newborn rats. As this class of chemicals possibly requires metabolism to act as toxicants, the relatively mature metabolic enzyme status of young rats would be expected to provide toxic intermediates by metabolic activation to a greater extent than in newborn rats, as evidenced by data for previously reported chemicals (Onkenhout *et al.* 1986; Kennedy *et al.* 1993). Other compounds such as acetaminophen, bromobenzene, and carbon tetrachloride have also been shown to not produce liver injury in neonatal animals at doses that are hepatotoxic to adults (Gregus & Klaassen 1998).

The ratios of pUETLs, doses inducing the same degree of toxicity in newborn and young rats, were almost the same as for pNOAELs with the direct toxicants, as shown in Table 7. However, newborn rats were considerably more susceptible to (hydroxyphenyl)methyl phenol when considering the pUETL, due to the much steeper dose–response curve in newborn rats, with a 100 mg/kg/day pNOAEL and half the animals dying at 200 mg/kg, compared with a 40 mg/kg/day pNOAEL and only one death in 12 animals at 1000 mg/kg for young rats. Although young rats showed stomach hyperplasia in addition to hepatotoxicity at 1000 mg/kg, the cause of newborn deaths at 200 mg/kg was unclear. With regard to trityl chloride, the pUETL for young rats was almost the same as for newborn although the latter were less susceptible. Such an anomaly has also been found for bromoalkanes previously analyzed. Another example of a chemical for which susceptibility differs at low and high doses is chlorpyrifos, the maximum tolerated dose in 17-day-old rats being reported to be five times less than that in adults following oral exposure (Moser & Padilla 1998), but the differential sensitivity not appearing in low-dose exposure (Pope & Liu 1997). Thus as there are several chemicals of which dose–response curve in newborn rats was obviously steeper than that in young rats, pUETL ratios should be also taken into account for the susceptibility of newborn rats as the second endpoint marker.

In conclusion, newborn rats were 2–5 times more susceptible than young rats in terms of both the pNOAEL and the pUETL in most cases. One exception was that young rats were clearly more susceptible than their newborn counterparts for trityl chloride.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Office of Chemical Safety, Pharmaceutical and Medical Safety Bureau, Ministry of Health, Labor and Welfare, Japan and also deeply appreciate the efforts of the six Japanese contract laboratories in performing the actual animal toxicity studies.

REFERENCES

Bartlett MS (1937) Properties of sufficiency and statistical tests. *P Roy Soc Lond A Mat* **160**: 268–282.

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

Daniel FB, Robinson M, Olson GR, York RG, Condie LW (1993) Ten and ninety-day toxicity studies of 2-chlorophenol in Sprague-Dawley rats. *Drug Chem Toxicol* **16**: 277–291.

Dourson M, Charnley G, Scheuplein R (2002) Differential sensitivity of children and adults to chemical toxicity. II. Risk and regulation. *Regul Toxicol Pharmacol* **35**: 448–467.

Dunnett CW (1964) New tables for multiple comparisons with a control. *Biometrics* **20**: 482–491.

Fisher RA (1973) *Statistical Methods for Research Workers*, 14th edn. Hafner Publishing Co, New York.

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

Gow PJ, Ghabrial H, Smallwood RA, Morgan DJ, Ching MS (2001) Neonatal hepatic drug elimination. *Pharmacol Toxicol* **88**: 3–15.

Gregus Z, Klaassen CD (1998) Hepatic disposition of xenobiotics during prenatal and postnatal development. In: Polin RA, Fox WF, (eds). *Fetal and Neonatal Physiology*. Saunders, Philadelphia, pp. 1472–1493.

Hasegawa R, Koizumi M, Hirose A (2004) Principles of risk assessment for determining the safety of chemicals: Recent assessment of residual solvents in drugs and di (2-ethylhexyl) phthalate. *Congenit Anom Kyoto* **44**: 51–59.

Hashimoto Y, Moriguchi Y, Oshima H, Kawaguchi M, Miyazaki K, Nakamura M (2001) Measurement of estrogenic activity of chemicals for the development of new dental polymers. *Toxicol In Vitro* **15**: 421–425.

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

Hollander M, Wolfe DA (1973) *Nonparametric Statistical Methods*. John Wiley and Sons, New York.

IPCS (International Programme on Chemical Safety) (2002) *Global Assessment of the State-of-the-Science of Endocrine Disruptors*. Damstra T, Barlow S, Bergman A, Kavlock R, van der Kraak G (eds). World Health Organization, Geneva.

Kennedy CH, Cohen KB, Bechtold WE *et al.* (1993) Effect of dose on the metabolism of 1,1,2,2-tetrabromoethane in F344/N rats after gavage administration. *Toxicol Appl Pharmacol* **119**: 23–33.

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.

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

Landrigan PJ, Kimmel CA, Correa A, Eskenazi B (2004) Children's health and the environment: Public health issues and challenges for risk assessment. *Environ Health Perspect* **112**: 257–265.

Lee PC (1998) Disruption of male reproductive tract development by administration of the xenoestrogen, nonylphenol, to male newborn rats. *Endocrine* **9**: 105–111.

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* **18**: 50–60.

Moser VC, Padilla S (1998) Age- and gender-related differences in the time course of behavioral and biochemical effects produced by oral chlorpyrifos in rats. *Toxicol Appl Pharmacol* **149**: 107–119.

Onkenhout W, Van Bergen EJ, Van der Wart JH, Vos GP, Buijs W, Vermeulen NP (1986) Identification and quantitative determination of four different mercapturic acids formed from 1,3-dibromopropane and its 1,1,3,3-tetradeutero analogue by the rat. *Xenobiotica* **16**: 21–33.

Pope CN, Liu J (1997) Age-related differences in sensitivity to organophosphorus pesticides. *Environ Toxicol Pharmacol* **4**: 309–314.

Rich KJ, Boobis AR (1997) Expression and inducibility of P450 enzymes during liver ontogeny. *Microsc Res Tech* **39**: 424–435.

Steel RD (1959) A multiple comparison rank sum test: Treatment versus control. *Biometrics* **15**: 560–572.

Stroheker T, Chagnon MC, Pinnert MF, Berges R, Canivenc-Lavier MC (2003) Estrogenic effects of food wrap packaging xenoestrogens and flavonoids in female Wistar rats: A comparative study. *Reprod Toxicol* **17**: 421–432.

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

Takahashi O, Oishi S (2001) Testicular toxicity of dietary 2,2-bis (4-hydroxyphenyl) propane (bisphenol A) in F344 rats. *Arch Toxicol* **75**: 42–51.

Yamasaki K, Takeyoshi M, Yakabe Y *et al.* (2002) Comparison of reporter gene assay and immature rat uterotrophic assay of twenty-three chemicals. *Toxicology* **170**: 21–30.

REVIEW ARTICLE

Principles of risk assessment for determining the safety of chemicals: Recent assessment of residual solvents in drugs and di(2-ethylhexyl) phthalate

Ryuichi Hasegawa[1], Mutsuko Koizumi[1], and Akihiko Hirose[2]
Divisions of [1]Medicinal Safety Science and [2]Risk Assessment, National Institute of Health Sciences, Tokyo, Japan

ABSTRACT **Risk assessment of chemicals is essential for the estimation of chemical safety, and animal toxicity data are typically used in the evaluation process, which consists of hazard identification, dose–response assessment, exposure assessment, and risk characterization. Hazard identification entails the collection of all available toxicity data and assessment of toxicity endpoints based on findings for repeated dose toxicity, carcinogenicity or genotoxicity and species-specificity. Once a review is compiled, the allowable lifetime exposure level of a chemical is estimated from a dose–response assessment based on several measures. For non-carcinogens and non-genotoxic carcinogens, the no-observed-adverse-effect-level (NOAEL) is divided by uncertainty factors (e.g. with environmental pollutants) or safety factors (e.g. with food additives) to derive a tolerable daily intake (TDI) or acceptable daily intake (ADI), respectively. These factors include interspecies and individual differences, duration of exposure, quality of data, and nature of toxicity such as carcinogenicity or neurotoxicity. For genotoxic carcinogens, low dose extrapolation is accomplished with mathematical modeling (e.g. linearized multistage model) from the point of departure to obtain exposure levels that will be associated with an excess lifetime cancer risk of a certain level. Data for levels of chemicals in food, water and air, are routinely used for exposure assessment. Finally, risk characterization is performed to ensure that the established 'safe' level of exposure exceeds the estimated level of actual exposure. These principles have led to the evaluation of several existing chemicals. To establish a guideline for residual solvents in medicine, the permitted daily exposure (PDE), equivalent to TDI, of N,N-dimethylformamide was derived on the basis of developmental toxicity (malformation) and of N-methylpyrrolidone on the basis of the developmental neurotoxicity. A TDI for di(2-ethylhexyl)phthalate was derived from assessment of testicular toxicity.**

Key Words: chemical risk assessment, DEHP, guideline for solvents in medicine, risk assessment

INTRODUCTION

Theophrastus Bombastus von Hohenheim (Philippus Aureolus, 1493–1541), better known as Paracelsius, wrote 'dosis sola facit venenum': all substances are poisons; there is none which is not a poison'. In other words, all chemicals can produce a toxic effect at some level and duration of exposure; however, the point at which toxicity may occur is unknown for the majority of chemicals that are utilized in society. The prediction of health effects of chemicals on the basis of available toxicity information is called risk assessment.

In the case of pharmaceutical development, candidate medicines must be tested in healthy volunteers and patients, and adverse events (toxicity) as well as efficacy must be thoroughly evaluated before any approval is given. However, chemicals can not be tested in humans for toxicity evaluation. Therefore, human health effects of agents such as pesticides, food additives, drug excipients, environmental chemicals and industrial chemicals must be estimated on the bases of the results mostly from animal toxicity studies. In the article presented here, principles of chemical risk assessment are briefly described, and two examples of actual risk assessment (residual solvents in medicines and a major plasticizer) are introduced, focusing especially on malformations due to exposure.

Correspondence: Ryuichi Hasegawa PhD, Division of Medicinal Safety Science, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan. Email: hasegawa@nihs.go.jp
Received October 16, 2003; revised and accepted January 19, 2004.

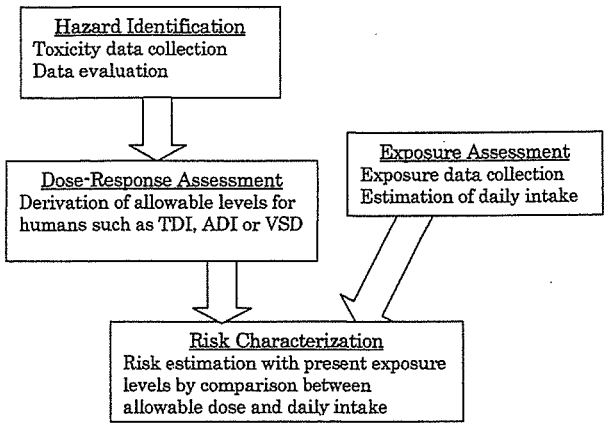

Fig. 1 Process of risk assessment. ADI, acceptable daily intake; TDI, tolerable daily intake; VSD, virtually safe dose.

Chemical risk assessment methodology

Risk assessment of chemicals is essential for the estimation of chemical safety for public health. The process consists of hazard identification, dose–response assessment, exposure assessment, and risk characterization (NRC 1983; Faustman & Omenn 2001). Toxicity assessment (hazard identification and dose–response assessment) and exposure assessment are generally conducted independently, and are merged at the final step, risk characterization (Fig 1). These practices include data collection, evaluation, and assessment, but are not necessarily the conduct of laboratory experiments. When the information is insufficient for risk assessment, additional research may be recommended.

Hazard identification

The first step entails the collection and evaluation of the available toxicity data. All information should be obtained from peer reviewed articles and if available, from pertinent reviews. Major toxicity endpoints such as short-term and long-term repeated dose toxicity, carcinogenicity, genotoxicity and reproductive/developmental toxicity are assessed. Other toxicity-related information, such as acute effects, irritation (in the eyes and skin), skin sensitization, toxicokinetics (absorption, distribution, metabolism and excretion), structure-activity relationships, and mode of action are also important to understand the toxicity profile of chemicals. Based on the available data, a no-observed-adverse-effect-level (NOAEL) or a lowest-observed-adverse-effect-level (LOAEL) for each endpoint is established, and judgment of genotoxicity and/or carcinogenicity is concluded. The following two important issues should be taken into account.

Evaluation of carcinogenicity in humans

Epidemiological information is the most important data source for assessment of human health, and is mainly derived from the following three kinds of studies:

1 Cross-sectional studies (relationship analysis between exposure and disease at a single time point in a specified population);
2 Cohort study (prospective examination of disease incidence in exposed and nonexposed populations); and
3 Case-control study (retrospective examination of exposure in disease-bearing and non-bearing populations).

Based on this information, in addition to animal carcinogenicity data, the IARC (International Agency for Research on Cancer) classifies chemicals into the following groups:

Group 1: Carcinogenic to humans;
Group 2A: Probably carcinogenic to humans;
Group 2B: Possibly carcinogenic to humans;
Group 3: Not classifiable as to its carcinogenicity to humans; and
Group 4: Probably not carcinogenic to humans.

The US EPA (Environmental Protection Agency) and the EC (European Commission) also have their own classifications into Groups A, B, C, D and E; and Categories 1, 2 and 3, respectively.

Such official conclusions regarding human carcinogenesis should be considered as key elements in hazard identification, in addition to separate full data analyzes of both epidemiological and animal carcinogenicity studies.

Animal toxicity cannot always be extrapolated to humans

Toxicity observed in some specific animal may not occur in humans. If the toxic mechanism was evidenced not to take place in humans, it could preclude the extrapolation to humans. Three typical examples are described below.

1 Rodents (rats and mice) are much more sensitive to peroxisome proliferators (fibrates, phthalates, etc.) than primates (cynomolgus monkeys and marmosets) and guinea pigs (IARC 1995). Peroxisome proliferation takes place via binding to a PPAR-α (peroxisome proliferator activated receptor-α) and no liver tumors have been induced by strong peroxisome proliferators in PPAR-α knockout mice (Peters *et al.* 1997; Ward *et al.* 1998). The m-RNA expression of PPAR-α in the livers of humans and guinea pigs is much lower than in the livers of rats and mice. Based on recent results from molecular biological studies the IARC re-classified di(2-ethylhexyl) phthalate (DEHP) from Group 2B to Group 3 in 2000 (IARC 2000).

2 α_{2U}-Globulin-related renal damages and tumors are male rat specific (Schnellman 2001). The protein is only produced in the male rat liver, appears in the blood and is excreted via the urine. When a chemical that can bind to α_{2U}-globulin is present in the blood, complexes are formed and reabsorbed to proximal epithelial cells in the kidneys after glomerular filtration. In epithelial cells, they become incorporated into lysosomes and accumulate over time due to

retarded degradation, leading to proximal tubular necrosis and finally tumors. Antibody-immunostaining can confirm this toxic mechanism. The most typical examples are unleaded gasoline, 2,2,4-trimethylpentane, d-limonene, lindane and 1,4-dichlorobenzene.

3 It is generally considered that thyroid hormone levels in humans are insensitive to chemical exposure, whereas in animals, especially male rats, they are extremely sensitive (Capen 2001). There are at least two major reasons. First, humans, monkeys and dogs have thyroxine-binding globulin in the blood, whereas rats, mice and chickens do not. Thus, rapid reduction of thyroid hormone levels in the blood can occur in the latter group, stimulating the release of thyroid hormone and leading to thyroid hypertrophy when hepatic metabolizing enzymes are induced. Second, thiourea and aniline derivatives inhibit thyroperoxidase in the thyroids of rats, mice and dogs, but not in humans, non-human primates and chickens.

Dose–response assessment

It is generally believed that there are two types of dose–response profiles. In one, toxic effects do not occur below a certain dose (i.e. threshold). In another, effects occur until the dose level reaches zero (i.e. non-threshold). Allowable lifetime exposure levels to a chemical at which no appreciable health risk would be expected over a lifetime are estimated via the different approaches for threshold and non-threshold cases.

For threshold cases, the NOAEL is divided by uncertainty factors (UF) or safety factors (SF) to derive a tolerable daily intake (TDI) or acceptable daily intake (ADI), respectively, as follows:

$$\frac{\text{NOAEL}}{\text{UF/SF}} = \text{TDI/ADI}$$

Usually, UF and TDI are used for undesirable chemicals, such as environmental pollutants and industrial chemicals, whereas SF and ADI are applied for permissible chemicals such as pesticides and food additives. However, UF and SF, TDI and ADI have basically the same meanings. UF/SF includes 5 variation components: interspecies differences; individual (intraspecies) differences; duration of exposure; quality of data; and nature of toxicity.

Interspecies differences

A factor of 10 or a body surface correction is used. Although it is generally difficult to compare toxicity levels between humans and experimental animals, information is available derived from cases in which anticancer drugs have been used for chemotherapy because they are administered up to dose levels, at which severe toxicity (the maximum tolerable) appears in patients. Data on 18 anticancer drugs in humans and from experiments with animals (rats, mice, hamsters, dogs and monkeys) showed that dose levels that induced the maximum tolerable effects were the same in humans and animals when the doses were expressed as mg/m^2 body surface area (Freireich *et al.* 1966). According to the following formula, the differences of body surface area/body weight between humans and animals are given as (human body weight)$^{1/3}$/(animal body weight)$^{1/3}$ (Freireich *et al.* 1966) as shown below:

$$\frac{\text{Body Surface}}{\text{Body Weight}} = \frac{\text{K}}{10^4 \times \text{W}^{1/3}} (\text{m}^2/\text{g})$$

$$\frac{\text{Animal (surface/weight)}}{\text{Human (surface/weight)}} \cong \frac{\text{Wh}^{1/3}}{\text{Wa}^{1/3}}$$

K is the correction factor (approximately the same in humans and animals); W is body weight (g); Wh is human body weight; and Wa is animal body weight.

For example, the human dose by body surface correction can be derived by dividing the animal dose (mg/kg body weight) by 11.4 ($60000^{1/3}/40^{1/3}$) for mice, 5.6 ($60000^{1/3}/350^{1/3}$) for rats, 2.7 ($60000^{1/3}/3000^{1/3}$) for monkeys, and 1.8 ($60000^{1/3}/10000^{1/3}$) for dogs, when the body weights are 60 kg for humans, 40 g for mice, 350 g for rats, 3 kg for monkeys, and 10 kg for dogs.

Individual differences

A factor of 10 has been commonly used from empiric findings.

Duration of exposure

Concerning the lifetime exposure, a 2-year period for rat and mouse studies for repeated dose toxicity is required. For shorter periods, factors of 2 for 1 year, 5 for 6 months and 10 for 3 months are applied.

Quality of data

If there is insufficient data for a NOAEL to be established, a lack of sufficient information in the literature, or a small number of animals, a factor of up to 10 is applied on the judgment of toxicological experts.

Alternatively, a benchmark dose approach might be used if a NOAEL is not established (Crump 1984). Figure 2 illustrates the benchmark dose approach for a 10% response. First, a dose–response curve is obtained by the application of curve-fitting technology (mathematical dose–response model) to experimental data. Second, an estimated dose for a certain incidence of toxicity or a certain percentage change in a toxicity parameter, and the lower confidence limit dose with 90–99% confidence are obtained. The latter value is a benchmark dose. A 5% incidence is usually applied for developmental toxicity and a 10% change (increase or decrease) is applied for other toxicity parameters mostly with 95% confidence limits. Advantages of the benchmark dose

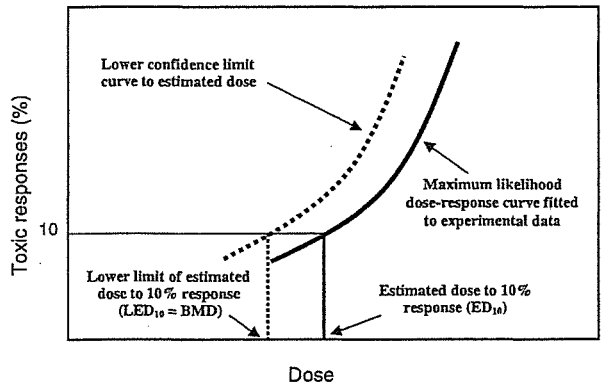

Fig. 2 Illustrated benchmark dose (BMD).

approach are: (i) its comparability to NOAEL (Farland & Dourson 1992; Allen *et al.* 1993); (ii) inclusion of the full dose–response curve and variability (number of animals and confidence limit); and (iii) generation of a realistic value because it is within the experimental range versus extrapolation from a high to a low dose. Regarding disadvantages, the benchmark dose approach is not generally applied to pathological changes because these may appear with various levels of severity and the diagnosis may change as the disease progress. The computational program compiled by US EPA can be downloaded from http://www.epa.gov/ncea/bmds.htm

Nature of toxicity

A factor of 10 is used for non-genotoxic carcinogenesis, neurotoxicity with pathological changes and malformations without maternal toxicity.

For non-threshold cases (genotoxic carcinogenesis), low dose extrapolation is accomplished with mathematical modeling (e.g. linearized multistage model) from the point of departure to obtain exposure levels that will be associated with an excess lifetime cancer risk of a certain level (generally 1 in 100 000). Usually, the lower confidence limit (95%) for the estimated dose level is called the virtually safe dose (VSD).

Exposure assessment

Measurement data for chemicals in food, water and air, are routinely used for exposure assessment. A 'market basket' methodology is often applied for estimating exposure through the food in the general population. Chemical analysis of outdoor and indoor air can be conducted, but it is very difficult to obtain constant and reliable values because winds can cause large fluctuations, and target sites and rooms can differ. Although chemical contents in drinking water can be determined in an easy and stable manner, the exposure allocation compared to other media is generally very low except with high contamination for a specific reason. Occupational exposure might provide the highest levels where sufficient protection is not in place, although the exposure data may not be generally available because they are not in the public domain.

Risk characterization

Finally, risk characterization is performed to ensure that the allowable lifetime exposure level exceeds the estimated level of exposure. Generally, practical safe levels of chemicals in food, air, water and household materials are established by regulatory authorities on the basis of TDI, ADI or VSD. If the estimated/measured levels exceed the safe level, regulatory action may be conducted case by case, concerning the excess and duration. This action is not a part of risk assessment but rather risk management.

Risk assessment is not simple because several complicating factors may be present and the political situation may interfere with the results. As recent international activities, Toxic Equivalent Factors (TEFs) and a TDI of dioxins have been established (van den Berg *et al.* 1998; WHO 1998; JECFA 2002) and revision of the WHO drinking water quality guideline is now in its final stages (WHO 2003). With these toxicity assessments, the above principles were used. New approaches such as subdivision of uncertainty factors into kinetics and dynamics (EHC 1994; Renwick & Lazarus 1998), application of benchmark dose for extrapolation of carcinogenesis assessment (US EPA, 2003) instead of mathematical models (such as a linearized multistage model), and margin of exposure or safety (Faustman & Omenn 2001), however, are also now being applied. As for Japanese risk assessment activities, we have contributed to the establishment of a Japanese drinking water standard (MHLW Japan 2003), with a proposal for risk assessment for drinking water contaminants (dichloroacetic acid, 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone [MX], formaldehyde and methyl tertiary-butyl ether [MTBE]) (Hasegawa *et al.* 1999; Hirose *et al.* 1999, 2001, 2002), and establishment of a TDI for dioxins (Hirose *et al.* 1998) and phthalates (Koizumi *et al.* 2000, 2001b, 2002b). We have also conducted infant toxicity studies on 18 chemicals to compare toxicity levels and profiles between infants and young animals. We have already reported outcomes of detailed evaluation of data for four chemicals (Koizumi *et al.* 2001a, 2002a, 2003) and are now processing the other data that has been obtained. These analytical conclusions should provide valuable information on what UF/SF is sufficient or appropriate for dose–response assessment in view of child health.

Example 1: Derivation of permitted daily exposure (PDE) of residual solvents in drug materials and products

As one activity of ICH (International Conference on Harmonization of Technical Requirements for Registration of

Table 1 Key permitted daily exposure derivation and overall assessment for N,N-dimethylformamide

Teratogenicity study data (Hass *et al.*, 1994) and PDE derivation	
Method	Russian rabbits were given 46.4, 68.1 or 200 mL/kg during the organogenic period.
Results	There was no increase in uterine deaths but decreased fetal weight was noted at 200 mL/kg along with hydrocephalus at 68.1 and 200 mL/kg, as well as umbilical hernia at high dose. No maternal effects at 68.1 mL/kg
NOAEL	46.4 mL/kg
PDE calculation	46.4 mL/kg = 46.4 × 0.9445 = 43.8 mg/kg $\text{PDE} = \frac{43.8 \times 50}{2.5 \times 10 \times 1 \times 10 \times 1} = 8.76\ \text{mg/day}$ F1: 2.5 used for species differences from rabbits F4: 10 used for malformations without maternal toxicity
Overall assessment	
IARC	Not classifiable as to its carcinogenicity to humans (Group 3)
Genotoxicity	Negative results in *in vitro* studies (six reports)
Carcinogenicity (no tumors)	Rat PDE (oral) = 30 mg/day
Reproductive/developmental toxicity	Rabbit PDE (gavage) = 8.8 mg/day, rabbit PDE (skin) = 400 mg/day, rat PDE (skin) = 1200 mg/day
General toxicity	Rat PDE (diet) = 14.1 mg/day, rat PDE (ip) = 56.7 mg/day
Human data	No chronic data available
Conclusion	PDE = 8.8 mg/day based on generation of malformations

IARC, International Agency for Research on Cancer; NOAEL, no-observed-adverse-effect-level; PDE, permitted daily exposure.

Pharmaceuticals for Human Use), a guideline for residual solvents in drug materials and products was established on the basis of PDE derivation (Connelly *et al.* 1997) and is presently in the process of maintenance. To derive a PDE, a TDI approach in chemical risk assessment has been used because there is no risk assessment concept for medicines and genotoxic materials are not basically permitted for use in humans. The following equation is applied with a modifying factor (MF) instead of a UF.

$$\text{PDE} = \frac{\text{NOAEL} \times \text{Body Weight}}{\text{MF}} (\text{mg/day})$$

MF consists of F1 (interspecies differences), F2 (individual differences), F3 (duration of exposure), F4 (nature of toxicity) and F5 (quality of data). For F1, only a body surface correction is applied. For F4, a factor of 1 is applied to reproductive toxicity with maternal (general) toxicity, 5 to reproductive toxicity without maternal toxicity and malformations with maternal toxicity and 10 to malformations without maternal toxicity. The benchmark dose approach is not applied to F5. 50 kg is employed as the body weight for patients.

As one example of 52 PDEs established in 1997, Table 1 shows PDE derivation for *N,N-dimethylformamide* from the data of developmental toxicity study (Merkle & Zeller 1980). Because malformation (hydrocephalus) without maternal toxicity was observed in a rabbit study, factors of 2.5 and 10 were used for F1 and F4, respectively. From the overall assessment, the PDE was concluded to be 8.8 mg/day, as the lowest of all calculated values.

A maintenance process has been used since 1999 for unrecognized or new information to established PDE or new solvents. In late 1999, reproductive/developmental toxicity data for two solvents were submitted to and assessed by the ICH expert working group, and one PDE was revised (Connelly *et al.* 2003). Table 2 gives an assessment summary of newly submitted data for *N-methylpyrrolidone*. A rat developmental inhalation study showed a lowering of body weight in offspring up to 5 weeks after birth and impairment of higher cognitive functions at 150 p.p.m. (Hass *et al.* 1994). This study was conducted at only a single dose level but sufficient neurotoxicity examinations were performed. The toxicity is potentially serious because it is unclear if it is permanent or reversible. Furthermore, it is not determined

Table 2 Assessment of developmental neurotoxicity for N-methylpyrrolidone

Items	Developmental toxicity study data (Hass *et al.*, 1994) and permitted daily exposure derivation
Method	Wistar rats were exposed by inhalation to 150 p.p.m. for 6 h/day, daily from days 7–20 of gestation and were then allowed to litter.
Toxicity	No maternal toxicity was detected and litter size was unaffected by the treatment.
Abnormalities	No physical abnormalities were described. The offspring were reduced in body weight, the difference being statistically significant up to week 5 after birth.
Development	Pre-weaning development was impaired as was higher cognitive function related to solving of difficult tasks. Basal function of the central nervous system was normal and there were no effects on learning of low grade tasks.
NOAEL	Not established.
PDE calculation	$150\ \text{ppm} = \frac{150 \times 99.3}{24.45} = 608.16\ \text{mg/m}^3 = 0.608\ \text{mg/L}$
	$\text{For continuous dosing} = \frac{0.608 \times 6}{24} = 0.152\ \text{mg/L}$
	$\text{Daily dose} = \frac{0.152 \times 290}{0.33} = 133.58\ \text{mg/kg}$
	$\text{PDE} = \frac{133.58 \times 50}{5 \times 10 \times 1 \times 5 \times 5} = 5.3\ \text{mg/day}$
	F4: 5 used for impairment of only higher cognitive function of offspring
	F5: 5 used for NOAEL not being established

NOAEL, no-observed-adverse-effect-level, PDE, permitted daily exposure.

if the delayed development could be due to the lower body weight of the pups. However, the expert working group decided to be cautious in its interpretation and safety decision.

Ethylene glycol is a solvent that induces renal toxicity at low dose, and malformation and acute toxicity at high dose. Although a PDE of 6.2 mg/day on the basis of malformation was established in 1997, a higher value was proposed on metabolic consideration. The metabolic pathway described in Fig. 3 shows that the step from glycolic acid to glyoxylic acid catalyzed by lactate dehydrogenase is rate-limiting, meaning that high dose induce accumulation of glycolic acid, leading to skeletal malformation. In fact, glycolic acid may accumulate above 200 mg/kg BW for mice and 1000 mg/kg BW for rats, which were related with the increased incidences of skeletal malformation in mice and rats, respectively. (Neeper-Bradley *et al.* 1995; Frantz *et al.* 1996). However, there is a possibility of higher sensitivity of lactate dehydrogenase to glycolic acid in humans than mice and we were therefore not able to accept a proposal of a higher PDE. If sufficient evidence for an inhibitory profile in humans were provided, metabolic consideration could be taken into account for risk assessment. This is an example of risk assessment for malformations concerning metabolic characteristics.

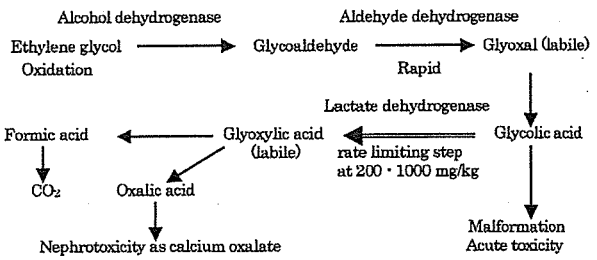

Fig. 3 Metabolic pathway of ethylene glycol.

Example 2: Establishment of a TDI for di(2-ethylhexyl) phthalate (DEHP) based on reproductive/developmental toxicity

Many kinds of phthalate esters have long been used as plasticizers. Among them, DEHP is the highest production volume chemical, found in various kinds of media. In 2001, the Division of Food Testing at our Institute reported extremely high amounts of phthalate esters, especially DEHP, to be present in certain cooked foods from convenience stores (Tsumura *et al.* 2001). Because the average amount of DEHP was 1700 μg in one Japanese style lunch package, the cause of contamination was examined. As the result, polyvinyl chloride gloves containing DEHP used in the final stage of food packing in factories were implicated.

It has been clearly shown that DEHP has two major toxicities, causing hepatic tumors and reproductive/developmental toxicity, at least in rodents. Hepatic tumors are excluded from the derivation of TDI, as the IARC (2000) concluded that hepatic tumor due to DEHP in rodents is not relevant to other animal species, including humans (Class 2B to 3), because of the association with peroxisome proliferation. Other toxic effects of DEHP can be divided into four regarding endpoints: (i) changes in male reproductive organs; (ii) alteration in female reproductive organs; (iii) reproduction toxicity; and (iv) developmental anomalies. As for male reproductive organs, the most critical toxicity data in rat testes give a NOAEL of 3.7 mg/kg/day in juveniles (Poon *et al.* 1997). It was reported that suppression of ovulation, disruption of estrus cycle and low plasma 17β-estradiol concentration occurred only at high dose level of DEHP in a single dose study (Davis *et al.* 1994). A NOAEL of 419 mg/kg/day for female reproductive organ toxicity was provided by Poon *et al.* 1997. Among a series of reproductive studies with continuous breeding in US NTP, a CD^{-1} mouse study showed the lowest NOAEL of 14 mg/kg/day (Lamb *et al.* 1987). Malformations such as hydrocephalus, cleft palate and skeletal malformation were found (Shiota *et al.* 1980) and a NOAEL of 44 mg/kg/day for developmental toxicity was apparent (Tyl *et al.* 1988).

The TDI was derived from dividing the NOAEL by a UF of 100 because the experimental conditions were sufficient for quality of data in all four cases, values for each endpoint being derived as follows:

1. Male reproductive organ toxicity
   NOAEL: 3.7 mg/kg/day → TDI: 40 μg/kg/day
2. Female reproductive organ toxicity
   NOAEL: 419 mg/kg/day → TDI: 4.2 mg/kg/day
3. Reproduction toxicity
   NOAEL: 14 mg/kg/day → TDI: 140 μg/kg/day
4. Effect on Development
   NOAEL: 44 mg/kg/day → TDI: 440 μg/kg/day

As to specific characteristics of DEHP, it is well known that there are strong species differences regarding testicular toxicity. DEHP caused severe seminiferous tubular atrophy of testes in rats and guinea pigs, and weak seminiferous tubular atrophy in mice, but none in hamsters (Gray *et al.* 1982). In two studies using marmosets and cynomolgus monkeys, no testicular toxicity was apparent (Kurata *et al.* 1998; Pugh *et al.* 2000). Considering these species differences, the Japanese government concluded a TDI of DEHP ranging from 40 to 140 μg/kg/day because it was unclear why the compound did not induce testicular toxicity in monkeys and whether this was relevant to humans (Koizumi *et al.* 2000, 2001b). As the exposure level of DEHP (118 μg/kg/day) in the worst case scenario exceeded the TDI, it was decided to ban the use of polyvinyl chloride gloves containing DEHP for food treatment.

At the same time, Gray *et al.* (2000) reported abnormalities of male reproductive organs to be observed in male offspring with exposure to several phthalate esters (500 or 750 mg/kg/day), but not to dimethyl or diethyl phthalates, from day 14 of pregnancy to lactation day 3 (Table 3). A NOAEL of 50 mg/kg/day for only di(n-butyl) phthalate was established from dose–response experiments (Mylchreest *et al.* 2000) before the establishment of any TDI for DEHP in Japan. Then, the 42nd Annual Meeting of the Society of Toxicology, Gray *et al.* (2003) described that slight antiandrogenic effects such as shortening of the anogenital distance and lowering of reproductive organ weights were observed at even 11 mg/kg/day of DEHP with maternal exposure from day 8 of pregnancy to post natal days 63–65. These seem more critical parameters than juvenile testicular toxicity. However, Gazouli *et al.* (2002) reported that DEHP can lower testosterone production *in vivo* via reduction of peripheral-type benzodiazepine receptor (PBR) expression

Table 3 Effects of phthalate esters on male offspring with exposure between late gestation to early lactation

Chemicals	Dose (mg/kg/day)	Effects on reproductive organs of male offspring	Reference
DMP	500 or 750	No effects	Gray *et al.* (2000)
DEP	500 or 750	No effects	Gray *et al.* (2000)
DBP	50	No effects	Mylchreest *et al.* (2000)
DBP	100–	Areola and nipple appearance	Mylchreest *et al.* (2000)
DBP	250–	Hypospadias, cryptorchidism and atrophy of accessory organs	Mylchreest *et al.* (2000)
DBP	500 or 750	Abnormal reproductive organs and cryptorchidism	Gray *et al.* (2000)
DEHP	500 or 750	Abnormal reproductive organs and cryptorchidism	Gray *et al.* (2000)
BBP	500 or 750	Abnormal reproductive organs and cryptorchidism	Gray *et al.* (2000)
DINP	750	Slightly abnormal reproductive organs	Gray *et al.* (2000)

BBP, n-butylbenzyl phthalate; DBP, di(n-butyl) phthalate; DEP, diethyl phthalate; DINP, di(iso-nonyl) phthalate; DMP, dimethyl phthalate.

in Leydig cells of the wild type mouse testis but it did not exert this effect in PPARα knockout mice. Therefore, there is a possibility that this antiandrogenic effect is rodent specific and would not take place in humans. The possible mechanisms underlying species differences should be clarified by further research.

As described above, species differences should need to be carefully considered for DEHP risk assessment, taking special account of appropriate mechanistic information.

REFERENCES

Allen B, Kavlock RJ, Kimmel CA, Faustman EM (1993) Comparison of quantitative dose–response modeling approaches for evaluating fetal weight changes I Segment II developmental toxicity studies. *Teratology* **47**: 41. (Abstract)

van den Berg M, Birnbaum L, Bosveld ATC, Brunström B, Cook P, Feeley M *et al.* (1998) Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect* **106**: 775–792.

Capen CC (2001) Toxic responses of the endocrine system. In: Klaassen CD (ed.) *Casarrett and Doull's Toxicology, the Basic Science of Poisons*, 6th edn. McGraw-Hill, New York, pp. 711–759.

Connelly JC, Hasegawa R, McArdle JV, Tucker ML (1997) ICH guideline residual solvents. *Pharmeuropa* **9** (Suppl. 1): S1–S68.

Connelly J, Osterberg RE, Tucker ML, Hasegawa R (2003) ICH guideline residual solvents: update. *Pharmeuropa* **15**: 365–372.

Crump KS (1984) A new method for determining allowable daily intakes. *Fundam Appl Toxicol* **4**: 854–871.

Davis BJ, Maronpot RR, Heindel JJ (1994) Di-(2-ethylhexyl) phthalate suppresses estradiol and ovulation in cycling rats. *Toxicol Appl Pharmacol* **128**: 216–223.

EHC (1994) Assessing human health risks of chemicals: derivation of guidance values for health-based exposure limits. *Environmental Health Criteria*, **Vol. 170**. International Programme on Chemical Safety, World Health Organization, Geneva.

Farland W, Dourson ML (1992) Noncancer health endpoints: approaches to quantitative risk assessment. In: Cothern R (ed.) *Comparative Environmental Risk Assessment*. Lewis Publishers, Boca Raton, Florida, pp. 87–106.

Faustman EM, Omenn GS (2001) Risk assessment. In: Klaassen CD (ed.) *Casarrett and Doull's Toxicology, the Basic Science of Poisons*, 6th edn. McGraw-Hill, New York, pp. 83–104.

Frantz SW, Beskitt JL, Grosse CM, Tallant MJ, Dietz FK, Ballantyne B (1996) Pharmacokinetics of ethylene glycol. II. Tissue distribution, dose-dependent elimination, and identification of urinary metabolites following single intravenous, peroral or percutaneous doses in female Sprague-Dawley rats and CD-1 mice. *Xenobiotica* **26**: 1195–1220.

Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE (1966) Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep* **50**: 219–244.

Gazouli M, Yao ZX, Boujrad N, Corton JC, Culty M, Papadopoulos V (2002) Effect of peroxisome proliferators on Leydig cell peripheral-type benzodiazepine receptor gene expression, hormone-stimulated cholesterol transport, and steroidogenesis: Role of the peroxisome proliferator-activator receptor alpha. *Endocrinology* **143**: 2571–2583.

Gray LE, Barlow NJ, Furr JR, Brock J, Silva MJ, Barr DB *et al.* (2003) Transgenerational effect di (2-ethylhexyl) phthalate in the male rat. *Toxicologist* **72**: 1376. (Abstract)

Gray LE, Ostby J, Furr J, Price M, Veeramachaneni DNR, Parks L (2000) Perinatal exposure to the phthalates in DEHP, BBP and DINP but not DEP, DMP or DOTP alters sexual differentiation of the male rat. *Toxicol Sci* **58**: 350–365.

Gray TJ, Rowland IR, Foster PM, Gangolli SD (1982) Species differences in the testicular toxicity of phthalate esters. *Toxicol Lett* **11**: 141–147.

Hasegawa R, Hirose A, Nishikawa A, Kurebayashi H, Ema M, Kurokawa Y (1999) Toxicity evaluation of dichloroacetic acid and estimation of the tolerable daily intake through oral route. *J Japan Soc Water Environ* **22**: 821–826. (In Japanese)

Hass U, Lund SP, Elsner J (1994) Effects of prenatal exposure to N-methylpyrrolidone on postnatal development and behavior in rats. *Neurotoxicol Teratol* **16**: 241–249.

Hirose A, Hasegawa R, Kurokawa Y (1998) Toxicity of dioxins in the body. *J Toxicol Sci* **23**: 93–106. (In Japanese)

Hirose A, Kamata E, Nishikawa A, Kurebayashi H, Ema E, Ando M *et al.* (2001) Toxicity evaluation of formaldehyde and its risk assessment for drinking water via oral and inhalation exposure. *J Japan Soc Water Environ* **24**: 308–316. (in Japanese).

Hirose A, Nishikawa A, Ema M, Kurebayashi H, Yamada M, Hasegawa R (2002) Toxicological properties on methyl tertiary-butyl ether (MTBE). *J Japan Soc Water Environ* **25**: 491–496. (In Japanese)

Hirose A, Nishikawa A, Kinae N, Hasegawa R (1999) 3-Chloro-4-(dichloromethyl) -5-hydroxy -2 (5H) -furanone (MX): toxicological properties and the risk assessment in drinking water. *Rev Environ Health* **14**: 103–120.

IARC (1995) *Peroxisome Proliferation and its Role in Carcinogenesis (Technical Report no. 24)*. IARC Press, Lyon.

IARC (2000) Di (2-ethylhexyl) phthalate. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Some Industrial Chemicals*, **Vol. 77**. World Health Organization, Lyon, pp. 41–148.

JECFA (2002) Polychlorinated dibenzodioxins, polychlorinated dibenzofurans, and coplanar polychlorinated biphenyls. *WHO Food Additives Series*, 48.

Koizumi M, Ema M, Hirose A, Hasegawa R (2000) Recent studies on toxic effects of phthalates to reproduction and development. *Jpn J Food Chem* **7**: 65–73. (In Japanese)

Koizumi M, Ema M, Hirose A, Kurokawa Y, Hasegawa R (2001b) No observed adverse effect levels of phthalate esters on reproductive and developmental toxicity, the differences with age and species in testicular toxicity, and tolerable daily intake of DEHP. *Jpn J Food Chem* **8**: 1–10. (In Japanese)

Koizumi M, Ema M, Hirose M, Ohno Y, Inoue T, Hasegawa R (2002b) Toxicity evaluation of DINP and estimation of the tolerable daily intake. *Jpn J Food Chem* **9**: 39–45. (In Japanese)

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

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

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

Kurata Y, Kidachi F, Yokoyama M, Toyota N, Tsuchitani M, Katoh M (1998) Subchronic toxicity of di (2-ethylhexyl) phthalate in common marmosets: lack of hepatic peroxisome proliferation, testicular atrophy, or pancreatic acinar cell hyperplasia. *Toxicol Sci* **42**: 49–56.

Lamb JCIV, Chapin RE, Teague J, Lawton AD, Reel JR (1987) Reproductive effects of four phthalic acid esters in the mouse. *Toxicol Appl Pharmacol* **88**: 255–269.

Merkle J, Zeller H (1980) Untersuchungen von Acetamiden und Formamiden auf embryotoxische und teratogene Wirkung bei Kaninchen. [Studies on acetamides and formamides for embryotoxic and teratogenic activities in the rabbit (author's transl)]. *Arzneimittelforschung* **30**: 1557–1562. (In German.)

MHLW Japan (2003) *Report on the Re-evaluation Japanese Drinking Water Standard*, 28 April, 2003. Japanese Ministry of Health, Labour and Welfare, Tokyo. (In Japanese.)

Mylchreest E, Wallace DG, Cattley RC, Foster PM (2000) Dose-dependent alterations in androgen-regulated male reproductive development in rats exposed to Di(n-butyl) phthalate during late gestation. *Toxicol Sci* **55**: 143–151.

Neeper-Bradley TL, Tyl RW, Fisher LC, Kubena MF, Vrbanic MA, Losco PE (1995) Determination of a no-observed-effect level for developmental toxicity of ethylene glycol administered by gavage to CD rats and CD-1 mice. *Fundam Appl Toxicol* **27**: 121–130.

National Research Council in USA (NRC) (1983) *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press, Washington, DC.

Peters JM, Cattley RC, Gonzalez FJ (1997) Role of PPAR alpha in the mechanism of action of the nongenotoxic carcinogen and peroxisome proliferator Wy-14,643. *Carcinogenesis* **18**: 2029–2033.

Poon R, Lecavalier P, Mueller R, Valli VE, Procter BG, Chu I (1997) Subchronic oral toxicity of di-n-octyl phthalate and di (2-ethylhexyl) phthalate in the rat. *Food Chem Toxicol* **35**: 225–239.

Pugh G, Isenberg JS, Kamendulis LM, Ackley DC, Clare LJ, Brown R *et al.* (2000) Effects of di-isononyl phthalate, di-2-ethylhexyl phthalate, and clofibrate in Cynomolgus monkeys. *Toxicol Sci* **56**: 181–188.

Renwick AG, Lazarus NR (1998) Human variability and non-cancer risk assessment- an analysis of the default uncertainty factor. *Regul Toxicol Pharmacol* **27**: 3–20.

Schnellman RG (2001) Toxic Responses of the Kidney. In: Klaassen CD (ed.) *Casarrett and Doull's Toxicology, the Basic Science of Poisons*, 6th edn. McGraw-Hill, New York, pp. 491–514.

Shiota K, Chou MJ, Nishimura H (1980) Embryotoxic effects of di-2-ethylhexyl phthalate (DEHP) and di-n-buty phthalate (DBP) in mice. *Environ Res* **22**: 245–253.

Tsumura Y, Ishimitsu S, Kaihara A, Yoshii K, Nakamura Y, Tonogai Y (2001) Di (2-ethylhexyl) phthalate contamination of retail packed lunches caused by PVC gloves used in the preparation of foods. *Food Addit Contam* **18**: 569–579.

Tyl RW, Price CJ, Marr MC, Kimmel CA (1988) Developmental toxicity evaluation of dietary di(2-ethylhexyl) phthalate in Fischer 344 rats and CD-1 mice. *Fundam Appl Toxicol* **10**: 395–412.

US EPA (2003) Notice of availability and opportunity to provide comment on the draft final guidelines for carcinogen risk assessment and the draft supplemental guidance for assessing cancer susceptibility from early- life exposure to carcinogens. *Federal Register (3 March 2003)* **68** (41): 10012–10015.

Ward JM, Peters JM, Perella CM, Gonzalez FJ (1998) Receptor and nonreceptor-mediated organ-specific toxicity of di (2-ethylhexyl) phthalate (DEHP) in peroxisome proliferator-activated receptor alpha-null mice. *Toxicol Pathol* **26**: 240–246.

WHO (1998) Executive summary: Assessment of the health risk of dioxins: re-evaluation of the Tolerable Daily Intake (TDI). WHO Consultation (25–29 May 1998), Geneva, Switzerland. (http://www.who.int/pcs/docs/dioxin-exec-sum/exe-sum-final.doc).

WHO (2003) *WHO Guidelines for Drinking Water Quality*, Draft, 3rd edn, World Health Organization, Geneva, Switzerland. http://www.who.int/water_sanitation_health/dwq/guidelines3rd/en/print.html

Available online at www.sciencedirect.com
SCIENCE DIRECT
Mutation Research 588 (2005) 129–135
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals

Makoto Hayashi [a,*], Eiichi Kamata [b], Akihiko Hirose [b], Mika Takahashi [b], Takeshi Morita [c], Makoto Ema [b]

[a] *Division of Genetics and Mutagenesis, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*
[b] *Division of Risk Assessment, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*
[c] *Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*

Received 12 June 2005; received in revised form 20 September 2005; accepted 26 September 2005
Available online 28 October 2005

Abstract

Genotoxicity is one of the important endpoints for risk assessment of environmental chemicals. Many short-term assays to evaluate genotoxicity have been developed and some of them are being used routinely. Although these assays can generally be completed within a short period, their throughput is not sufficient to assess the huge number of chemicals, which exist in our living environment without information on their safety. We have evaluated three commercially available in silico systems, i.e., DEREK, MultiCASE, and ADMEWorks, to assess chemical genotoxicity. We applied these systems to the 703 chemicals that had been evaluated by the Salmonella/microsome assay from CGX database published by Kirkland et al. [1]. We also applied these systems to the 206 existing chemicals in Japan that were recently evaluated using the Salmonella/microsome assay under GLP compliance (ECJ database). Sensitivity (the proportion of the positive in Salmonella/microsome assay correctly identified by the in silico system), specificity (the proportion of the negative in Salmonella/microsome assay correctly identified) and concordance (the proportion of correct identifications of the positive and the negative in Salmonella/microsome assay) were increased when we combined the three in silico systems to make a final decision in mutagenicity, and accordingly we concluded that in silico evaluation could be optimized by combining the evaluations from different systems. We also investigated whether there was any correlation between the Salmonella/microsome assay result and the molecular weight of the chemicals: high molecular weight (>3000) chemicals tended to give negative results. We propose a decision tree to assess chemical genotoxicity using a combination of the three in silico systems after pre-selection according to their molecular weight.

Keywords: In silico; (Quantitative) structure-activity relationship; (Q)SAR; Chemical genotoxicity; Decision tree

1. Introduction

It is said that more than 20,000 chemicals are in use in Japan. Among them, only approximately 10% are thought to have been assessed for human hazard based on data from in vitro and in vivo bioassays. According to the "Law Concerning the Evaluation of Chemical Substances and Regulation of Their Manufacture, etc." [2], the Salmonella/microsome (Ames) assay, in vitro chromosomal aberration assay (or alternatively mouse lymphoma TK assay), and 28-day repeat dose toxicity test in rodents are obligatory to notify new chemicals for production/import at a level of more than 10 t per year.

* Corresponding author.

1383-5718/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2005.09.009

To screen the remaining 18,000 chemicals for human hazard by application of this three-test battery is not realistic from the time and economical point of view. We need a much higher-throughput system to assess these chemicals, at least for prioritization of those chemicals that should be submitted to biological testing. To assess human hazard for regulatory purposes, in silico systems are now beginning to be used [3]. Here, we evaluated three commercially available in silico (quantitative) structure-activity relationship ((Q)SAR) systems and tried to construct a decision tree for prioritization of which chemicals need in vitro and/or in vivo testing. Also, within the drug discovery process, integrated computational analysis has been proposed to be incorporated as a toxicity prediction tool [4].

Kirkland et al. [1] published a database (CGX database, see http://www.lhasalimited.org/cgx) for nearly 1000 carcinogens and non-carcinogens with results of representative in vitro genotoxicity assays, i.e., Salmonella/microsome assay (Ames), mouse lymphoma TK assay using L5178Y cells (MLA), and in vitro chromosomal aberration assay or in vitro micronucleus assay (CA/MN). We used 703 chemicals that had been assessed in the Ames assay for evaluation of the three in silico systems, i.e., DEREK, MultiCASE (MCase), and ADMEWorks (AWorks). We also used a database (the ECJ database) that we constructed from chemicals existing in Japan that had recently been assessed in the Ames assay, in vitro chromosomal aberration assay, and 28 day repeat dose rodent toxicity test and/or reproductive and developmental toxicity test for their safety evaluation under GLP compliance. The ECJ database consisted of 206 chemicals but only 26 chemicals were positive by the Ames assay. Initially we evaluated both sensitivity and specificity of these three systems using the ECJ database of 206 chemicals [5].

We selected these three in silico systems because of their different modes of analysis. DEREK is a rule-based system [6], MCase [7] is a database/substructure based system, and AWorks is a QSAR. We applied these systems individually to assess gene-mutation induction on the 703 and 206 chemical sets described above and evaluated their sensitivity, specificity, concordance, and applicability (how many chemicals could be assessed), independently.

It is known that high molecular weight polymers tend not to induce gene mutation and chromosomal aberrations mainly because they cannot enter the target cells to react with DNA, or other bio-molecules necessary for genetic stability. We analyzed 194 Ames positive chemicals (confidential source) for the effect of molecular weight.

2. Materials and methods

2.1. Data sources for chemicals assessed

Of about 1000 chemicals, 703 that had been assessed in the Ames test were chosen from the CGX database published by Kirkland et al. [1]. All chemical structures were re-drawn using Chemdraw Ultra (Cambridge Soft Corporation, USA) and converted to MOL files before application to each system. We also used the database of 206 chemicals evaluated in the MHLW project "Safety Examination of Existing Chemicals and Safety Programmes in Japan" (ECJ database). The test summary for each of these chemicals can be seen at http://wwwdb.mhlw.go.jp/ginc/html/db1.html. In addition, we collected 194 Ames positive chemicals from a confidential source and investigated the relationship between gene mutation induction and molecular weight, with identification of any active side chain that might have contributed to the positive result in the Ames assay.

2.2. In silico systems used and definition of positive and negative responses

We used DEREK (Lhasa Ltd., UK) version 8.0.1. When the system gave an evaluation as "certain", "probable" or "plausible" we considered this as "positive", and when the system gave "equivocal", "doubted", "improbable", "impossible", or "no alert" we considered this as "negative". We used MCase (Multicase Co. Ltd.) version mc4pc. When the system gave "active" or "marginal" we considered this as "positive", and when the system gave "Inactive" we considered this as "negative". In the case of AWorks (Fujitsu Kitakyushu, Co. Ltd., version 2.0), we considered as "positive" when system evaluation was "positive", and considered as "negative" when the system evaluation was "negative". We excluded chemicals from further analysis when DEREK or AWorks gave no answer, or the evaluation was "inconclusive" by MCase.

2.3. Definition of sensitivity, specificity, concordance, and applicability

We calculated sensitivity, specificity, concordance, and applicability as follows:

$$\text{sensitivity} = \frac{N_{A+S+}}{N_{A+}} \times 100, \qquad \text{specificity} = \frac{N_{A-S-}}{N_{A-}} \times 100,$$

$$\text{concordance} = \frac{N_{A+S+} + N_{A-S-}}{N_{\text{eval}}} \times 100,$$

$$\text{applicability} = \frac{N_{\text{eval}}}{N_{\text{all}}} \times 100$$

where N_{A+} is number of chemicals revealing positive in Ames assay; N_{A-} is number of chemicals negative in Ames assay; N_{A+S+} is number of chemicals revealing positive by both Ames assay and in silico evaluation; N_{A-S-} is number of chemicals negative in both Ames assay and in silico evaluation; N_{eval} is

Table 1
Performance of in silico systems

	Ames result	+	–	Total	Sensitivity (%)	Specificity (%)	Concordance (%)	Applicability (%)
CGX database								
DEREK	+	288	64	352				
	–	69	267	336	81.8	79.5	80.7	97.9
	Total	357	331	688				
MCase	+	235	32	267				
	–	6	249	255	88.0	97.6	92.7	74.3
	Total	241	281	522				
AWorks	+	267	89	356				
	–	149	187	336	75.0	55.7	65.6	98.4
	Total	416	276	692				
ECJ database								
DEREK	+	19	7	26				
	–	21	159	180	73.1	88.3	86.4	100.0
	Total	40	166	206				
MCase	+	13	7	20				
	–	13	133	146	65.0	91.1	88.0	80.6
	Total	26	140	166				
AWorks	+	19	7	26				
	–	54	124	178	73.1	69.7	70.1	99.0
	Total	73	131	204				

MCase: MultiCASE; AWorks: ADMEWorks.

number of chemicals evaluated; and N_{all} is total number of chemicals subjected.

3. Results

Among the set of 703 CGX chemicals with published Ames data, 358 were positive and 345 were negative. The results of the in silico evaluation are summarized in Table 1. The highest sensitivity, specificity, and concordance with Ames assay results was provided by MCase, then followed by DEREK. However, the systems that showed the best applicability were AWorks and (almost the same) DEREK, then followed by MCase. For the database of 206 ECJ chemicals, 26 were positive and 180 were negative. The outcomes of the in silico analyses are summarized in Table 1. The pattern of performance was very similar to that with the 703 chemicals in the CGX database.

Fig. 1 shows the cumulative percent of Ames positive chemicals against molecular weight. It can be seen that 87.1% of those positive chemicals had molecular weights less than 1000, and 96.4% had molecular weights less than 3000; in other words, only 3.6% of the chemicals with a molecular weight >3000 gave a positive response in the Ames assay. Seven of 194 Ames positive chemicals had a molecular weight >3000 and four of these seven polymers had epoxy groups.

When we combined the in silico systems, the performance was different from that when assessed individually (Table 2). If we considered the in silico mutagenicity as positive (or negative) when two or more systems gave positive (or negative) evaluations, 87.8

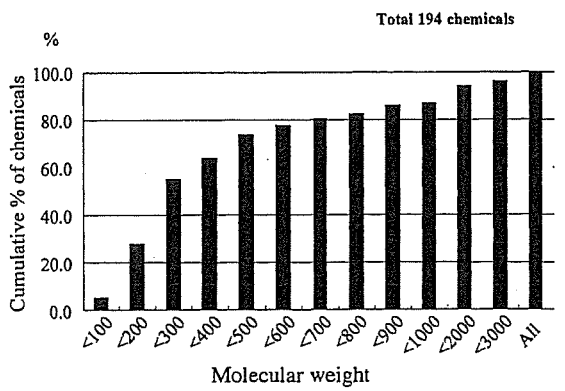

Fig. 1. Cumulative percentage of chemicals based on their molecular weight. 194 Ames positive chemicals were analyzed. 7/194 chemicals were more than 3000 molecular weight and Ames positive and 4/7 contained epoxy groups.

Table 2
Performance of in silico systems after combined

CGX database

Ames \ In silico	++ or +++	-- or ---	Total	Sensitivity (%)	Specificity (%)	Concordance (%)	Applicability (%)
+	279	40	319				
-	42	249	291	**87.8**	**85.6**	**86.7**	**86.8**
Total	321	289	610				
	+++	---					
+	166	1	167				
-	3	127	130	**99.4**	**97.7**	**98.7**	**42.2**
Total	168	129	297				

ECJ database

Ames \ In silico	++ or +++	-- or ---	Total	Sensitivity (%)	Specificity (%)	Concordance (%)	Applicability (%)
+	19	7	26				
-	23	147	170	**73.1**	**86.5**	**84.7**	**95.1**
Total	42	154	196				
	+++	---					
+	13	2	15				
-	5	94	99	**86.7**	**94.9**	**93.9**	**55.3**
Total	18	96	114				

Table 3
Performances of DEREK and MCase in several published papers.

Target compounds	In silico system	Sensitivity (%)	Specificity (%)	Concordance (%)	Applicability (%)	Reference
394 Drugs	DEREK	52	75	74	94[a]	[11]
	MCase	48	93	90	91[a]	
217 Non-drugs	DEREK	86	50	81	100[a]	[10]
	MCase	91	62	83	100[a]	
520 Drug candidates	DEREK	28	80	72	100	[13]
	MCase	50	86	81	41	
	DEREK + MCase	29	95	88	29	
	DEREK + Mcase + TOPKAT	75	96	95	15	
123 Drug candidates	DEREK	8[b]	31[c]	61	100[d]	[4]
	MCase (A2H)	13[b]	15[c]	72	97[d]	
	Topcat (Ames Mut)	18[b]	15[c]	67	98[d]	
	DEREK + MCase	6[b]	19[c]	75	97[d]	
	DEREK + Mcase + TOPKAT	5[b]	9[c]	86	46[d]	
94 Non-drugs	DEREK	63	81	76	100	[13]
	MCase	40	90	76	75	
	DEREK + MCase	47	100	85	56	
	DEREK + Mcase + TOPKAT	55	100	86	37	
516 Non-drugs	DEREK	6[b]	24[c]	70	100[d]	[4]
	MCase (A2H)	7[b]	12[c]	81	98[d]	
	Topcat (Ames Mut)	25[b]	19[c]	56	97[d]	
	DEREK + MCase	2[b]	16[c]	82	98[d]	
	DEREK + Mcase + TOPKAT	7[b]	10[c]	83	43[d]	

[a] Calculated by us
[b] % False negative.
[c] % False positive.
[d] (1-Indeterminate).

and 73.1% sensitivity, 85.6 and 86.5% specificity, 86.7 and 84.7% concordance, and 86.8 and 95.1% applicability were obtained for the CGX and ECJ databases, respectively. If we considered the in silico mutagenicity as positive (or negative) only when all three systems gave positive (or negative) evaluations, all performance measures (sensitivity, specificity, etc.) increased up to 98.7 and 93.9%. However, applicability decreased to 42.2 and 55.3%, which meant only about half of the chemicals in the CGX and ECJ databases could be evaluated. One chemical, *o*-phenylphenol [90-43-7], was positive in the Ames test but negative by all three in silico systems and three chemicals, carboxymethylnitrosourea [60391-92-6], methidathion [950-37-8], 1-nitroso-3,5-dimethyl-4-benzoylpiperazine [61034-40-0], were negative in the Ames test although all three in silico system gave positive evaluation for mutagenicity in the CGX database. When we used the ECJ database, 2-amino-1-naphthalenesulfonic acid [81-16-3] and 2-vinylpyridine [100-69-6] were positive in the Ames test but negative by all three in silico systems and there was no chemical that was negative in the Ames assay and all positive in in silico system. These exceptional chemicals are listed in Table 3 together with such chemicals taken from literatures.

4. Discussion

It is important to construct a strategy for efficient evaluation of the toxicity of a large number of existing chemicals. Even so-called short-term assays, e.g., Ames assay and in vitro chromosomal aberration assay, can practically assess only 100 chemicals per year according to our experiences in Japan. In this case, it will take 180 years to assess the outstanding 18,000 existing chemicals for genotoxicity, and it will take even longer when repeat dose toxicity tests are also performed, as these are not short-term assays. We therefore need higher-throughput systems to assess chemical safety, or at least to set priorities for those chemicals that should be tested in in vitro and/or in vivo tests. In silico systems have the capability for high throughput but have not yet been well validated for assessment of human hazard, although some regulatory bodies have started to use these methods.

Correlation between the Ames assay result and molecular weight could be explained by the lack of membrane permeability of high molecular weight chemicals, making it more difficult for them to reach target molecules such as DNA and proteins that contribute to the fidelity of cell division. Therefore, only a few chemicals with molecular weight >3000 gave positive responses in the Ames assay. This phenomenon is also true for induction of chromosomal aberrations in vitro (data not shown). The other important issue is the contribution of epoxy group in the polymer. Although of molecular weight >3000, some polymers with an epoxy group gave positive results in both the Ames and chromosomal aberration assays. Epoxy embedding reagents employed in electron microscopy (e.g., epon and araldite) have been reported as mutagenic in the Ames assay [8]. According to these findings, we should include a step to evaluate molecular weight and existence of any epoxy groups in the molecule.

In the present study, we used the CGX database recently published by Kirkland et al. [1] for microbial mutagenicity data on 358 carcinogens and 345 non-carcinogens for validation of three commercially available in silico (Q)SAR systems. When applied individually, MCase gave high sensitivity, specificity, and concordance compared to other two systems. One of the reasons may be because the CGX database contained many results from the U.S. National Toxicology Program (NTP), and the learning dataset of MCase would have used many of the same results. Therefore, some of them were evaluated by direct matching. Moreover, the applicability of MCase was relatively low compared with the other systems in this study (Table 1). MCase judged 119 chemicals as inconclusive and one chemical as marginal, and could not evaluate 67 chemicals. Such selectivity in MCase may contribute to the high concordance. On the other hand, the other systems were not influenced directly by the NTP data. We applied the in silico systems to another dataset, the ECJ database, that does not contain the NTP data and we obtained similar patterns of sensitivity, specificity, etc.

Each in silico system showed different outcomes on some chemicals complimentary by some extent. These different evaluation patterns were mainly due to the different evaluation rules. The DEREK is a rule-based system, AWorks is a discriminant-based system mainly depending on physicochemical descriptors, and MCase is a hybrid system based on a database. Therefore, we concluded that in silico evaluation could be optimized by combining the evaluations from the three systems. Sensitivity, specificity and concordance were increased when we combined the three in silico systems to make a final conclusion of mutagenicity (Table 1). Concordance was much higher after combining but the applicability became poor (42.2%). When two of the in silico systems gave the same evaluations, the applicability (86.8%) was good but the concordance was lower (86.7%) than when all three were combined (98.7).

Recently, several in silico studies for prediction of mutagenicity have been conducted on drugs or non-

Table 4
Exceptional chemicals that showed Ames test gave positive but all three in silico systems (DEREK, MCase, TOPKAT/AWorks) gave negative and Ames test gave negative but all three systems gave positive

Compound	CAS	Ames test	DEREK	MCase	TOPKAT/Aworks	Source[a]
Bupropion	34911-55-2	+	–	–	–	1
Citalopram	59729-33-8	+	–	–	–	1
Naloxone	465-65-6	+	–	–	–	1
Oxcarbazepime	28721-07-5	+	–	–	–	1
Quetiapine	111976-69-7	+	–	–	–	1
Rabeprazole	117976-89-3	+	–	–	–	1
Zolmitriptan	139264-17-8	+	–	–	–	1
2-(2-Methylpropyl) thiazole	18640-74-9	+	–	–	–	2
2-Chloropyridine	109-09-1	+	–	–	–	2
Pyrogallol	87-66-1	+	–	–	–	2
o-Phenylphenol	90-43-7	+	–	–	–	3
2-Amino-1-naphthalenesulfonic acid	81-16-3	+	–	–	–	3
2-Vinylpyridine	100-69-6	+	–	–	–	3
Fosfomycin	23155-02-4	–	+	+	+	1
Toremifene	89778-26-7	–	+	+	+	1
Poly (2-hydroxypropyl methacrylate)	25703-79-1	–	+	+	+	2
Carboxymethylnitrosourea	60391-92-6	–	+	+	+	3
Methidathion	950-37-8	–	+	+	+	3
1-Nitroso-3,5-dimethyl-4-benzoylpiperazine	–	+	+	+	3	

[a] 1: Synder et al. [11] (with TOPKAT), 2: White et al. [13] (with TOPKAT), 3: this study (with AWorks).

drug chemicals with commercially available programs, e.g., DEREK, MCase or TOPKAT, or newly developed computational approaches [4,9–12]. The performances of DEREK and MCase in several of these studies are summarized in Table 4. Generally, similar performance in sensitivity, specificity, concordance, and applicability were shown between DEREK and MCase but with some exceptions, e.g., sensitivity in 520 drug candidates [13], specificity in 516 non-drugs [4], and applicability in 520 pharmaceutical drug candidates and 94 non-drugs [13]. These differences might be due to the chemical class of target compounds in each database. However, there was no remarkable difference in performance whether the chemical was intended for use as a pharmaceutical, agricultural, or industrial agent. Our results on performance of in silico systems showed similarity with the published analyses. With respect to the combination of in silico prediction systems, White et al. [13] reported that combination improved the overall accuracy and specificity, but sensitivity was barely above the 50% level (Table 4). On the other hand, their analysis showed quite low applicability in the combination of three prediction systems, DEREK, MCase and TOPKAT. Our analysis of the combination of DEREK, MCase and AWorks showed good improvements in sensitivity, specificity and concordance, but applicability was low, especially in the 3-system combination.

Exceptional chemicals that gave positive Ames results but were negative in all three in silico systems (DEREK, MCase, TOPKAT/AWorks), and those that were negative in the Ames test but gave positive evaluations in all three systems, are summarized in Table 4. This table, which includes data from Synder et al. [11] and White et al. [13] shows there are 19 exceptional chemicals from both drug and non-drug families. Although it would be unrealistic to expect zero exceptions using this approach, further improvement of the prediction systems is needed. We do not have good reasons to explain the discordance, therefore we will verify the results from both sides, i.e., in silico system and Ames test.

Considering these outcomes, we propose a decision tree (Fig. 2), in order to evaluate chemical induction of gene mutation. We may use the decision tree to prioritize chemicals to be assayed by in vitro and/or in vivo tests. A final goal being that eventually, chemical mutagenicity will be evaluated by in silico systems alone for regulatory use. The decision tree consists of three steps; namely to assess the molecular weight, the existence of epoxy groups, and the in silico evaluation for genotoxicity. Based on the purpose of the in silico evaluation, the tree might be altered by the different final call of the in silico evaluation, i.e., regarding as positive (negative) all three systems show positive (negative). The choice of definition for final call applying to the decision tree should be based on the balance between accuracy of eval-

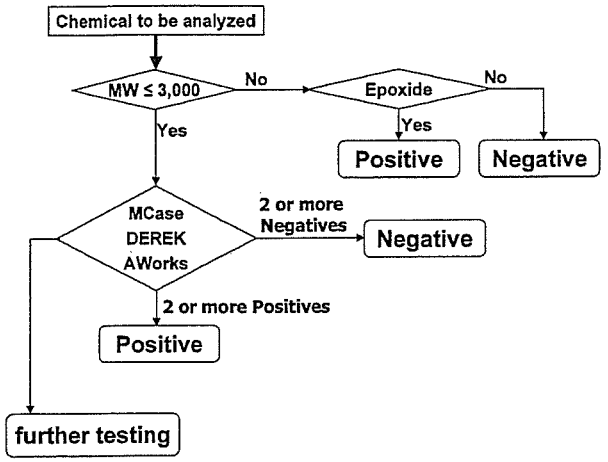

Fig. 2. Decision tree. In in silico evaluation, when two or more give positive then the final call is "positive" and two or more negative then call "negative".

uation and applicability, which are especially important for regulatory purpose. The decision should be made on a case-by-case basis depending upon the purpose of the decisions to be made.

Acknowledgements

The authors thank Dr. D. Kirkland, Covance, for his critical review and kind English edition and Mr. T. Ehara, MHLW, for his invaluable discussion. Authors also want to thank Ms. H. Akiyama (CTC laboratories, Japan), Akamatsu and Naitoh (Charles River Japan, Japan), and Kitajima, Suiroi, and Yuta (Fujitsu Kitakyushu, Japan) for their technical assistance. This work supported by the Health and Labour Sciences Research Grants (H15-Chemistry-003).

References

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

[2] Law Concerning the Evaluation of Chemical Substances and Regulation of Their Manufacture, etc., Law No. 117, 16 October 1973 as last amended by Law No.49, 28 May 2003.

[3] M.T.D. Cronin, J.S. Jaworska, J.D. Walker, M.H.I. Comber, C.D. Watts, A.P. Worth, Use of QSARs in international decision-making frameworks to predict health effects of chemical substances, Environ. Health Perspect. 111 (2003) 1391–1401.

[4] G.M. Pearl, S. Livingstone-Carr, S.K. Durham, Integration of computational analysis as sentinel tool in toxicological assessments, Curr. Topoics Med. Chem. 1 (2001) 247–255.

[5] A. Hirose, M. Takahashi, M. Kamata, M. Ema, M. Hayashi, Development of genotoxicity predicting QSAR system for registered and existing industrial chemicals in Japan, Toxicol. Appl. Pharmacol. 197 (2004) 358.

[6] N. Greene, P.N. Judson, J.J. Langowski, C.A. Marchant, Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR, SAR QSAR Environ. Res. 10 (1999) 299–314.

[7] H.S. Rosenkranz, A.R. Cunningham, Y.P. Zhang, H.G. Claycamp, O.T. Macina, N.B. Sussman, S.G. Grant, G. Klopman, Development, characterization and application of predictive-toxicology models, SAR QSAR Environ. Res. 10 (1999) 277–298.

[8] M.P. Murray, J.E. Cummins, Mutagenic activity of epoxy embedding reagents employed in electron microscopy, Environ. Mutagen. 1 (1979) 307–313.

[9] N.F. Cariello, J.D. Wilson, B.H. Britt, D.J. Wedd, B. Burlinson, V. Gombar, Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Deductive estimate of risk from existing knowledge. Toxicity prediction by computer assisted technology, Mutagenesis 17 (4) (2002) 321–329.

[10] J.R. Votano, M. Parham, L.H. Hall, L.B. Kier, S. Oloff, A. Tropsha, Q. Xie, W. Tong, Three new consensus QSAR models for the prediction of Ames genotoxicity, Mutagenesis 19 (5) (2004) 365–377.

[11] R.D. Snyder, D.E. Ewing, L.B. Hendry, Evaluation of DNA intercalation potential of pharmaceuticals and other chemicals by cell-based and three-dimensional computational approaches, Environ. Mol. Mutagen. 44 (2) (2004) 163–173.

[12] R.D. Snyder, G.S. Pearl, G. Mandakas, W.N. Choy, F. Goodsaid, I.Y. Rosenblum, Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules, Environ. Mol. Mutagen. 43 (3) (2004) 143–158.

[13] A.C. White, R.A. Mueller, R.H. Gallavan, S. Aaron, A.G. Wilson, A multiple in silico program approach for the prediction of mutagenicity from chemical structure, Mutat. Res. 539 (1–2) (2003) 77–89.