Table 5
Results of GHS classification including re-evaluation on selected chemicals in the Japanese GHS Classification Project (J-GHS).

No.	Chemical Name [CAS] (J-GHS ID)	EU Mut cat.	MAK GCM cat.	GHS GCM cat.	#1 1.1	#1 1.2	#1 1.3	#2 2.1	#2 2.2	#2 2.3	#2 2.4	#3 3.1	#3 3.2	#3 3.3	#3 3.4	#3 3.5	#4 4.1	#4 4.2	#4 4.3	#4 4.4	#5 5.1	#5 5.2	#5 5.3	#5 5.4	#5 5.5	#6 6.1	#6 6.2	#6 6.3	#6 6.4
1	Acrylamide [79-06-1] (0001)	2	2	1B	+	+	+	+, –	+			+, –	+	+, –	+		+	–		+	+	+				+		+, –	–
2	Acrylic acid [79-10-7] (0002)			NC			–					–														+		+, –	–
3	Acrylonitrile [107-13-1] (0003)			2			–	–				–		–	+			–			+	–			–	+	+	+, –	–
	Re-evaluation for J-GHS			NC			–	–				–		–	(+)	–		–			(+), –	–	+, –		–	+	+	+, –	+, –
4	Aniline [62-53-3] (0007)	3		2			Inc							+								+	+	+		+		+	–
5	Antimony hydride (Stibine) [7803-52-3] (0010)	NL	3B*	CNP																									
6	Cadmium oxide [1306-19-0] (0015)	3	3A**	2										–		+, –									+, –				–
7	Vanadium(V) oxide [1314-62-1] (0026)	3	2***	1B			+							+, –						+		–		+, –		+	+	–	+, –
8	Phenol [108-95-2] (0061)	3		1B				+				+, –		+, –						–		+	–	–		–		+, –	–
	Re-evaluation for J-GHS			2				(+)				+, –		+, –						–		+	–	–		+	+	+, –	–
9	Formaldehyde [50-00-0] (0069)		5	2			Inc, –	–				+, –		+, –								+, –		+		+		+	+
10	o-Anisidine [90-04-0] (0083)	3		2										–	+						–		–	–		+		+	–
11	Glycidol [556-52-5] (0098)	3		2								+		+								+				+		+	+
12	Vinyl chloride [75-01-4] (0113)			2			–					+	–	+		+						+		+		+		+, –	+
13	1,4-Dioxane [123-91-1] (0125)			NC										+, –										+, –		–	–	–	–
14	4,4'-Methylenebis(2-chloroaniline) [101-14-4] (0130)			2										+								+	+			–		+	+
15	3,3'-Dichlorobenzidine [91-94-1] (0138)			2								+		+						+		+	+						+
16	Nitrilotriacetic acid [139-13-9] (0170)	NL	–§	1B			–	+				–		–								–				–	+	+, –	–
	Re-evaluation for J-GHS			NC			–	(+)				–		(+), –								–				+, –	+	+, –	–
17	1,3-Dibromopropane [109-64-8] (0539)	NL	NL	CNP																						+			+
18	Ethanol [64-17-5] (0662)		5§§	1B			+, –	+, –				–		+, –								+				–		–	–
	Re-evaluation for J-GHS			NC			(+), –	(+), –			(+)	–		(+), –			(+)					(+), –			(+)	–		–	–
19	Sodium chlorite [7758-19-2] (1109)	NL	NL	2								–		+, –												+			+, –
	Re-evaluation for J-GHS			NC								–		(+), –												+			+, –
20	1-Chloro-2-nitrobenzene [88-73-3] (1184)			2																				+		+, –		–	+, –
	Re-evaluation for J-GHS			CNP																				(+)		+, –		–	+, –

Abbreviation: EU Mut cat.: EU mutagen category by Annex I (ECB, 2008); MAK GCM cat.: MAK germ cell mutagen category by MAK List (DFG, 2007); GHS GCM cat.: GHS germ cell mutagen category; CNP: classification not possible; NC: not classified; NL: not listed; Test results: +: positive; (+): positive in special case, or questionable, non-relevant or non-conclusive positive, –: negative, Inc: inconclusive.
Mutagenicity/genotoxicity tests (some cases included modified methods).
#1 *In vivo* heritable germ cell mutagenicity tests in mammals.
1.1: Mouse specific locus test; 1.2: Mouse heritable translocation test; 1.3: Rodent dominant lethal test.
#2 *In vivo* germ cell mutagenicity tests in mammals.
2.1: Chromosomal aberration test in spermatogonia; 2.2: Micronucleus test in spermatid cells; 2.3: Gene mutation test in germ cells of transgenic rodents; 2.4: Analysis of aneuploidy in sperm cells of exposed people.
#3 *In vivo* somatic cell mutagenicity tests in mammals.
3.1: Chromosome aberration test in bone marrow cells or peripheral lymphocytes; 3.2: Mouse spot test; 3.3: Micronucleus test in hematopoietic cells; 3.4: Gene mutation test in somatic cells of transgenic rodents; 3.5: Metaphase or micronucleus formation analysis of peripheral lymphocytes of exposed people.
#4 *In vivo* germ cell genotoxicity tests in mammals.
4.1: Sister chromatid exchange (SCE) test in spermatogonia; 4.2: Unscheduled DNA synthesis (UDS) test in testicular cells; 4.3: Assays of covalent binding or adduct formation to germ cell DNA; 4.4: Assays of DNA damage in germ cells (comet assay, alkaline elution assay, etc.).
#5 *In vivo* somatic cell genotoxicity tests in mammals.
5.1: UDS test in liver; 5.2: SCE test in bone marrow cells or peripheral lymphocytes; 5.3: Assays of covalent binding or adduct formation to somatic cell DNA; 5.4: Assays of DNA damage in somatic cells (comet assay, alkaline elution assay, etc.); 5.5: SCE analysis of peripheral lymphocytes of exposed people.
#6 *In vitro* mutagenicity tests.
6.1: Chromosomal aberration test in cultured mammalian cells; 6.2: Micronucleus test in cultured mammalian cells; 6.3: Gene mutation test in cultured mammalian cells; 6.4: Reverse mutation test in bacteria (Ames test).
* As antimony [7440-36-0] and its inorganic compounds (except for stibine).
** As cadmium [7440-43-9] and its inorganic compounds (inhalable fraction).
*** As vanadium [7440-62-2] and its inorganic compounds.
§ Nitriloacetic acid and its sodium salt.
§§ Changed from Category 2.

GHS Category 2 was assigned for this chemical based on only one positive result in rat T-cell mutation assay (IPCS, 2002). Acrylonitrile is not classified as a mutagen or GCM in EU Annex I or the MAK List, respectively.

The T-cell mutation assay (column 3.4 in Table 4) is not a standard test and the reliability of the result is questionable, therefore, re-evaluation was performed on this chemical with data used initially and obtained additionally. The information came only from an abstract. Therefore, it was decided that this data should not be used for the classification.

Several data can be obtained from *in vivo* genotoxicity tests with somatic cells. A positive result in liver UDS test using liquid scintillation counting method (column 5.1) (IPCS, 2002; CERI-NITE, 2003; ECB, 2004a) was used initially. In the course of re-evaluation, a negative result was found in liver UDS test using the autoradiograph method (IPCS, 2002; ECB, 2004a). EU Risk Assessment Report (EURAR) discussed that "liquid scintillation counting method is not regarded as the most reliable means of establishing evidence of DNA-repair, preference being given to autoradiographical techniques (ECB, 2004a)." For DNA binding *in vivo* (column 5.3), Concise International Chemical Assessment Document (CICAD) reported inconsistent results (IPCS, 2002). With respect to *in vitro* mutagenicity tests (column 6), EURAR (ECB, 2004a) suggested that "Positive findings *in vitro* are not reliably reflected in the *in vivo* situation, because acrylonitrile or its active metabolites do not reach target tissues *in vivo*, possibly due to the detoxification of the epoxide metabolite cyanoethylene oxide via a glutathione conjugation pathway".

In the dominant lethal test in rats (column 1.3), one negative result is cited in hazard assessment report by Chemicals Evaluation and Research Institute (CERI) and National Institute of Technology and Evaluation (NITE) (CERI-NITE, 2003) and EURAR (ECB, 2004a). Two negative results in mice treated by i.p. or inhalation were found in addition (IARC, 1999a; IPCS, 2002). A negative result in the mouse spermatogonial chromosomal aberration test, two negative results in mouse bone marrow chromosomal aberration test were used initially (column 3.1), and two negative results in rodent chromosomal aberration test reported in IARC (1999a) or CICAD (IPCS, 2002) were cited additionally. In the rodent micronucleus test (column 3.3), two negative results in mice treated by i.p. were used. On the other hand, CICAD (IPCS, 2002) reported negative results in mice and inconclusive results in rats treated by multiple routes from a collaborative study by Morita et al. (1997). A negative result in rat spermatocyte UDS test was reported (IARC, 1999a; ECB, 2004a).

The following conclusion is made in the re-evaluation; acrylonitrile gave clear negative results in the rodent dominant lethal and micronucleus tests in spite of the mutagenic activities reported *in vitro*. Positive results in T-cell mutation cannot be evaluated, and the positive in rat liver UDS test is not regarded as reliable. Therefore, we propose that acrylonitrile should be assigned NC based on the practical decision tree.

Several issues were identified through classification of this chemical: (1) lack of understanding of reliability of test method: T-cell mutation assay is not a standard test; (2) insufficient review of documents: T-cell mutation study is abstract information only, and; (3) shortage of information collection: presence of negative result in rat liver UDS test and non-citation of IARC (1999a).

5.4. Aniline [62-53-3] (J-GHS ID 0007), Cat. 2

Negative results were obtained from bacterial mutagenicity tests. Findings in rodent dominant lethal tests were negative in general, but the final evaluation of the test result is considered to be inconclusive due to slight but statistically significant but slight, toxicologically significant effect at the highest dose (ECB, 2004b). A rodent erythrocyte micronucleus test gave positive results, as did several DNA endpoints in *in vivo* genotoxicity tests with somatic cells. These positive findings *in vivo* were supported by *in vitro* positives with mammalian cells (CERI-NITE, 2004c; ECB, 2004b). GHS Category 2 was applied. EU Annex I gives Category 3 for mutagenicity, but no classification for GCM in the MAK List.

It should be noted that the EU Risk Assessment Report provides a full discussion on the responses in micronucleus tests and the dominant lethal test. Weak positive effects were found in micronucleus tests, which were limited to high doses in the toxic range, and the result of the dominant lethal test is considered to be inconclusive in spite of general negative findings (ECB, 2004b). The conclusion of the report is that the available data of mutagenicity are not sufficient to classify aniline as a Category 2 mutagen in the EU classification, but as a Category 3 mutagen, due to the positive results from several *in vivo* and *in vitro* tests, especially in the bone marrow micronucleus test with rats. Aniline induces methaemoglobinanemia, which might lead erythrocyte degradation, resulting in the induction of micronuclei as a result of increases in cell division to replace lost erythrocytes (Tweats et al., 2007). If the involvement of increases in erythropoiesis after aniline treatment is resolved, the classification will be reconsidered.

5.5. Antimony hydride (Stibine) [7803-52-3] (J-GHS ID 0010), CNP

No data was found in the data source used. "Classification Not Possible (CNP)" was assigned. Antimony hydride (Stibine) is not listed in the EU Annex I. "Antimony and its inorganic compounds" were classified MAK GCM Category 3B, but stibine was excluded as an exception.

5.6. Cadmium oxide [1306-19-0] (J-GHS ID 0015), Cat. 2

A negative result was obtained in a rodent micronucleus test. Conflicting results were reported in cytogenetic and SCE evaluations of peripheral lymphocytes of exposed people (IARC, 1994; ECB, 2003; NTP, 2005a). Based on the positive findings in somatic cell mutagenicity in humans, GHS Category 2 was assigned. EU classification of cadmium oxide is mutagenicity Category 3 in Annex I, but this chemical was classified MAK GCM Category 3A as cadmium and its inorganic compounds (inhalable fraction) in the List.

5.7. Vanadium (V) oxide [1314-62-1] (J-GHS ID 0026), 1B

A rodent dominant lethal test was positive though conflicting results were obtained in rodent micronucleus tests. Positive results were obtained from *in vivo* germ and somatic cell genotoxicity tests and also *in vitro* mutagenicity tests (IPCS, 2001; NTP, 2005d). Based on the positive finding in the dominant lethal test, GHS Category 1B was given. EU classification gave mutagenicity Category 3 in Annex I. On the other hand, this chemical was classified MAK GCM Category 2 as vanadium and its inorganic compounds in the List.

5.8. Phenol [108-95-2] (J-GHS ID 0061), Changed to Cat. 2 from Cat. 1B

Positive results were obtained in cytogenetic analysis with mouse spermatogonia or spermatocytes. Negative and positive results were obtained in rodent chromosomal aberration or micronucleus tests with somatic cells. An assay of DNA damage in germ cells gave a negative result. *In vivo* genotoxicity tests with somatic cells and *in vitro* mutagenicity tests showed negative results generally (CERI-NITE, 2005a; NTP, 2005c). Based on the positive results in cytogenetic analysis in germ cells *in vivo*, GHS Category 1B was assigned. Phenol is classified as a mutagen, Cat-

egory 3 in the EU Annex I, but not classified as GCM in the MAK List.

Conflicting results obtained in rodent cytogenetic analysis with somatic cells raised questions about positive findings in chromosomal aberration tests with mouse germ cells. Therefore, re-evaluation was performed on this chemical. Positive findings in chromosomal aberration tests with mouse germ cells (Bulsiewicz, 1977) (column 2.1), cited from CERI-NITE (2005a), gave support to the classification of GCM. However, IARC (1999a) and EURAR (ECB, 2006b) have not cited these data. On the other hand, US EPA cited the data in the toxicological review document in support of summary information on the IRIS (EPA, 2002a), and stated a comment of "inconsistencies in reporting" to the data (EPA, 2002b), including that one of the inconsistencies is found in dosing concentration. The germ cell cytogenetic analysis used was not a standard test, and it had no statistical analysis.

Both positive and negative results were reported in rodent cytogenetic analysis with somatic cells (column 3.1 and 3.3) (IARC, 1999a; EPA, 2002b; CERI-NITE, 2005a; NTP, 2005c; ECB, 2006b). A mouse chromosomal aberration test gave a positive result up to an intraperitoneal injection (i.p.) dose of 300 mg/kg, but the test used a non-standard protocol (NTP, 2005c). In contrast, negative results were obtained in the rat at doses up to 180 mg/kg by i.p. and up to 510 mg/kg by oral administration (p.o.) (CERI-NITE, 2005a). Six mouse micronucleus tests showed positive results at doses up to 300 mg/kg by i.p. or p.o. dosing including three treatments, while three tests showed negative results at doses up to 250 mg/kg by i.p. or p.o. EURAR discussed that "Phenol should be regarded as a somatic cell mutagen, and that the high dose positive results in micronucleus tests might be due to phenol-induced hypothermia (ECB, 2006b)." It is also reported that a single i.p. dose of phenol to mice at 300 mg/kg produced a significant and prolonged hypothermia and a significant increase in micronuclei (Spencer et al., 2007).

A DNA damage test using alkaline elution in testicular cells was negative in rats treated by i.p. injection up to 79 mg/kg (column 4.4) (EPA, 2002b; CERI-NITE, 2005a; ECB, 2006b). With respect to *in vivo* genotoxicity tests with somatic cells, a positive result was reported in sister chromatid exchange analysis in mice treated by the i.p. route at doses up to 300 mg/kg (column 5.2,) (NTP, 2005c). However, it used non-standard protocol. Other endpoints including DNA adduct formation (column 5.3) or DNA damage induction (column 5.4) were negative in rats (p.o. dosing of 75 mg/kg/day for days) or mice (i.p. dosing of 75 mg/kg), respectively (IARC, 1999a; CERI-NITE, 2005a; ECB, 2006b). A negative result from an *in vitro* chromosomal aberration test (column 6.1) (CERI-NITE, 2005a) was cited initially, but phenol is recognized as positive in chromosomal aberration tests *in vitro* (EPA, 2002b; Kirkland et al., 2005; ECB, 2006b). In addition, several positive results from *in vitro* micronucleus tests were also reported (column 6.2) (IARC, 1999a; EPA, 2002b; Kirkland et al., 2005; ECB, 2006b).

The following conclusion is made in the re-evaluation. The positive result in the germ cell cytogenetic analysis is not considered as sufficient evidence due to this being a non-standard test without statistical analysis. Though phenol induces micronuclei in rodent somatic cells, recent published data demonstrated that the induction of micronuclei was exclusively associated with phenol-induced hypothermia (Spencer et al., 2007), suggesting that the increase in micronuclei may not be a result of any intrinsic direct genotoxic effects of phenol. On the other hand, Tweats et al. (2007) pointed out that the response is somewhat higher than with other compounds that induce hypothermia and it would be informative to ascertain if this response can be reversed by maintaining the core temperature of the treatment. The mechanism by which hypothermia induces micronuclei is not clearly established, but may involve disturbance of the mitotic spindle. Therefore, we propose that phenol should be classified as Category 2 in the practical decision tree in case of positive in somatic mutagenicity but negative in germ cell genotoxicity *in vivo*. This would need re-evaluation if further information becomes available.

Practical issues identified are (1) insufficient review; no critical review of a positive finding in mouse germ cells, and (2) shortage of information collection; non-citation of IARC (1999a) and EPA (2002b). It is not possible to cite recent publications of EURAR (ECB, 2006b), Spencer et al. (2007), and Tweats et al. (2007) at that time of the project operation.

5.9. Formaldehyde [50-00-0] (J-GHS ID 0069), Cat. 2

Negative or inconclusive results were obtained from rodent dominant lethal tests or a spermatogonial chromosomal aberration test. Cytogenetic evaluation as measured by the induction of micronuclei in bone marrow or peripheral blood cells showed negative results in mice treated by p.o. dosing or intravenous injection (i.v.) (Morita et al., 1997). On the other hand, positive findings were observed in a chromosome aberration test with pulmonary lavage cells and a micronucleus test with gastrointestinal or nasal cells *in vivo*. Positive results were also obtained from *in vivo* genotoxicity tests and *in vitro* mutagenicity tests (IPCS, 1989; OECD, 2004b; CERI-NITE, 2005c). Based on the positive findings in somatic cells *in vivo*, GHS Category 2 was assigned. Formaldehyde is not categorized as a mutagen in EU Annex I, but as Category 5 in MAK GCM classification in the list. MAK GCM category 5 is a relative new category (Adler et al., 2000; DFG, 2007) that is defined as "germ cell mutagens or suspected substances (according to the definition of Category 3A and 3B), the potency of which is considered to be so low that their contribution to genetic risk for man is expected not to be significant". At present, acetaldehyde, formaldehyde and ethanol are in Category 5 in the MAK List. A major limitation of many studies is the well established cytotoxic effects of formaldehyde.

It should be noticed that there is evidence from recent papers and assessment on formaldehyde that the effects of it are due to DNA–protein adducts, which are shown to have a threshold (Schmid and Speit, 2006; Speit et al., 2007; UKCOM, 2007). The conclusions from UK Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment are as follows: "There was no convincing evidence from *in vivo* mutagenicity studies in experimental animals and from biomonitoring studies of genotoxicity in workers exposed to formaldehyde for a direct *in vivo* systemic mutagenic effect of inhaled formaldehyde. A secondary mechanism might be involved in the genotoxic effects documented in peripheral blood lymphocytes in the biomonitoring studies reviewed. For occupational and environmental exposure to formaldehyde, the pattern of metabolism and distribution of formaldehyde indicate that a threshold for *in vivo* systemic mutagenicity is likely (UKCOM, 2007)."

We agree with the conclusions by UKCOM. Systemic exposure after inhalational exposure of formaldehyde is negligible and thus would not present a germ cell hazard. However, local genotoxic effects in somatic cells should be taken consider in hazard communication.

5.10. o-Anisidine [90-04-0] (J-GHS ID 0083), Cat. 2

Negative results were obtained from a rodent micronucleus test and several *in vivo* genotoxicity tests with somatic cells. On the other hand, a transgenic mouse mutation test showed a positive result. The positive finding was supported by the positive results from *in vitro* gene mutation tests (DFG, 1998; IARC, 1999b; ECB, 2002a). Based on the positive result from the transgenic gene mutation model assay, GHS Category 2 was assigned. *o*-Anisidine

is categorized as a Category 3 mutagen in the EU Annex I, but not classified as GCM in the MAK List.

5.11. Glycidol [556-52-5] (J-GHS ID 0098), Cat. 2

Positive results in a rodent cytogenetic evaluation were supported by the positive findings *in vitro*. There were no data on mutagenicity or genotoxicity in germ cells (ACGIH, 2001; Bingham et al., 2001b; CERI, 2002a; DFG, 2003b). GHS Category 2 was assigned. Glycidol is categorized as a Category 3 mutagen in the EU Annex I, but not classified as GCM in the MAK List.

5.12. Vinyl chloride [75-01-4] (J-GHS ID 0113), Cat. 2

A negative result was obtained from a rodent dominant lethal test. However, positive results were obtained from *in vivo* mutagenicity and genotoxicity tests with somatic cells including a human epidemiological study, and from *in vitro* mutagenicity tests (ECETOC, 1998; ATSDR, 2004; CERI-NITE, 2004d). Based on the positive results from *in vivo* tests with somatic cells, GHS Category 2 was assigned. Vinyl chloride is not categorized as a mutagen or GCM in the EU Annex I or the MAK List, respectively.

5.13. 1,4-Dioxane [123-91-1] (J-GHS ID 0125), NC

Almost all tests conducted were negative. One positive result was obtained in a mouse bone marrow micronucleus test, but it was not confirmed by other reports. Conflicting results were obtained in DNA damage tests in rodents, and where positive findings were observed these were only at high doses. All *in vitro* mutagenicity tests were negative (CERI-NITE, 2004a). Based on the negative results in *in vivo* and *in vitro* tests, NC was assigned. 1,4-Dioxane is not categorized as a mutagen or GCM in the EU Annex I or the MAK List, respectively. Several evaluation documents have been issued on 1,4-dioxane (NICNAS, 1998; ECB, 2002b; DFG, 2003a; ATSDR, 2006). These documents were not used in the J-GHS, because only one review document written in Japanese was used by the Classifier. Negative results in a rodent liver UDS assay and a positive result in rodent liver micronucleus test were also described in the EU (ECB, 2002b) and US documents (ATSDR, 2006). However, the positive result at high doses in the liver micronucleus test is non-relevant to humans.

A practical issue identified is a shortage of information collection, though it gave no influence of the classification result.

5.14. 4,4′-Methylenebis(2-chloroaniline) [101-14-4] (J-GHS ID 0130), Cat. 2

Positive results were obtained from rodent micronucleus tests and *in vivo* genotoxicity tests with somatic cells. There were no data on mutagenicity or genotoxicity in germ cells (IARC, 1993; CERI-NITE, 2005b; NTP, 2005b). Based on the positive findings from *in vivo* tests with somatic cells, GHS Category 2 was assigned. 4,4′-Methylenebis(2-chloroaniline) is not categorized as a mutagen or GCM in the EU Annex I or the MAK List, respectively.

5.15. 3,3′-Dichlorobenzidine [91-94-1] (J-GHS ID 0138), Cat. 2

Positive results were obtained from rodent cytogenetic evaluations and *in vivo* genotoxicity tests with somatic cells. There was no data on mutagenicity or genotoxicity in germ cells (IARC, 1982; DFG, 1992b; ATSDR, 1998; IPCS, 1998; CERI, 2002b). Based on the positive findings from *in vivo* tests with somatic cells, GHS Category 2 was assigned. 3,3′-Dichlorobenzidine is not categorized as a mutagen or GCM in the EU Annex I or the MAK List, respectively.

5.16. Nitrilotriacetic acid [139-13-9] (J-GHS ID 0170), Changed to NC from Cat. 1B

A dominant lethal test, chromosomal aberration test, micronucleus test and SCE analysis with rodent cells all gave negative results. However, aneuploidy was detected in mouse spermatocytes with nitrilotriacetic acid trisodium salt (CAS No. 5064-31-1). Both negative and positive results were obtained in *in vitro* mutagenicity tests in which the trisodium salt was mainly used (IARC, 1999b; CERI, 2002c). Based on the positive finding for germ cell aneuploidy, GHS Category 1B was assigned. Nitrilotriacetic acid and its trisodium salt are not listed in the EU Annex I and the MAK List.

Due to the positive finding in germ cells, despite the negatives in somatic cells *in vivo*, re-evaluation was performed on these chemical using additional references. Mouse dominant lethal tests gave negative results (column 1.3), but nitrilotriacetic acid trisodium salt induced meiotic aneuploidy (hyperhaploidy) in mouse spermatocytes (column 2.1) (Costa et al., 1988). This result was supported by positive finding in a rat kidney micronucleus test (column 3.3) (Robbiano et al., 1999). On the other hand, a negative result was reported in a mouse chromosomal aberration test (aneuploidy, column 3.1), mouse bone marrow micronucleus test, and mouse SCE test (column 5.2) (IARC, 1990, 1999b). *In vitro* chromosomal aberration tests showed negative or positive results in CHO cells and human peripheral lymphocytes or in rat kangaroo kidney cells, respectively (column 6.1) (IARC, 1999b). *In vitro* micronucleus tests gave a positive result in hamster CL-1 cells and primary kidney cells from rats and humans (column 6.2) (IARC, 1999b; Robbiano et al., 1999).

The following conclusion is made in the re-evaluation. Standard *in vivo* mutagenicity tests including dominant lethal test, chromosomal aberration test, and micronucleus tests showed negative results. A positive result was obtained in a test which measured mouse germ cell aneuploidy, and this is supported by a positive finding in an *in vivo* rat kidney micronucleus test. On the other hand, it is not supported by the mouse chromosomal aberration test (negative in aneuploidy) and mouse bone marrow micronucleus test. Some *in vivo* tests used the trisodium salt, which might have different toxicokinetics, and the *in vivo* rat kidney micronucleus test is not a standard test. Therefore, a positive result for mouse germ cell aneuploidy is not considered to be sufficient evidence, and also there is no clear evidence of *in vivo* somatic cell mutagenicity. Overall, we propose that nitrilotriacetic acid or its trisodium salt should be assigned NC in GHS GCM.

Practical issues complicate the classification identified are (1) insufficient review; no critical review of the positive finding for mouse germ cell aneuploidy, and (2) shortage of information collection; non-citation of IARC (1990) and Robbiano et al. (1999).

5.17. 1,3-Dibromopropane [109-64-8] (J-GHS ID 0539), CNP

Positive results were obtained from both an *in vitro* chromosomal aberration test and an Ames test (JECDB, 2006a,b). No information was obtained in *in vivo* mutagenicity or genotoxicity tests. Classifier requested experts to judge classification of this chemical based on the practical decision tree (Fig. 2). Experts reviewed these data and information on other dihaloalkanes or dihaloalkenes including 1,2-dibromoethane, 1,2-dichloropropane, 1,2-dichloroethane, 1,3-dichloropropane and 1,3-dichloropropene (IARC, 1999a; HSDB, 2002). These halogenated compounds are usually

positive in *in vitro* mutagenicity tests, but are negative in *in vivo* micronucleus tests, mouse dominant lethal tests, and the spermatocyte UDS assay. There is no clear evidence or suggestion that 1,3-dibromopropane should be classified as Category 2. Therefore, the experts concluded that 1,3-dibromopropane should be assigned as CNP. 1,3-Dibromopropane is not listed in the EU Annex I, and not categorized as a GCM in the MAK List.

5.18. Ethanol [64-17-5] (J-GHS ID 0662), Changed to NC from Cat. 1B

Conflicting results were obtained in *in vivo* heritable and germ cell mutagenicity tests with rodents. Some rodent dominant lethal tests showed positive results as well as positive results in chromosomal aberration (aneuploidy) tests in mouse spermatocytes, and negative results were also obtained in both tests in different studies. Chromosomal aberration tests in mammalian bone marrow cells showed negative results. Conflicting results were obtained in *in vivo* micronucleus tests with rodent somatic cells, and a positive result was obtained in an SCE test in mouse fetal hepatocytes. All *in vitro* mutagenicity tests were negative (IARC, 1988; DFG, 1999; Bingham et al., 2001a). Based on positive results in germ cells including dominant lethal tests, GHS Category 1B was assigned. Ethanol is not categorized as mutagen in the EU Annex I, but as GCM Category 5 in the MAK List.

Due to the mixture of positive and negative findings reported in mutagenicity or genotoxicity tests from *in vivo* germ or somatic mutagenicity tests and from all of the *in vitro* mutagenicity tests evaluated, re-evaluation was performed.

Conflicting results were obtained in rodent dominant lethal tests (column 1.3). Waters et al. also summarized positive results of dominant lethal tests both in rats and mice (Waters et al., 1994). However, the following conclusions for dominant lethal effect of ethanol were drawn in the review documents: (1) dominant lethal mutations were observed in mice given bolus doses of highly concentrated ethanol solutions. For rats, such findings have only been reported after very high doses which produced marked symptoms of systemic toxicity (DFG, 1999); (2) using a weight of evidence approach, it was concluded that ethanol does not induce dominant lethality in assays using standard regulatory approved methodologies (Phillips and Jenkinson, 2001; OECD, 2004a), and; (3) the majority of studies showing positive results can be criticized on the grounds of inadequate numbers of animals or on the methods used to score, evaluate or distinguish between early and late fetal deaths (OECD, 2004a).

In the chromosomal aberration test with germ cells including spermatocytes or spermatogonia, conflicting results were obtained (column 2.1). For aneugenic effects of ethanol in germ cells, the following conclusions were drawn in the review documents: (1) mutagenic potential seems to be weak, is limited to the induction of aneuploidy and could be demonstrated to date only with very high doses of at least 5 g/kg, which produced systemic toxicity, and only in mice (DFG, 1999); (2) findings could be due to an aneugenic effect during meiosis but convincing evidence is lacking and alternative non-genotoxic mechanisms are feasible (Phillips and Jenkinson, 2001), and; (3) many variables may affect the outcomes (Pacchierotti and Ranaldi, 2006).

The cytogenetic analysis with somatic cells showed negative results, however, positive findings were also reported (column 3.1 and 3.3). Significant increases of micronuclei in bone marrow cells from rats given a diet containing ethanol might be due to the increase of cell division as is induced by erythropoietin (Phillips and Jenkinson, 2001). For micronucleus induction by ethanol in bone marrow cells, the review documents concluded that there is no convincing evidence that ethanol induces micronuclei in the bone marrow of rodents (Phillips and Jenkinson, 2001; OECD, 2004a).

The majority of studies on sister chromatid exchange (SCE) induction *in vivo* were positive, although the effects have generally been small (column 4.1 and 5.2) (IARC, 1988; DFG, 1999; Phillips and Jenkinson, 2001). Negative results have also reported for SCE induction (column 5.2). For SCE induction by ethanol, Phillips and Jenkinson pointed out the possibility of disturbance of the metabolism, nutritional or hormonal status of the animal treated at high dose levels of ethanol (Phillips and Jenkinson, 2001).

A human monitoring study found an association between alcohol drinking and aneuploidy in sperm from young men (column 2.4) (Robbins et al., 1997), which has been reviewed by Phillips and Jenkinson (2001) and Pacchierotti and Ranaldi (2006). Several studies for SCE induction in humans (column 5.5) have suggested that alcoholics have higher SCE frequencies in their lymphocytes than non-alcoholics. However, these studies frequently failed to control for smoking and other confounding factors which may have influenced the results (IARC, 1988).

All standard *in vitro* mutagenicity tests showed negative results (column 6) (IARC, 1988; DFG, 1999; Phillips and Jenkinson, 2001; OECD, 2004a).

A recent IARC monograph mentioned that numerous reports have shown that human alcoholics have a higher frequency of chromosomal aberrations, SCE and micronuclei in their peripheral lymphocytes. The data from studies in animals suggest that ethanol causes DNA damage in target tissues (IARC, 2007). However, the following overall conclusions are suitable for the purpose of classification of industrial chemicals: (1) it is concluded that there is no significant evidence that ethanol is a genotoxic hazard according to the criteria normally applied for the purpose of classification and labeling of industrial chemicals (Phillips and Jenkinson, 2001), and; (2) the balance of evidence is that ethanol is not genotoxic. There is very little evidence to suggest that ethanol is genotoxic in somatic cells and it may have a very limited capacity to induce genetic changes *in vivo* but under very specific circumstances and at very high doses achievable in humans only by deliberate oral ingestion (OECD, 2004a). Therefore, we propose that ethanol should be assigned NC in GHS GCM. Deliberate high oral intake of ethanol in alcoholic beverages over a long period might be genotoxic; however, GHS does not require foods or beverages to be labeled to indicate the presence of hazardous materials (UN, 2007).

Practical issues identified are (1) shortage of information collection; Non-citation of Phillips and Jenkinson (2001) and Screening Information Data Set (SIDS) (OECD, 2004a), (2) discrepancy of international or national review evaluation depend on year; for example, change from MAK GCM Category 2 (DFG, 1999) to Category 5 (Adler et al., 2000; DFG, 2007), and (3) necessity of careful evaluation of single positive findings in some tests, especially with non-standard methods.

5.19. Sodium chlorite [7758-19-2] (J-GHS ID 1109), Changed to NC from Cat. 2

Conflicting results were obtained in rodent cytogenetic evaluations. A mouse micronucleus test by intraperitoneal injection gave a positive result, but a negative result was obtained in a test using gavage dosing in a chromosomal aberration test. Positive results were obtained in *in vitro* mutagenicity tests. There were no data on mutagenicity or genotoxicity in germ cells (IARC, 1991; ECB, 2000; EPA, 2000; RTECS, 2003). Based on the positive result in the micronucleus test by i.p. dosing, GHS Category 2 was assigned. Sodium chlorite is not listed in the EU Annex I or the MAK List.

Since the route of administration yields different results, a re-evaluation was performed. Mouse bone marrow chromosomal aberration tests and mouse bone marrow micronucleus test were negative by single or multiple oral administrations (column 3.1 and 3.3) (Meier et al., 1985; Hayashi et al., 1988; IARC, 1991;

EPA, 2000). On the other hand, single i.p. dosing induced micronuclei in mice (Hayashi et al., 1988; IARC, 1991; EPA, 2000; RTECS, 2003). An IARC review noted the importance of the chemical properties of sodium chlorite: in aqueous acid solutions, chlorite forms chlorous acid, which rapidly decomposes to chlorine dioxide, chlorate and chloride (IARC, 1991). Sodium chlorite is not considered to be a somatic cell mutagen at relevant exposures in humans, based on the negative result in micronucleus test by p.o. dosing, in which sodium chlorite decomposes in the stomach. Therefore, we propose that sodium chlorite should be assigned NC in GHS GCM.

Practical issues identified are (1) insufficient review; no consideration of chemical properties, and (2) discrepancy of weighting of evidence; positive when dosed i.p., but negative *in vivo* when dosed by the p.o. route.

5.20. 1-Chloro-2-nitrobenzene [88-73-3] (J-GHS ID 1184), Changed to CNP from Cat. 2

Data from one *in vivo* genotoxicity test with rodents is available. DNA single strand breaks in liver and kidneys were observed in mice (DFG, 1992a; IARC, 1996; CERI, 1999; OECD, 2001). Conflicting results were obtained in chromosomal aberration tests with cultured mammalian cells and in bacterial reverse mutation tests. In this case (i.e., positive in *in vivo* germ or somatic cell genotoxicity test which is supported by positive in *in vitro* mutagenicity test), an expert review should have been requested originally (see Fig. 2). However, GHS Category 2 was assigned without expert judgment. 1-Chloro-2-nitrobenzene is not categorized as mutagen or GCM in the EU Annex I or the MAK List, respectively.

Conflicting results were shown in *in vitro* mutagenicity tests. In addition, a single positive finding in *in vivo* genotoxicity test should be viewed with care. Also, no expert review was conducted in this case. Therefore, a follow-up review by experts was conducted. The positive finding in DNA damage as DNA single strand breaks was identified by the alkaline elution technique after i.p. dosing to male mice (column 5.4). A dose–effect relationship was not obtained, and the route of administration should be taken into consideration (OECD, 2001). An *in vitro* mammalian cell gene mutation test was negative using the V79 *hprt* assay (column 6.3) (DFG, 1992a; OECD, 2001). Positive effects in a cytogenetic test *in vitro* and a bacterial reverse mutation test were weak (column 6.1 and 6.4) (DFG, 1992a; OECD, 2001). There is no definitive evidence of *in vivo* somatic cell mutagenicity by 1-chloro-2-nitrobenzene. Therefore, we propose that 1-chloro-2-nitrobenzene should be assigned as CNP.

6. Issues identified in classification

Several issues in classification are revealed in the review process of acrylonitrile, phenol, nitriloacetic acid, ethanol, sodium chlorite and 1-chloro-2-nitrobenzene.

6.1. Sources of information on mutagenicity or genotoxicity of chemicals

The GHS is based on currently available information on toxicity of chemicals and it does not require any additional testing to classify a chemical substance. Categorization is based on the criteria for assessing classification and on the existing/available test data/information. Therefore, it is important to know where to find the information necessary for classification and, more importantly, how to correctly interpret these data. Several types of information sources are available. These include review documents, peer-reviewed papers, industry based reports, abstracts, or databanks, etc. The most reliable source is international or national review documents in terms of the quality, availability and suitability of information that has to be used in decision making. Peer-reviewed papers and industry based reports have high quality and suitability, but low availability. Databanks have high availability, but low quality. Abstracts should not be used for classification without any supportive information. One of the major factors of the different classifications was the different sources used. These resulting classifications may differ to a significant degree, leading to varying hazard communication. The age of the data differ among sources. Newer information will be available from more recent documents, and this information could result in changed assessment of chemicals. Therefore, the timeliness of data is an important consideration as previous classifications may be revised when new data becomes available. The use of data in one major information source (Table 2) led to unsuitable classification in the J-GHS. It is important that all available information (e.g., review documents) should be collected and used in a complementary fashion. It is noted that the other information sources, e.g., USEPA GENE-TOX database (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX) and UKCOM statements (http://www.iacom.org.uk/statements/index.htm) were not used in the J-GHS (Table 2). These sources are also useful for GCM classification. Classification based on old or limited information will possess lower reliability. The evaluation of test results in each information source should be checked with multiple sources of information, if available. Original peer-reviewed papers are the best source for assessing difficult and comprehensive test results: these should be included in information collection, if possible. Different classification results from different information source sets have also been noted by The United Nations Institute for Training and Research (UNITAR, 2008b). It might be necessary to develop an internationally-constructed and maintained information database for general Classifiers.

If mutagenicity/genotoxicity data are not available in the list of information sources, Classifiers should search original peer-reviewed papers using PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), for example. In the case of no data in any available information sources, the substance will be assigned as "No data to make classification [as Classification not possible]".

6.2. Expertise of Classifiers

The GHS is originally designed as a self-classification system. The hazard classification process under the GHS is highly technical in nature, and requires a certain background and level of expertise of Classifiers to perform it accurately. If a Classifier lacks understanding of the GHS classification criteria, the effort should be repeated or reviewed carefully before communicating hazard or risk. Many mutagenicity or genotoxicity tests exist today; and many test results have been reported. Research oriented and non-validated studies are included. Classifiers should understand well the classification criteria and suitable materials for classification. A summary or abstract may omit important information or discussion. Therefore, Classifiers should fully study review documents, text, tables, and/or figures, and not only the summary or abstract. Evaluation and interpretation of the test results sometimes differ between the authors of original papers and of review documents. In addition, Classifiers should note that classification will be conducted based on hazard identification and assessment, not on risk characterization for humans. The people who have the scientific knowledge, experimental skill and expertise in toxicology (preferably genotoxicity) should become Classifiers. A deep knowledge of standard mutagenicity testing protocols is also required.

6.3. Expert judgment on data quality and weight of evidence

The GHS criteria for determining health hazards do not depend entirely on test methods. However, the methods used should sci-

entifically sound, validated and accepted according to international scientific standards (e.g., peer-review, authorization by international bodies). Though not all endpoints of the existing substances have been investigated, there is a huge data base of published literature and of files in research institutes and in industry. Not all these data have been developed under standardized test methods or according to the requirements of Good Laboratory Practice (GLP), i.e. they are not equally adequate, valid and relevant. However, it may be difficult for those who are not genetic toxicologists to evaluate the quality of test data. There are a variety of types of test methods and multiple test results in mutagenicity or genotoxicity tests. The determination of the quality of test data is a critical point for the classification of GCM. Therefore the evaluation of data quality frequently has to be done by an expert. Expert judgment plays an important role in making weight of evidence determinations in interpreting data for hazard classification of substances. When multiple data for one endpoint exist, the so-called "total weight of evidence approach" must be applied. GHS defines weight of evidence as follows: all available information bearing on the determination of toxicity is considered together, including the results of valid *in vitro* tests, relevant animal data, and human experience such as epidemiological and clinical studies and well-documented case reports and observations. Both positive and negative test results are assembled together in the weight of evidence determination. However, a single positive study performed according to good scientific principles and with statistically and biologically significant positive results may justify classification (UN, 2007). Genetic toxicology experts must consider all available data (both positive and negative), weigh it with respect to validity, and finally reach a conclusion. Useful data/information can come from different sources, e.g. from human experience, from experimental investigations in animals, from *in vitro* tests, or from similar substances (UN, 2007; UNITAR, 2008a).

Recent analysis demonstrates that *in vitro* mammalian cell tests have high sensitivity to carcinogens (i.e., above 65%), while showing very low specificity (i.e., below 45%) to non-carcinogens that results in false positive findings (Kirkland et al., 2005). Species-specific differences in metabolism are known in animals used in *in vivo* tests. Different mutagenic or genotoxic responses might be obtained. Therefore, it is important to understand the mechanism and/or mode of action of mutagenicity, and to use weight of evidence approaches for assessing the results.

Any discrepancy in classification will be based on the different weighting evidence used from expert to expert. Also, scientific progress will affect expert judgments. Styrene provides an example of the impact of these factors. A draft EU Risk Assessment Report on styrene has been published in November 2007 (ECB, 2007a). This report contains in depth discussions on whether styrene can be classified as a Category 3 mutagen in the EU classification scheme (corresponding to Category 2 GCM in GHS) (ECB, 2006a,c, 2007b). Exhaustive data collection and review were performed. The conclusion is "based on standard regulatory tests, there is no convincing evidence that styrene possesses significant mutagenic/clastogenic potential *in vivo* from the available data in experimental animals". Therefore, the EU classification of styrene as a mutagen Category 3 is not justified. In the J-GHS, styrene was classified as a Category 2 GCM based on the positive findings both *in vivo* micronucleus tests and a human biomonitoring study on micronucleated peripheral lymphocytes. At the time of the J-GHS effort, the draft EU Risk Assessment Report had not been published. We agree with the conclusion in the EU draft report for the purpose of classification and labeling at present time. Epidemiological studies will be needed for further discussion on the effect of styrene to industrial workers. The case of styrene reveals that harmonization of expert judgment is not easy, and is not static.

Recently, many discussion papers on data evaluation by weight of evidence and mode of action approaches have been published (Butterworth, 2006; Hoffmann and Hartung, 2006; Kirkland et al., 2007; Thybaud et al., 2007a,b). Determination of a genotoxic mode of action depends on mutagenicity and/or genotoxicity tests with bacteria, cell cultures, and whole animals. In addition, differences in data quality for published data are dramatic. A weight of evidence approach is essential to judge whether the chemical should be classified as GCM (Butterworth, 2006). Harmonization or consensus of weighting of mutagenicity or genotoxicity tests is needed for non-experts in GCM classification.

6.4. Decision logic and practical decision tree

The practical decision tree (Fig. 2) has been prepared by combining the decision logic of Fig. 1 with GHS classification criteria of Table 1 for germ cell mutagenicity. The tree flows from upstream (i.e., evidence of heritable germ cell mutagenicity) to downstream (i.e., evidence of germ cell mutagenicity *in vivo*, somatic cell mutagenicity *in vivo*, and then mutagenicity *in vitro*). The decision tree is simple and clear when definitive test result(s) exist. However, the results of heritable germ cell mutagenicity tests including the dominant lethal test are emphasized in the decision tree. The positive findings of a dominant lethal test had an impact in the classification of ethanol despite a lack of clear supportive evidence of mutagenicity in *in vivo* and *in vitro* tests (see Section 5.18). In addition, positive findings are given more weight than negative ones for decision making in the tree. Some of the results of classification in J-GHS were overestimated. A complete data set of mutagenicity or genotoxicity tests including *in vitro* tests is not usually available, and the results are sometimes inconsistent. *In vitro* tests employ a metabolic activation system (e.g., rat liver S9) to mimic *in vivo* situation. However, the S9 might be able to produce a metabolite that is not produced in humans. In such case, a positive response in *in vitro* tests in the presence of S9 would be irrelevant for humans. In addition rat liver S9 may not be able to produce a metabolite(s) that is formed in humans. If this is known, genotoxicity data should be sought for such metabolites, in their own right. Ideally tests should be supported by positive controls similar in structure to the test compounds in question. For evaluation of *in vivo* tests, the dose levels, route of administration, and target tissue exposure should be considered. As positive responses as a result of changes in core body temperature and increases in erythropoiesis following prior toxicity to erythroblasts are known in rodent micronucleus tests (Tweats et al., 2007), the consideration to such secondary effects is also needed, as these effects do not necessarily indicate genotoxicity by the test chemical under normal exposure conditions. Therefore, careful evaluation of the test results (negative or positive) is important (see Section 5.20). Though the decision tree is useful for the GCM classification a "total weight of evidence approach" using all available data is required for classification. Expert judgment is required in some cases in the practical decision tree (Fig. 2).

7. Conclusions

The usefulness of various information sources (Table 2), the examples of different mutagenicity or genotoxicity tests (Table 3), the practical decision tree (Fig. 2), and the definition of GCM (Table 4) for GHS classification have been demonstrated. GHS criteria for germ cell mutagenicity have been adopted in worldwide, and are becoming standard for hazard classification. In Japan, the GHS system has been employed in the Amended Industrial Safety and Health Law for labeling of hazardous chemicals (enforcement date, December 1, 2006). In the EU, the European Commission adopted the "Proposal for Regulation of the European Parliament

and of the Council on classification, labeling and packaging of substances and mixtures, and amending Directive 67/548/EEC and Regulation (EC) No 1907/2006". The proposed Regulation would align the EU system of classification, labeling and packaging substances and mixtures with the GHS. The preparation of detailed guidance for the application of the GHS criteria is under development for the REACH framework (the Registration, Evaluation and Authorisation of Chemicals, a new regulatory approach for chemicals in the EU) (ECHA, 2008a,b,c). After implementation of GHS in each country, the classification of chemicals will be conducted by (hopefully) experts in classification within chemical suppliers, i.e., manufacturers or importers. Illustrations, practical guide and supporting explanations for classification of GCM are helpful for them in order to classify chemicals using scientifically principles. Classification should be performed by Classifiers with high expertise using high quality information sources. It is clear that suitable classification depends on the weight of evidence and reliability of the data. Genetic toxicologists as experts should consider data quality and reliability, and critically review several authoritative documents including original articles to support the classification of chemicals. Finally, it is noted that the results of GHS classification are not inflexible; they will be revised by the consideration of new information including new test data and/or by the elucidation of the mechanism or mode of action of the chemical.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgment

This work was supported in part by the Health and Labor Sciences Research Grants (H18-Chemistry-General-009 and H21-Chemistry-General-004).

References

ACGIH (American Conference of Governmental Industrial Hygienists), 2001. Documentation of the threshold limit values for chemical substances, Glycidol.

Adler, I.D., Anrae, U., Kreis, P., Neumann, H.-G., Thier, R., Wild, D., 2000. Recommendation for the categorization of germ cell mutagens. Int. Arch. Occup. Environ. Health 73, 428–432.

ATSDR (US Agency for Toxic Substances and Disease Registry), 1998. Toxicological profile, 3,3′-Dichlorobenzidine. Available from: <(http://www.atsdr.cdc.gov/toxprofiles/tp108.pdf> (accessed 20.4.2009).

ATSDR, 2004. Toxicological profile, Vinyl chloride, draft. Available from: <http://www.atsdr.cdc.gov/toxprofiles/tp20.pdf> (2007, final version, accessed 20.4.2009).

ATSDR, 2006. Toxicological profile, 1,4-Dioxane. Available from: <http://www.atsdr.cdc.gov/toxprofiles/tp187.pdf> (2007, draft for public comment, accessed 20.4.2009).

Bingham, E., Cohrssen, B., Powell, C.H. (Eds.), 2001a. Patty's Toxicology, fifth ed., vol. 6, Ethanol. John Wiley and Sons, New York, pp. 382–394.

Bingham, E., Cohrssen, B., Powell, C.H. (Eds.), 2001b. Patty's Toxicology, fifth edition, vol. 6, Glycidol. John Wiley and Sons, New York, pp. 1132–1134.

Bulsiewicz, H., 1977. The influence of phenol on chromosomes of mice (Mus Musculus) in the process of spermatogenesis. Fol. Morph. 36, 13–22.

Butterworth, B.E., 2006. A classification framework and practical guidance for establishing a mode of action for chemical carcinogens. Regul. Toxicol. Pharmacol. 45, 9–23.

CERI (Chemicals Evaluation and Research Institute), 1999. Chemical hazard data sheet, 98-21, o-Chloronitrobenzene (in Japanese). Available from: <http://qsar.cerij.or.jp/SHEET/F98_21.pdf> (accessed 20.4.2009).

CERI, 2002a. Chemical hazard data sheet, 2000-17, 2,3-Epoxy-1-propanol (Glycidol), revised (in Japanese). Available from: <http://qsar.cerij.or.jp/SHEET/F2000_17.pdf> (accessed 20.4.2009).

CERI, 2002b. Chemical hazard data sheet, 2000-19, 3,3′-Dichlorobenzidine, revised (in Japanese). Available from: <http://qsar.cerij.or.jp/SHEET/F2000_19.pdf> (accessed 20.4.2009).

CERI, 2002c. Chemical hazard data sheet, 2000-23, Nitrilotriacetic acid, revised (in Japanese). Available from: <http://qsar.cerij.or.jp/SHEET/F2000_23.pdf> (accessed 20.4.2009).

CERI-NITE (Chemicals Evaluation and Research Institute-National Institute of Technology and Evaluation), 2003. Hazard assessment report, 64, Acrylonitrile, ver. 1.0 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-64_1.1.pdf> (2005, ver. 1.1, accessed 20.4.2009).

CERI-NITE, 2004a. Hazard assessment report, 13, 1,4-Dioxane, ver. 1.1 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-13_1.1.pdf> (accessed 20.4.2009).

CERI-NITE, 2004b. Hazard assessment report, 35, Acrylamide, ver. 1.0 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-35_1.1.pdf> (2006, ver. 1.1, accessed 20.4.2009).

CERI-NITE, 2004c. Hazard assessment report, 63, Aniline, ver. 1.0 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-63_1.1.pdf> (2006, ver. 1.1, accessed 20.4.2009).

CERI-NITE, 2004d. Hazard assessment report, 75, Chloroethylene (Vinyl chloride), ver. 1.1 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-75_1.1.pdf> (accessed 20.4.2009).

CERI-NITE, 2005a. Hazard assessment report, 32, Phenol, ver. 1.0 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-32.pdf> (accessed 20.4.2009).

CERI-NITE, 2005b. Hazard assessment report, 33, 3,3′-Dichloro-4,4′-diaminodiphenylmethane (4,4′-Methylenebis(2-chloroaniline)), ver. 1.0 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-33.pdf> (accessed 20.4.2009).

CERI-NITE, 2005c. Hazard assessment report, 71, Formaldehyde, ver. 1.1 (in Japanese). Available from: <http://www.safe.nite.go.jp/pdf/No-71_1.1.pdf> (accessed 20.4.2009).

Costa, R., Russo, A., Zordan, M., Pacchierotti, F., Tavella, A., Levis, A.G., 1988. Nitrilotriacetic acid (NTA) induces aneuploidy in *Drosophila* and mouse germline cells. Environ. Mol. Mutagen. 12, 397–407.

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

DFG (Deutsche Forschungsgemeinschaft), 1992a. Occupational Toxicants, 4, o-Chloronitrobenzene. WILEY-VCH, Weinheim, pp. 107–114.

DFG, 1992b. Occupational Toxicants, 5,3,3′-Dichlorobenzidine. WILEY-VCH, Weinheim, pp. 99–107.

DFG, 1998. Occupational Toxicants, 10, o-Anisidine. WILEY-VCH, Weinheim, pp. 1–13.

DFG, 1999. Occupational Toxicants, 12, Ethanol. WILEY-VCH, Weinheim, pp. 129–165.

DFG, 2003a. Occupational Toxicants, 20, 1,4-Dioxane. WILEY-VCH, Weinheim, pp. 105–133.

DFG, 2003b. Occupational Toxicants, 20, Glycidol. WILEY-VCH, Weinheim, pp. 179–190.

DFG, 2007. List of MAK and BAT values, Commission for the investigation of health hazards of chemical compounds in the work area, Report No. 43. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

ECB (European Chemicals Bureau), 2000. International Uniform Chemical Information Database (IUCLID), Sodium chlorite. Available from: <http://ecb.jrc.it/esis/esis.php?PGM=hpv&DEPUIS=autre> (accessed 30.6.2005).

ECB, 2002a. EU Risk Assessment Report, 15, o-Anisidine. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/o-anisidinereport025.pdf> (accessed 20.4.2009).

ECB, 2002b. EU Risk Assessment Report, 21, 1,4-Dioxane. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/dioxanereport038.pdf> (accessed 20.4.2009).

ECB, 2002c. EU Risk Assessment Report, 28, Acrylic acid. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/acrylicacidreport028.pdf> (accessed 20.4.2009).

ECB, 2003. EU Risk Assessment Report. Cadmium Oxide, Draft: Final version, 75, 2007. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/cdoxidereport302.pdf> (final, vol. 75, cadmium oxide, Part II Human Health, accessed 20.4.2009).

ECB, 2004a. EU Risk Assessment Report, 32, Acrylonitrile. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/acrylonitrilereport029.pdf> (accessed 20.4.2009).

ECB, 2004b. EU Risk Assessment Report, 50, Aniline. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/anilinereport049.pdf> (accessed 20.4.2009).

ECB, 2006a. Classification proposal for styrene, Danish Environmental Protection Agency, November 2006, ECBI/19/06 Add. 1, part I. Available from: <http://ecb.jrc.it/classification-labelling/MEETINGS/public.htm> (accessed 30.7.2008).

ECB, 2006b. EU Risk Assessment Report, 64, Phenol. Available from: <http://ecb.jrc.ec.europa.eu/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/phenolreport060.pdf> (accessed 20.4.2009).

ECB, 2006c. Proposal for harmonized classification and labeling of styrene, U065 (UK), Styrene, ECBI/19/06 Add. 11. Available from: <http://ecb.jrc.it/classification-labelling/MEETINGS/public.htm> (accessed 30.7.2008).

ECB, 2007a. EU Draft Risk Assessment Report, Styrene, R034_0711_hh November 2007. Available from: <http://ecb.jrc.it/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/DRAFT/R034_0001_0711_env_hh.pdf> (accessed 30.7.2008).

ECB, 2007b. Industry response to the classification proposal for styrene, Submitted by the Danish Environmental Protection Agency, CEFIC Styrenics Steering Committee, January 2007, ECBI/19/06 Add. 6. Available from: <http://ecb.jrc.it/classification-labelling/MEETINGS/public.htm> (accessed 30.7.2008).

ECB, 2008. EU Annex I, Annex I to Directive 67/548/EEC on Classification and Labelling of Dangerous Substances. Available from: <http://ecb.jrc.it/classification-labelling/search-classlab/> (accessed 30.6.2008).

ECETOC (European Center of Ecotoxicology and Toxicology of Chemicals), 1998. Technical report 31, The mutagenicity and carcinogenicity of vinyl chloride: a historical review and assessment.

ECHA (European Chemicals Agency), 2008a. Guidance on information requirements and chemical safety assessment, Chapter R.3: information gathering, May 2008. Available from: <http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r3_en.pdf?vers=20_08_08> (accessed 20.4.2009).

ECHA (European Chemicals Agency), 2008b. Guidance on information requirements and chemical safety assessment, Chapter R.4: evaluation of available information, May 2008. Available from: <http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r4_en.pdf?vers=20_08_08> (accessed 20.4.2009).

ECHA (European Chemicals Agency), 2008c. Guidance on information requirements and chemical safety assessment, Chapter R.7a: endpoint specific guidance, May 2008. Available from: <http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08> (accessed 20.4.2009).

EPA (US Environmental Protection Agency), 2000. Integrated Risk Information System (IRIS), Chlorite (sodium salt), Toxicological review of chlorine dioxide and chlorite, EPA/635/R-00/007, September 2000. Available from: <http://cfpub.epa.gov/ncea/iris/index.cfm> (accessed 12.10.2007).

EPA, 2002a. IRIS, Phenol. Available from: <http://cfpub.epa.gov/ncea/iris/index.cfm> (accessed 15.10.2007).

EPA, 2002b. Toxicological review of phenol, September 2002, EPA/635/R-02/006. Available from: <http://cfpub.epa.gov/ncea/iris/index.cfm> (accessed 15.10. 2007).

European Communities, 2001. Commission directive 2001/59/EC of 6 August 2001, Offi. J. Eur. Commun. L225. (accessed 21.8.2001).

Hayashi, M., Kishi, M., Sofuni, T., Ishidate Jr., M., 1988. Micronucleus tests in mice on 39 food additives and eight miscellaneous chemicals. Food Chem. Toxicol. 26, 487–500.

Hoffmann, S., Hartung, T., 2006. Toward an evidence-based toxicology. Hum. Exp. Toxicol. 25, 497–513.

HSDB (Hazardous Substance Data Bank), 2002. 1,3-Dichloropropane, HSDB Number 5482, Last revision data, Oct 2002. Available from: <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB> (accessed 12.6.2008).

IARC (International Agency for Research on Cancer), 1982. IARC Monographs on the evaluation of the carcinogenic risks to humans, 29, Some Industrial Chemicals and Dyestuffs. IARC, Lyon, France.

IARC, 1988. IARC Monographs on the evaluation of the carcinogenic risks to humans, 44, Alcohol Drinking. IARC, Lyon, France.

IARC, 1990. IARC Monographs on the evaluation of the carcinogenic risks to humans, 48, Some Flame Retardants and Textile Chemicals, and Exposures in the Textile Manufacturing Industry. IARC, Lyon, France.

IARC, 1991. IARC Monographs on the evaluation of the carcinogenic risks to humans, 52, Chlorinated Drinking-water; Chlorination By-products; Some Other Halogenated Compounds; Cobalt and Cobalt Compounds. IARC, Lyon, France.

IARC, 1993. IARC Monographs on the evaluation of the carcinogenic risks to humans, 57, Occupational Exposures of Hairdressers and Barbers and Personal Use of Hair Colourants; Some Hair Dyes, Cosmetic Colourants, Industrial Dyestuffs and Aromatic Amines. IARC, Lyon, France.

IARC, 1994. IARC Monographs on the evaluation of the carcinogenic risks to humans, 58, Beryllium, Cadmium, Mercury, and Exposures in the Glass Manufacturing Industry. IARC, Lyon, France.

IARC, 1996. IARC monographs on the evaluation of the carcinogenic risks to humans, 65, Printing processes and printing inks, carbonblack and some nitro compounds. IARC, Lyon, France.

IARC, 1999a. IARC Monographs on the evaluation of the carcinogenic risks to humans, 71, Re-evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide. IARC, Lyon, France.

IARC, 1999b. IARC monographs on the evaluation of the carcinogenic risks to humans, 73, Some Chemicals that Cause Tumours of the Kidney or Urinary Bladder in Rodents, and Some Other Substances. IARC, Lyon, France.

IARC, 2007. IARC Monographs on the evaluation of the carcinogenic risks to humans, 96, Consumption of Alcoholic Beverages and Ethyl Carbamate (Urethane). Available from: <http://monographs.iarc.fr/ENG/Meetings/96-alcohol.pdf> (accessed 20.4.2009).

IPCS (International Program on Chemical Safety), 1989. Environmental Health Criteria (EHC), 89, Formaldehyde. Available from: <http://www.inchem.org/documents/ehc/ehc/ehc89.htm> (accessed 20.4.2009).

IPCS, 1998. Concise International Chemical Assessment Document (CICAD), 2, 3,3′-Dichlorobenzidine. Available from: <http://www.who.int/ipcs/publications/cicad/en/cicad02.pdf> (accessed 20.4.2009).

IPCS, 2001. Concise International Chemical Assessment Document (CICAD), 29, Vanadium pentoxide and other inorganic vanadium compounds. Available from: <http://www.who.int/ipcs/publications/cicad/en/cicad29.pdf> (accessed 20.4.2009).

IPCS, 2002. Concise International Chemical Assessment Document (CICAD), 39, Acrylonitrile. Available from: <http://www.who.int/ipcs/publications/cicad/cicad39_rev.pdf> (accessed 20.4.2009).

JECDB (Japan Existing Chemical Data Base), 2006a. In vitro chromosomal aberration test of 1,3-dibromopropane in cultured Chinese hamster cells (in Japanese). Available from: <http://dra4.nihs.go.jp/mhlw_data/home/paper/paper109-64-8f.html> (accessed 2.9.2006).

JECDB, 2006b. Reverse mutation test of 1,3-dibromopropane in bacteria (in Japanese). Available from: <http://dra4.nihs.go.jp/mhlw_data/home/paper/paper109-64-8e.html> (accessed 2.9.2006).

Kirkland, D., Aardema, M., Henderson, M., Müller, L., 2005. 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, 1–256.

Kirkland, D.J., Aardema, M., Banduhn, N., Carmichael, P., Fautz, R., Meunier, J.R., Pfuhler, S., 2007. *In vitro* approaches to develop weight of evidence (WoE) and mode of action (MoA) discussions with positive in vitro genotoxicity results. Mutagenesis 22, 161–175.

Meier, J.R., Bull, R.J., Stober, J.A., Cimino, M.C., 1985. Evaluation of chemicals used for drinking water disinfection for production of chromosomal damage and sperm-head abnormalities in mice. Environ. Mutagen. 7, 201–211.

Morita, T., Asano, N., Awogi, T., Sasaki, Y.F., Sato, S., Shimada, H., Sutou, S., Suzuki, T., Wakata, A., Sofuni, T., Hayashi, M., 1997. Evaluation of the rodent micronucleus assay in the screening of IARC carcinogens (groups 1, 2A, and 2B). The summary report of the 6th collaboration study by CSGMT/JEMS MMS. Mutat. Res. 389, 3–122. (Erratum, Mutat. Res. 391, 259–267, 1997).

Morita, T., Hayashi, M., Morikawa, K., 2006. Globally harmonized system on hazard classification and labeling of chemicals and other existing classification systems for germ cell mutagens. Genes Environ. 28, 141–152.

NICNAS (Australia National Industrial Chemical Notification and Assessment Scheme), 1998. PECAR, PEC/7, 1,4-Dioxane. Available from: <http://www.nicnas.gov.au/publications/car/PEC/PEC7/PEC7_Full_Report_PDF.pdf> (accessed 20.4.2009).

NITE (National Institute of Technology and Evaluation), 2005a. GHS Classification Manual, Ed. Interministerial Committee on GHS, ver. 20 October 2005. Available from: <http://www.safe.nite.go.jp/english/pdf/ghs_manual_e.pdf> (accessed 20.4.2009).

NITE, 2005b. Technical Guidance on the GHS classification, Ed. Interministerial Committee on GHS. Available from: <http://www.safe.nite.go.jp/english/pdf/ghs_guidance_e.pdf> (accessed 20.4.2009).

NTP (US National Toxicology Program), 2005a. Testing Information Data Base, Cadmium Oxide. Available from: <http://ntp.niehs.nih.gov:8080/index.html?col=010stat> (accessed 2.12.2005).

NTP, 2005b. Testing Information Data Base, 4,4′-Methylenebis(2-chloroaniline). Available from: <http://ntp.niehs.nih.gov:8080/index.html?col=010stat> (accessed 28.12.2005).

NTP, 2005c. Testing Information Data Base, Phenol. Available from: <http://ntp.niehs.nih.gov:8080/index.html?col=010stat> (accessed 28.12.2005).

NTP, 2005d. Testing Information Data Base, Vanadium Pentoxide. Available from: <http://ntp.niehs.nih.gov:8080/index.html?col=010stat> (accessed 12.12. 2005).

OECD, 2001. Screening Information Data Set (SIDS) Initial Assessment Report, 1-Chloro-2-nitrobenzene. Available from: <http://www.chem.unep.ch/irptc/sids/OECDSIDS/CHLORONITROB.pdf> (accessed 20.4.2009).

OECD, 2004a. Screening Information Data Set (SIDS) Initial Assessment Report, Ethanol. Available from: <http://www.chem.unep.ch/irptc/sids/OECDSIDS/64175.pdf> (accessed 20.4.2009).

OECD, 2004b. Screening Information Data Set (SIDS) Initial Assessment Report, Formaldehyde. Available from: <http://www.chem.unep.ch/irptc/sids/OECDSIDS/FORMALDEHYDE.pdf> (accessed 20.4.2009).

Pacchierotti, F., Ranaldi, R., 2006. Mechanisms and risk of chemically induced aneuploidy in mammalian germ cells. Curr. Pharm. Des. 12, 1489–1504.

Phillips, B.J., Jenkinson, P., 2001. Is ethanol genotoxic? a review of the published data. Mutagenesis 16, 91–101.

Pratt, I.S., Barron, T., 2003. Regulatory recognition of indirect genotoxicity mechanisms in the European Union. Toxicol. Lett. 140–141, 53–62.

Robbiano, L., Carrozzino, R., Puglia, C.P., Corbu, C., Brambilla, G., 1999. Correlation between induction of DNA fragmentation and micronuclei formation in kidney cells form rats and humans and tissue-specific carcinogenic activity. Toxicol. Appl. Pharamcol. 161, 153–159.

Robbins, W.A., Vine, M.F., Truong, K.Y., Everson, R.E., 1997. Use of fluorescence in situ hybridization (FISH) to assess effects of smoking, caffeine, and alcohol on aneuploidy load in sperm of healthy men. Environ. Mol. Mutagen. 30, 175–183.

RTECS (Registry of Toxic Effects of Chemical Substances), 2003. US National Institute for Occupational Safety and Health (NIOSH), Sodium chlorite, RTECS No. VZ4800000, Last revision data, August 2003. Available from: http://csi.micromedex.com (accessed 30.6.2005).

Schmid, O., Speit, G., 2006. Genotoxic effects induced by formaldehyde in human blood and implications for the interpretation of biomonitoring studies. Mutagenesis 22, 69–74.

Speit, G., Schmid, O., Fröhler-Keller, M., Lang, I., Triebig, G., 2007. Assessment of local genotoxic effects of formaldehyde in humans measured by the micronucleus test with exfoliated buccal mucosal cells. Mutat. Res. 627, 129–135.

Spencer, P.J., Gollapudi, B.B., Waechter Jr., J.M., 2007. Induction of micronuclei by phenol in the mouse bone marrow: I. Association with chemically induced hypothermia. Toxicol. Sci. 97, 120–127.

Thybaud, V., Aardema, M., Casciano, D., Dellarco, V., Embry, M.R., Gollapudi, B.B., Hayashi, M., Holsapple, M.P., Jacobson-Kram, D., Kasper, P., MacGregor, J.T., Rees, R., 2007a. Relevance and follow-up of positive results in *in vitro* genetic toxicity assays: an ILSI-HESI initiative. Mutat. Res. 633, 67–79.

Thybaud, V., Aardema, M., Clements, J., Dearfield, K., Galloway, S., Hayashi, M., Jacobson-Kram, D., Kirkland, D., MacGregor, J.T., Marzin, D., Ohyama, W.,

Schuler, Suzuki, H., Zeiger, E., 2008b. Expert working group on hazard identification and risk assessment in relation to *in vitro* testing. Strategy for genotoxicity testing: hazard identification and risk assessment in relation to *in vitro* testing. Mutat. Res. 627, 41–58.

Tweats, D.J., Blakey, D., Heflich, R.H., Jacobs, A., Jacobsen, S.D., Morita, T., Nohmi, T., O'Donovan, M.R., Sasaki, Y.F., Sofuni, T., Tice, R., 2007. Report of the IWGT working group on strategies and interpretation of regulatory *in vivo* tests. I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. Mutat. Res. 627, 78–91.

UKCOM (UK Committee on Mutagenicity), 2007. Formaldehyde: Evidence for Systemic Mutagenicity, COM/07/S5, November 2007. Available from: <www.advisorybodies.doh.gov.uk/com/formalde.htm> (accessed 10.12.2008).

UN (United Nations), 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), New York and Geneva. Available from: <http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_e.html> (accessed 20.4.2009).

UN, 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First revised edition, New York and Geneva. Available from: <http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html> (accessed 20.4.2009).

UN, 2007. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second revised edition, New York and Geneva. Available from: <http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html> (accessed 19.8.2008).

UNITAR (United Nations Institute for Training and Research), 2007. WSSD Global GHS Partnership Annual Report 2007. Available from: <http://www.unitar.org/cwm/publications/pag_ghs/pag13/PAG_13-3_Guide.pdf> (accessed 19.8.2008).

UNITAR, 2008a. Course 2: classifying chemicals according to the GHS, part 1: understanding classification, part 2: classification (Draft version 27 June 2008). Available from: <http://www.unitar.org/cwm/ghs/ghs14.html> (accessed 19.8.2008).

UNITAR, 2008b. Implementation of GHS, Development of lists of classification, UN/SCEGHS/15/INF.32, Fifteenth session, Geneva, 9–11 July 2008. Available from: <http://www.unece.org/trans/doc/2008/ac10c4/UN-SCEGHS-15-inf32e.pdf> (accessed 19.8.2008).

Waters, M.D., Stack, H.F., Jackson, M.A., Bridges, B.A., Adler, I.D., 1994. The performance of short-term tests in identifying potential germ cell mutagens: a qualitative and quantitative analysis. Mutat. Res. 341, 109–131.

Mutation Research 680 (2009) 31–42

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities—Report from an ECVAM workshop☆

Stefan Pfuhler[a,q,*], David Kirkland[b], Peter Kasper[c,q], Makoto Hayashi[d], Philippe Vanparys[e,q], Paul Carmichael[f,q], Stephen Dertinger[g], David Eastmond[h], Azeddine Elhajouji[i], Cyrille Krul[j], Andreas Rothfuss[k], Gabriele Schoening[l], Andrew Smith[m], Guenter Speit[n], Claire Thomas[o,q], Jan van Benthem[p,q], Raffaella Corvi[o,q]

[a] Procter & Gamble, Cosmital SA, Rte de Chesalles 21, CH-1723 Marly, Switzerland
[b] Covance Laboratories Ltd., Otley Road, Harrogate HG3 1PY, England, United Kingdom
[c] Federal Institute for Drugs and Medical Devices (BfArM), Kurt-Georg-Kiesinger-Allee 3, D-53175 Bonn, Germany
[d] Biosafety Research Center, Foods, Drugs and Pesticides, 582-2, Shioshinden, Iwata, Shizuoka 437-1213, Japan
[e] Altoxicon BVBA, Boskant 101, B-2350 Vosselaar, Belgium
[f] Unilever Colworth Science Park, Safety and Environmental Assurance Centre, Sharnbrook, Bedfordshire MK44 1LQ, England, United Kingdom
[g] Litron Laboratories, 200 Canal View Blvd., Rochester, NY 14623, USA
[h] Environmental Toxicology Graduate Program, 2109 Biological Sciences Building, University of California, Riverside, Riverside, CA 92521, USA
[i] Novartis Pharma AG, MUT-2881.5.38, CH-4002 Basel, Switzerland
[j] TNO Quality of Life, Utrechtseweg 48, PO Box 360, 3500 AJ Zeist, The Netherlands
[k] Bayer Schering Pharma AG, Nonclinical Drug Safety, 13342 Berlin, Germany
[l] European Chemicals Agency – ECHA, Directorate B.2, P.O. Box 400, FI-00121 Helsinki, Finland
[m] Health and Safety Executive (HSE), Redgrave Court, Merton Road, Bootle, Merseyside L20 7HS, England, United Kingdom
[n] Ulm University, Human Genetics, Oberer Eselsberg, Ulm D-89069, Germany
[o] In vitro Methods/European Centre for the Validation of Alternative Methods (ECVAM), Institute for Health and Consumer Protection (IHCP), JRC of the European Commission TP, 580, Via E. Fermi 2749, 21027 Ispra (Va), Italy
[p] National Institute for Public Health and the Environment, Antonie van Leeuwenhoeklaan 9, P.O. Box 1, 3720 BA Bilthoven, The Netherlands
[q] European Centre for the Validation of Alternative Methods (ECVAM) – Genotoxicity and Carcinogenicity Expert Team, Italy

ARTICLE INFO

Article history:
Received 30 March 2009
Received in revised form 28 July 2009
Accepted 27 August 2009
Available online 16 September 2009

ABSTRACT

In vivo genetic toxicology tests measure direct DNA damage or the formation of gene or chromosomal mutations, and are used to predict the mutagenic and carcinogenic potential of compounds for regulatory purposes and/or to follow-up positive results from *in vitro* testing. These tests are widely used and consume large numbers of animals, with a foreseeable marked increase as a result of the EU chemicals legislation (REACH), which may require follow-up of any positive outcome in the *in vitro* standard battery with appropriate *in vivo* tests, regardless of the tonnage level of the chemical.

A 2-day workshop with genotoxicity experts from academia, regulatory agencies and industry was hosted by the European Centre for the Validation of Alternative Methods (ECVAM) in Ranco, Italy from 24 to 25 June 2008. The objectives of the workshop were to discuss how to reduce the number of animals in standard genotoxicity tests, whether the application of smarter test strategies can lead to lower animal numbers, and how the possibilities for reduction can be promoted and implemented.

The workshop agreed that there are many reduction options available that are scientifically credible and therefore ready for use. Most of these are compliant with regulatory guidelines, i.e. the use of one sex only, one administration and two sampling times *versus* two or three administrations and one sampling time for micronucleus (MN), chromosomal aberration (CA) and Comet assays; and the integration of the MN endpoint into repeat-dose toxicity studies. The omission of a concurrent positive control in routine CA and MN tests has been proven to be scientifically acceptable, although the OECD guidelines still require this; also the combination of acute MN and Comet assay studies are compliant with guidelines, except for sampling times.

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

* Corresponding author. Tel.: +41 26 435 2520.
E-mail address: pfuhler.s@pg.com (S. Pfuhler).

1383-5718/$ – see front matter
doi:10.1016/j.mrgentox.2009.09.002

Keywords:
Genotoxicity *in vivo*
Reduction
Animal use

Based on the data presented at the workshop, the participants concluded that these options have not been sufficiently utilized to date. Key factors for this seem to be the uncertainty regarding regulatory compliance/acceptance, lack of awareness, and an in many cases unjustified uncertainty regarding the scientific acceptance of reduction options. The workshop therefore encourages the use and promotion of these options as well as the dissemination of data related to reduction opportunities by the scientific community in order to boost the acceptance level of these approaches. Furthermore, experimental proof is needed and under way to demonstrate the credibility of additional options for reduction of the number of animals, such as the integration of the Comet assay into repeat-dose toxicity studies.

1. Introduction

In vivo genetic toxicology tests are used to predict the mutagenic and carcinogenic potential of compounds for regulatory purposes and/or to follow-up positive results from *in vitro* testing. They measure direct DNA damage, its repair or the formation of gene or chromosomal mutations following induction of DNA damage by test compounds. These tests are widely used for pharmaceuticals, industrial chemicals, pesticides, biocides, food additives and cosmetic ingredients. The results form the scientific basis for risk assessment and are used for classification and labelling (C&L) of chemical substances in the EU (the Dangerous Substances Directive 67/548/EEC [1] and Regulation (EC) No. 1272/2008 on the classification, labelling and packaging of substances and mixtures [2]), and across the world (UN Globally Harmonised System (GHS) [3]).

In the scientific community efforts are increasing to replace *in vivo* tests by appropriate *in vitro* tests, especially driven by regulations enforced within the European Union (EU) such as REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) [4] and the 7th Amendment to the Cosmetics Directive [5]. The 7th Amendment prohibits any acute *in vivo* genotoxicity tests for cosmetic ingredients after March 2009, thereby triggering search for innovative hazard and risk assessment concepts. Tens of thousands of substances being tested for genotoxic potential for the purposes of REACH registration, using a classical testing scheme, could potentially lead to a very high number of additional tests in animals. The requirements as described in the REACH Integrated Testing Strategy (ITS) (Guidance on information requirements and chemical safety assessment) [6] signal the follow-up of any positive outcome in the *in vitro* standard battery with appropriate *in vivo* tests, regardless of the tonnage level of the chemical.

Taking into account the problem of low specificity of the current standard *in vitro* tests to discriminate rodent non-carcinogens from carcinogens [7,8], it becomes clear that the improvement of current *in vitro* tests (i.e. reduction of "false positives"), and the search for new approaches to *in vivo* testing that could lead to a reduction in the number of animals, present key challenges for genetic toxicologists. In a workshop hosted by the European Centre for the Validation of Alternative Methods (ECVAM) in 2006, strategies to reduce false positive results from *in vitro* tests were discussed [9] and research is under way to address this. While this will likely reduce the number of *in vivo* studies required to follow-up positive outcomes from *in vitro* tests, additional efforts will be needed to ensure a reduction – rather than further increases – of the total number of animals. This is because of the very large number of chemical substances that will be subject to the REACH information requirements and the mandatory *in vivo* testing requirements that still apply to some categories of substances (e.g. pharmaceuticals, pesticides).

The ECVAM Task Force Genotoxicity and Carcinogenicity has recently conducted a survey on opportunities to reduce the number of animals in genotoxicity testing. A questionnaire was sent to industry and contract research organisations (CROs) with the goal to investigate current practice regarding animal use in the *in vivo* micronucleus and chromosomal aberration tests (e.g. number of animals, gender, use of negative and positive controls, etc.). The results of this survey revealed that the opportunities for reduction provided by the OECD guidelines and IWGT recommendations [10,11], although apparent for many years, have not been implemented generally [12]. For example, although there is the possibility to use one sex of test animal only and an option to omit some controls, the majority of laboratories still use animals of both sexes and a concurrent positive control with every assay. When asked whether a reduction of the size of the positive control group would be acceptable, all participants of the survey were prepared to follow this – provided that this would be accepted by the regulatory authorities. A further trigger for a subsequent workshop was the ongoing revision of the ICH guidelines for genotoxicity testing, which strongly promotes the integration of genotoxicity tests into repeat-dose toxicity (RDT) studies, an approach also mentioned as an option in the REACH ITS.

The 2-day workshop was hosted and sponsored by ECVAM in Ranco, Italy from 24 to 25 June 2008. Seventeen genotoxicity experts from academia, regulatory authorities and industry were invited to contribute their experiences. The objectives of the workshop were:

- To discuss how to reduce the number of animals in standard genotoxicity tests.
- To discuss whether the number of animals can be reduced by application of smarter test strategies.
- To find a way forward how these possibilities for reduction can be best promoted and implemented.

The overriding premise agreed by all participants was that a reduction of the number of animals in genotoxicity testing should not compromise the safety standards, and that a poorly conducted and poorly designed *in vivo* study (i.e. with too few animals) is a waste of animals.

2. Summaries of the presentations given at the workshop

Information from the presentations given by various participants relevant to decision-making is summarised below.

2.1. Regulatory background

Raffaella Corvi from ECVAM, Italy informed the group that the Joint Research Centre of the European Commission estimated that, under REACH, genotoxicity is among the endpoints for which the highest number of *in vivo* tests will be needed [13]. Stefan Pfuhler from Procter & Gamble, Switzerland, summarized the opportunities that the REACH ITS [6] offers to the toxicologist to reduce the number of animals used in this area. These include the integration of genotoxicity endpoints into repeat-dose toxicity studies "if scientifically justified", and the need for a second *in vivo* study only if it is required to conclude on the relevance of the positive results *in vitro* (as opposed to the performance of a mandatory

second *in vivo* study triggered by tonnage only [4]). Furthermore, the *in vivo* Comet assay is listed as a suitable follow-up for positive results from both *in vitro* gene mutation and chromosomal aberration (CA) endpoints. This may enable registrants to omit a second *in vivo* assay irrespective of the *in vitro* profile. It was recognised that the Comet assay does not detect possible aneugenic chemicals; these, however, can be picked up at the *in vitro* testing stage.

The impact of the revision of the ICH S2 guideline on reduction in animal usage was presented by Peter Kasper from the German Federal Institute for Drugs and Medical Devices (BfArM). The revised guideline, which at the time of the workshop was one step away from finalization, touches several areas that are expected to affect (lower) the number of animals to be used for genotoxicity testing of pharmaceuticals. The main changes compared to the guideline currently in place are:

- The highest concentration used in *in vitro* mammalian cell assays is decreased from 10 to 1 mM, which should reduce the number of irrelevant positive results and lead to a decrease of *in vivo* follow-up studies.
- One sex only approach as a default option for classical genotoxicity studies such as the bone marrow micronucleus (MN) test or the assay to detect unscheduled DNA synthesis (UDS) measured in hepatocytes *ex vivo*. The use of both sexes will only need to be considered if any existing data indicate a toxicologically meaningful sex difference in the species used.
- It is sufficient to include a positive control animal group (either concurrently or separately) only periodically, after a laboratory has established competence in the use of the assay.
- Integration of *in vivo* genotoxicity endpoints into repeat-dose toxicity (RDT) studies is preferable. In cases where this does not meet the criteria for a sufficient exposure, acute studies can be performed but testing of different endpoints and/or tissues should be combined into one study, where possible.

Two different scenarios for *in vivo* testing will be available for the pharmaceutical industry, see Fig. 1.

2.2. Use of one sex versus two sexes

David Kirkland from Covance, UK, presented data from 23 rat and 7 mouse MN studies where the positive control, cyclophosphamide (CPA), was compared in animals of both sexes. For CPA in

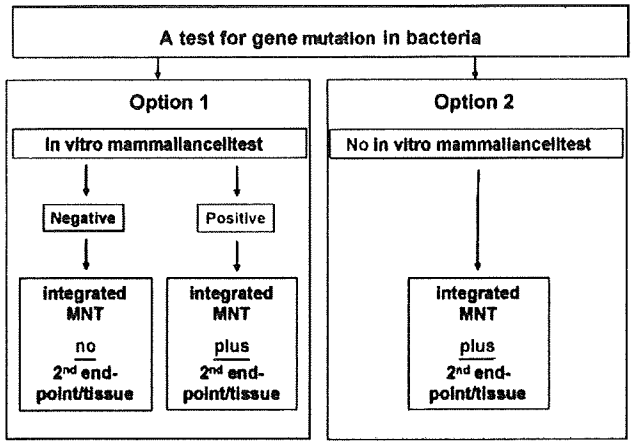

Fig. 1. Schematic of the draft revised ICH S2 (R1) guideline which provides the pharmaceutical industry with two options to perform the basic genotoxicity battery (integrated MNT = rodent micronucleus test with haematopoietic cells integrated into repeat-dose toxicity study).

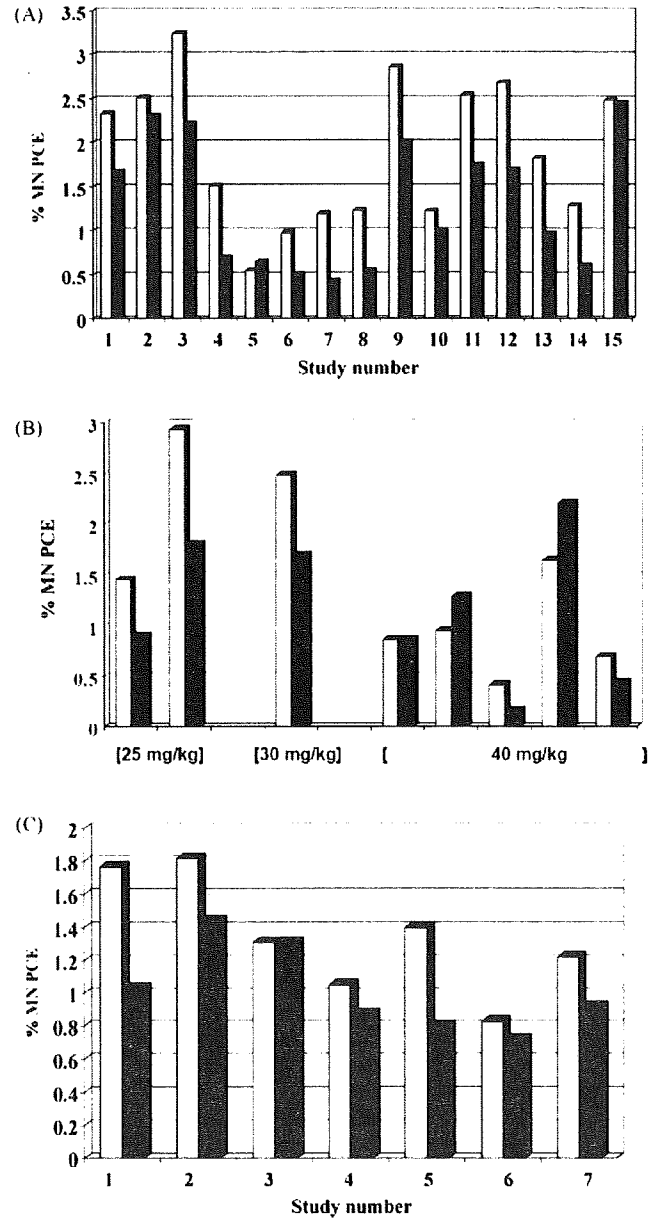

Fig. 2. Comparison of positive control group (cyclophosphamide) results between male (□) and female (■) rats from 23 rat and 7 mouse MN BM studies. All studies were carried out in the same laboratory. Studies compared were from male and female rats treated with 20 mg/kg cyclophosphamide (A), male and female rats treated with low, medium and high doses of cyclophosphamide (B) and male and female mice treated with 40 mg/kg cyclophosphamide (C).

rats, the MN response was greater in males than females in almost every study (Fig. 2A and B). The MN responses in female rats were similar or greater than in males only at 40 mg/kg but the responses in males were still positive. This dose is probably too high for routine use. In mice, the responses to CPA in males were never less than the responses in females (Fig. 2C).

Covance has also examined test chemical data from three rat studies using both sexes in which the test chemical gave a positive response. For these chemicals, positive responses were observed in all male rats whereas 1 out of 3 compounds were negative in female rats. Females only showed stronger MN responses than males when bone marrow toxicity was greater.

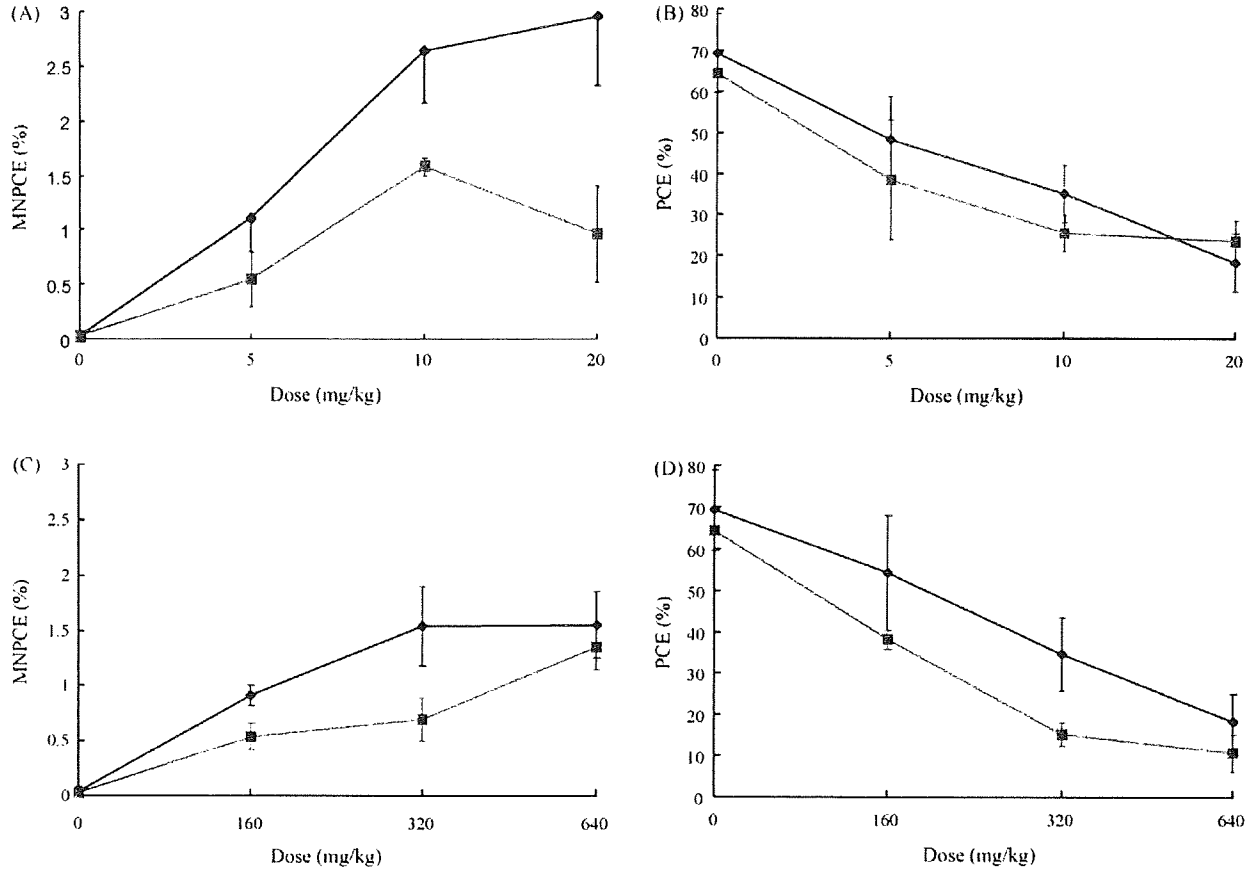

Fig. 3. Sex differences in the formation of MN in rats treated with cyclophosphamide (A and B) and cytarabine (C and D). Male (◆) and female (■) rats were dosed with test compounds by i.p. injection and the percentage MN in polychromatic erythrocytes (PCE) (A and C) and % PCEs among total erythrocytes (B and D) were measured 24 h after treatment.

Makoto Hayashi from the Biosafety Research Center in Shizuoka, Japan, presented data for CPA and Cytarabine (Ara-C) (Fig. 3) in male and female rats. Genotoxicity was clearly detected in rats of both sexes, the MN response being greater in males than in females. Cyrille Krul from TNO, The Netherlands, explained that the default strategy at TNO is to perform studies with one sex, namely males, if no substantial sex difference in toxicity occurs; otherwise the more sensitive sex is used. In their laboratory, exceptions have to be justified. Azeddine Elhajouji (Novartis) reported that 90% of all *in vivo* MN studies carried out in the last few years were single-sex studies, indicating that even under the premise that substantial sex differences in toxicity would trigger both sexes to be used, this can lead to huge savings in animal use. However, results from a recently conducted survey show that such a strategy is not current practice in industry. The ECVAM Task Force, as part of its activity to capture current status with regard to reduction, had asked laboratories performing *in vivo* MN and CA assays about the current practice in the industry [12], and the survey shows that the majority of the laboratories perform most or all of their studies using both sexes.

2.3. One treatment versus two or more treatments

Administering a test chemical to animals on two occasions and sacrificing those animals at one time point (for those endpoints where sacrifice is needed) obviously reduces the numbers

Table 1
Numbers of animals used in single-administration and two administration protocols (including positive controls), assuming single sex only.

Dose group	Two administration protocol[a]	One administration protocol[b]	
	No. of animals used	No. of animals used option 1	No. of animals used option 2
Vehicle	5	10	10
Low	5	5	10
Mid	5	5	10
High	5	10	10
Positive control	5	5	5
Total	25	35	45

Option 1 = all dose groups sampled at 24 h but only top dose and control at 48 h; option 2 = all dose groups sampled at 24 h and all dose groups except positive control sampled at 48 h.

[a] Single sampling time.
[b] Two sampling times.

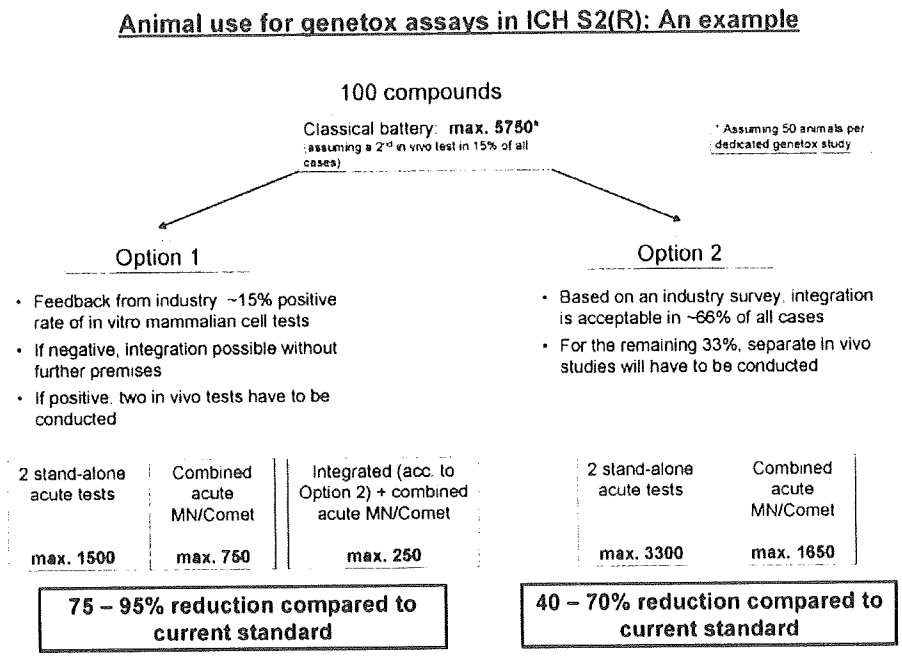

Fig. 4. Potential animal savings offered by the revised ICH S2 (R1) guideline, assuming the testing of 100 new compounds for their genotoxic properties.

of animals used compared with a single administration followed by two sacrifice times (Table 1). However, concerns had been expressed that more animals are needed to select the doses for a two-administration compared with a single-administration study. D. Kirkland therefore addressed this concern, and showed that for mean numbers of animals used in range finding the opposite is the case. Covance reviewed the numbers of animals used during range finding experiments for their last 12 studies using single- and double-administration protocols and found that for single-administration studies, the number of animals used ranged from 6 to 36 with a mean of 17, while for double-administration studies 3 to 36 animals were used, with a mean of 11. There was general consensus that for the MN and Comet assays no problems are expected when the test substance is dosed twice at the intervals suggested in OECD 474 and in the respective recommendations for the Comet assay [14,15]. Following this concept will lead to a substantial reduction in animal usage (see Table 1). However, the result of the ECVAM survey shows that to date, less than half of the labs actually utilize this option.

2.4. Omission of positive control groups

Philippe Vanparys from Altoxicon reported on the survey on the *in vivo* MN and CA test and on the use of positive and negative control animals [12]. The question whether concurrent positive controls are needed for every assay has long been debated and IWGT, following a meeting in Washington in 1999, has recommended that positive controls might not be needed [11]. The ECVAM TF survey shows where this stands today: 65% of those surveyed believed a concurrent positive control is not needed if the test facility has sufficient experience and 79% would accept a reduction of the number of animals in a concurrent positive control dose group. If regulators would accept the reduction in the number of animals, 100% of those surveyed would follow this, 59% would follow a total omission of the positive control animals if accepted by regulatory agencies. Amongst the workshop participants, Novartis has implemented the concept of a reduction in animal numbers for the positive control and has been using a positive control group of only three animals for the bone marrow MN test since 2001.

2.5. Integration of genotoxicity endpoints into repeat-dose toxicity testing

Stephen Dertinger from Litron Laboratories, USA, presented the latest status of flow-cytometric scoring of peripheral blood micronuclei, a prerequisite for efficient integration of the MN test into RDT testing. The efforts of the 9th and 13th Collaborative Study Group for the Micronucleus Test (CSGMT) were cited as important starting points that clearly established the feasibility of integration [16,17] and provided scientific proof that blood is a reliable tissue matrix for measuring micronuclei, even for rats that exhibit an efficient splenic filtration function [18,19]. Using flow cytometry, MacGregor and colleagues extended this work by demonstrating the feasibility of studying other species of toxicological interest – dogs and non-human primates [20,21]. In addition to this cross-species potential, other characteristics of the automated method that make it well-suited for integrated studies were outlined, including the scoring system's low blood-volume requirement, high throughput capacity, ability to discern mode of action based on MN size (fluorescence intensity) [22], and the objective nature of the scoring method, especially when biological standards are used for instrument calibration purposes, which controls intra- and inter-lab variability [23,24].

The joint industry initiative of pharmaceutical companies from the US, EU and Japan (Integration Initiative), which aims at investigating integration and animal saving options in the context of the revised ICH S2 guideline, was presented by Andreas Rothfuss from Bayer Schering Pharma, Germany. He emphasised the huge animal saving potential offered by the guideline options and provided an exemplary calculation for the ICH S2 (R) options 1 and 2 (Fig. 4). In the pharmaceutical industry, virtually no experience is available on the integration of the Comet assay into RDT studies with regard to sensitivity and specificity. A work plan defining experimental activities to address these concerns as well as questions related to the feasibility of integration has been suggested and is currently being discussed with the members of the Integration Initiative.

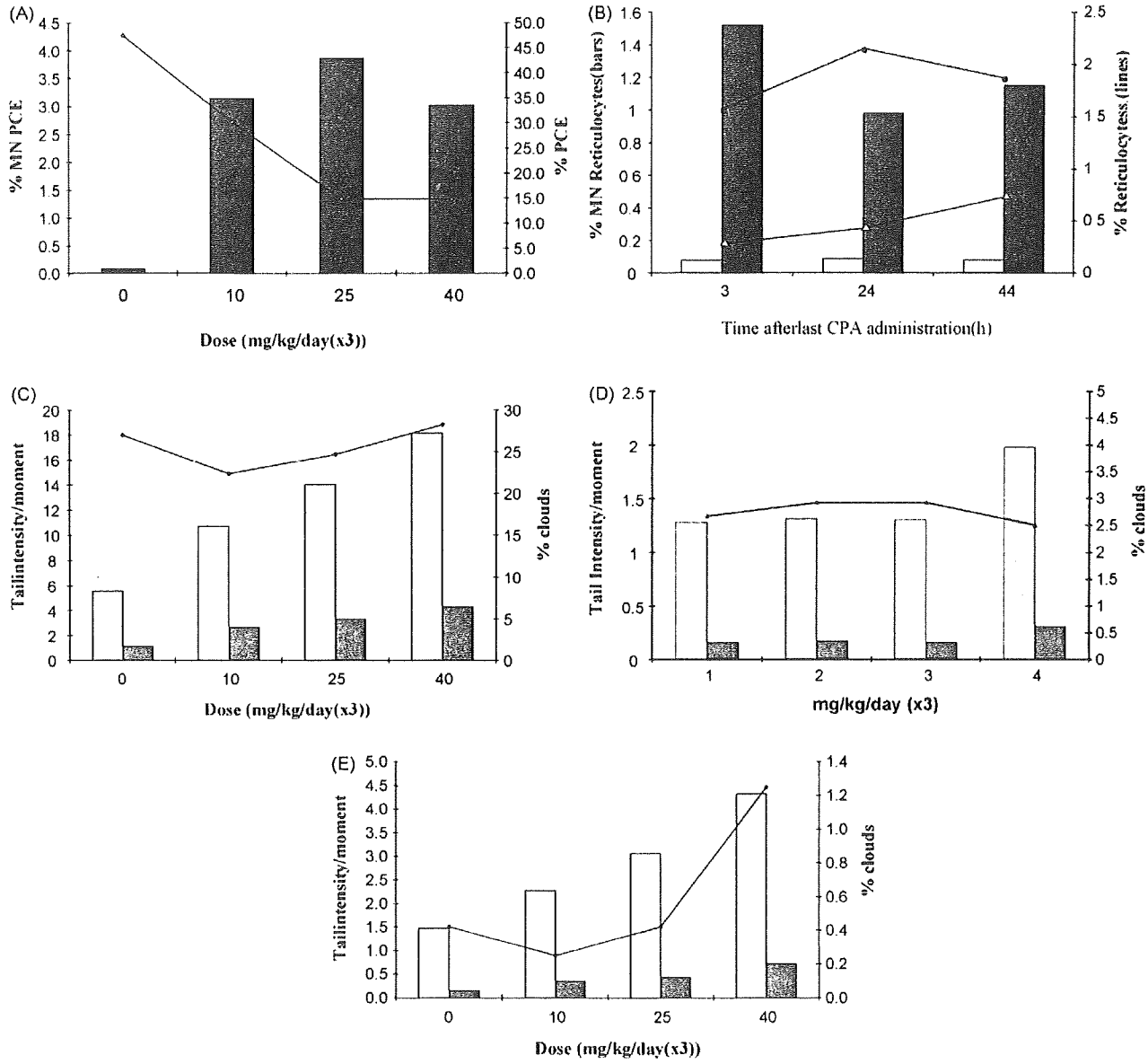

Fig. 5. MN/Comet data from a combination study performed with groups of 6 male rats (A, C, D and E) and 5 male rats (B) treated with cyclophosphamide (CPA) orally by gavage three times (48, 24 and 3 h before sacrifice). (A) MN data showing the percentage of MN polychromatic erythrocytes PCE (bars) and the percentage of PCE (line) in bone marrow; (B) MN data in rat reticulocytes measured by flow cytometry at different time points after the last administration of CPA. Rats were treated with vehicle control (□, ○) or 10 mg/kg CPA (▩, ●). Bars represent % MN in reticulocytes and lines represent the % reticulocytes among all erythrocytes. (C to E) Comet assay data. The percentage of clouds serves as indicator for cytotoxic effects (line). Tail intensity (□) and tail moment (▩) were analysed as indicators of DNA damage. Rats were treated with vehicle control or CPA. The graphs show data from the rat stomach (C), liver cells (D), and from peripheral blood cells.

Azeddine Elhajouji reported about Novartis's efforts to integrate the MN into 2-week toxicity studies. They have performed six pilot studies to date and have not observed any interference with clinical chemistry/haematology, toxicokinetics (TK) or with the sampling for histopathology. The sampling for the evaluation of micronuclei is very simple as freezing is possible after a short fixation procedure, which makes this compatible with multi-site studies.

2.6. Combination of standard assays into one study

Covance's efforts to check whether bone marrow MN and comets in stomach/liver/blood can be combined in one assay using a single sampling time were presented by D. Kirkland. Groups of six male rats were dosed on three consecutive days at 0, 24 and 45 h. A range of different chemicals (CPA, DMN, 2AAF, B(a)P, EMS, ENU, MMC and carbendazim (CBZ)) were tested, three dose levels per chemical. At 48 h (i.e. 3 h after the final dose), stomach, liver and blood were removed for Comet assay, and bone marrow was removed for analysis of MN in polychromatic erythrocytes. Additional groups of five male rats were also dosed as above, but blood samples were taken at 3, 24 and 44 h after the last dose in order to see if MN in peripheral blood reticulocytes will also give a positive response at the same (or later) sampling time. Samples were split, one being sent to Litron and the other kept at Covance for flow cytometry analysis of MN in young reticulocytes. Preliminary data for two chemicals, cyclophosphamide (CPA) and carbendazim

Table 2
Animal savings at Novartis related to the performance of combination studies of the Comet assay and the micronucleus test compared to pairs of studies for the individual assays. The impact of the use of one gender only is also demonstrated.

Information from studies	Total	Male/female	Male
Number of combination studies	7	1	6
Number of compounds tested	11	2	9

	Number of studies	Number of animals used (number of studies)	
In combination	7	51 (1)	210 (6)
Isolated MNT	4	43 (1)	93 (3)
Isolated Comet	4	38 (1)	75 (3)
Total number of animals used		**132 (2)**	**378 (9)**

Calculations	
Number of male and female animals per compound per MNT study	43
Number of male and female animals per compound per Comet study	38
Number of male and female animals needed to test 11 compounds	=(43 + 38) × 11 = 891
Number of male and female animals per compound for a combination of MNT and Comet	51
Number of male and female animals needed to test 11 compounds in combination studies	=51 × 11 = 561; saved **330** animals **(37% reduction)**
Number of male animals per compound for a combination of MNT and Comet	=210/6 = 35
Number of male animals needed to test 11 compounds	=35 × 11 = 385; saved **506** animals **(57% reduction)**

(CBZ) were shown (see Fig. 5A–E for CPA data). With this limited data set, positive genotoxicity responses can be found for MN in bone marrow and comets 3 h after a three administration protocol. The micronucleus frequency in reticulocytes from peripheral blood also increased at this 3-h sample for both compounds, which may be the better option than MN in bone marrow. These initial results suggest that different endpoints (MN, comets) can be measured at a single sample time in the same animals. Data have subsequently been obtained for DMN, EMS, ENU, and MMC. In all cases positive responses have been found either for MN in bone marrow, or in reticulocytes from peripheral blood at the 3-h sample, or for comets. The experiments with 2AAF and B(a)P are ongoing. When complete, these data will be submitted separately for publication.

A. Rothfuss reported that within the Integration Initiative repeat-dose toxicity studies are planned, amongst others, with liver-specific genotoxins to investigate the possibility of integration of the Comet assay as a complement to the MN assay and to profile the *in vivo* Comet assay against the UDS *in vivo*.

Novartis is already successfully using the combination of MN and Comet assays. Out of 11 Comet assays performed under GLP, seven were combination studies. Most of these studies (9) were carried out using males only. Performance of all studies as combination studies would have led to a 37% reduction in animal numbers, while a 57% reduction can be achieved by combining this with the use of one gender only (see Table 2).

A multiple-endpoint genotoxicity assay, combining Comet assay, MN test and gene mutation in transgenic animals was presented by M. Hayashi (Fig. 6, Table 3). Two to three male and female MutaMice were treated with three different compounds (etoposide, bleomycin and procarbazine). Etoposide, a topoisomerase inhibitor, showed a clear increase in micronuclei in the peripheral

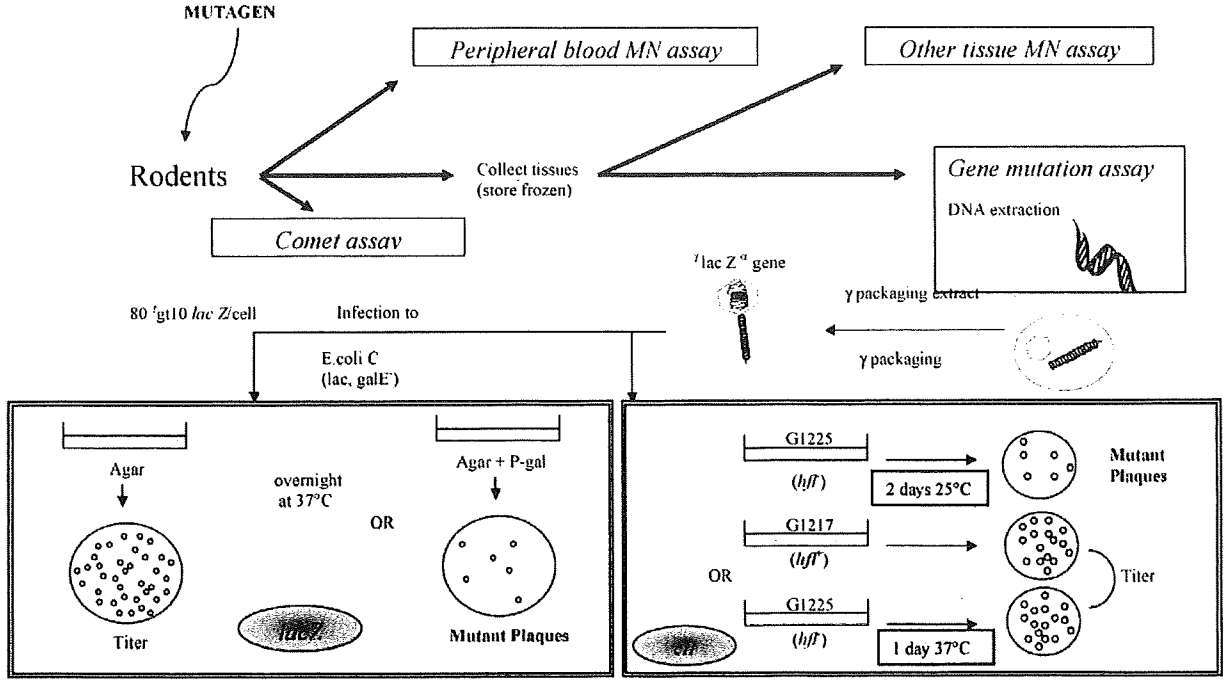

Fig. 6. Schematic of a multiple-endpoint assay combining the MN and Comet assays with the endpoint transgenic mutation (see Table 3).

Table 3
Multiple-endpoint assay in male and female transgenic animals (Muta™Mouse) for three test chemicals. DNA damage in blood was evaluated 3 h after administration of the test compound using the Comet assay (tail moment), induction of micronuclei 48 h after administration in reticulocytes (% micronucleated reticulocytes) and the frequency of *lacZ* mutations was determined in peripheral blood 14 days after the initial i.p. injection. Two to three male and female animals were used per group. Mice were treated with a single i.p. injection of test compound. Values are mean ± SD.

Test compound	Multiple-endpoint assay		
	Comet assay (3 h) Olive tail moment	MNT (48 h) MN/1000 reticulocytes	Mutation (14 day) mutation frequency (10^{-6})
Control (10 ml/kg saline)	1.2 ± 0.21	0.75 ± 0.5	9.24 ± 6.99
Etoposide (4 mg/kg)	1.16 ± 0.15	45.0 ± 14.5[a]	10.3 ± 3.19
Procarbazine HCl (200 mg/kg)	4.76 ± 0.18[a]	57.6 ± 16.8[a]	36.43 ± 36.88[a]
Bleomycin HCl (36 mg/kg)	2.55 ± 1.59[a]	4.2 ± 2.86[a]	34.81 ± 34.18[a]

[a] Represents a positive response.

blood while the other two compounds showed positive effects for all tree endpoints investigated.

3. Discussion and recommendations

Discussion of the issues and presentations took place in three break-out groups, followed by a final plenary session. There was general agreement that the current level of use of reduction opportunities is unsatisfactory. There are several actions that need to be taken in order to improve this situation, and the consensus recommendations are summarised in the following paragraphs. Some of the actions are short-term, i.e. to promote the use of reduction options that the participants considered ready to be implemented. Other actions are medium or long-term as they will require changes to the current OECD guidelines.

3.1. How to reduce the number of animals in standard genotoxicity tests?

3.1.1. Weight range

Many animals are wasted because laboratories order animals of a specific weight. When choosing the weight range for a study, the standard operating procedure (SOPs) should be as flexible as possible, to avoid unnecessary waste of animals at randomisation. We therefore encourage companies not to define too narrow a range for body-weight comparability and not to order an excessive number of animals for a study. In general, the criteria for a study should not be so strict that it becomes difficult to calculate the necessary number of animals.

3.1.2. Range finding study

If toxicity data are available, can range finder (RF) data from another facility be relied upon? With this in mind, although it may use additional animals, a small number of confirmatory RF experiments (e.g. 1–3 males and females each, as appropriate) may be worthwhile to confirm the maximum tolerated dose (MTD) and avoid possible failure of the main study.

Range finding (with or without prior data) should be undertaken by starting with the most likely dose to cause toxicity, using a small number of animals. If this does not define the MTD, then there is a need to increase or decrease the dose depending on the clinical effects of the first dose. It is a waste of animals to begin a range finder study with several dose groups and larger numbers of animals.

3.1.3. One sex versus two sexes

In the absence of data suggesting the use of both sexes, studies should use animals of one sex only. The available data suggest a slightly higher sensitivity to some positive control chemicals when males are used ([25] and Figs. 2 and 3). If at the time of the study, data show no substantial differences in toxicity between sexes, then use of a single sex should be sufficient (OECD 474 [26], paragraph 18). It is generally agreed that differences in toxicity of up to two-fold are often due to normal variability and are therefore not sufficient to demand the use of both sexes. Also, different severities of clinical effects in the two sexes would not usually justify the use of both sexes in the main study if the clinical signs themselves are similar. If available, other factors should also be taken into account, such as comparison of metabolic profiles and plasma concentrations. Some participants of the workshop suggested that where sex differences occur, the most sensitive sex can be used alone. However, if there are substantial differences between the sexes, we recommend that both sexes be used in the main study because there are insufficient data to date to identify which sex would be most likely to exhibit a positive genotoxic response. We do, however, encourage industry to evaluate available data to see whether the use of the most sensitive sex could be an option.

3.1.4. One administration and two sampling times versus two/three administrations and one sampling time

According to published literature and data presented at the workshop, the sensitivity of a protocol including a single administration with two sampling times is considered comparable with a protocol with two or three administrations and one sampling time. Whilst most comparisons of one administration *versus* two administration protocols have focused on clastogens [27], the aneugen carbendazim does elicit a positive response after three daily administrations, which provides further support for this assumption. The use of a two or three administration protocol can substantially reduce the number of animals (Table 1). Furthermore, this is a prerequisite for the combination of genotoxicity endpoints into a single study, since the kinetics of events leading to the measured endpoints differ significantly (e.g. primary DNA damage, as detected by the Comet assay, is an earlier event than MN formation, which requires the cells to go through cell division). Further saving of animals can be realized by measuring MN in blood (either through manual scoring or flow cytometry), which would allow for the use of pre-treatment samples to serve as negative controls instead of a concurrent vehicle control.

One concern voiced at the workshop was that dose-range finding for a two administration study may use more animals than for a single-administration study. However, according to the data presented at the workshop, this is not the case. We therefore strongly recommend the protocol using two or three administrations with one sampling time for acute MN and CA assays.

3.1.5. Positive controls

There was a consensus that, with sufficient experience of a laboratory, it is possible to omit the use of concurrent positive controls in routine MN and CA assays. For some other assays, e.g. the Comet assay and DNA adduct assays, it was the opinion of the workshop that at present, the technical and processing demands are such that concurrent positive controls should be used in these cases.

Table 4
Summary of results from the Japanese 13th Collaborative Study Group for the Micronucleus Test (CSGMT). The outcome of the testing of 15 different mutagens after daily administration of the compounds to rats for 28 days is shown for bone marrow (B.M.) and peripheral blood (P.B.), and for an initial measurement in P.B. at day 4.

Chemical name	Toxicity data/dose	Dose-range in this study (mg/kg/day)	Detection in the dose for general toxicity study			
			B.M. (4 week)	P.B. (4 week)	P.B. (4 day)	Overall call
2AAF	Mouse LD_{50}: 810 mg/kg	12–380; 60–240	○	○	△	○
B[a]P		31.3–250	○	△	X	○
CP	LD_{50}: 94 mg/kg	1–4; 15	●	●	○	●
DMH	LD_{50}: 100 mg/kg	5–50	X	X	X	X
2-Methylaziridine	LD_{50}: 19 mg/kg	2.5–20	○	○	○	○
MMS	LD_{50}: 225 mg/kg	3–30	●	●	●	●
MMC	LD_{50}: 30 mg/kg	0.08–0.5	X	X	△	△
Monocrotaline	LD_{50}: 66 mg/kg	3.75–30	X	X	○	○
Oil orange SS	LD_{50}: mg/kg	100–500	○	○	X	○
Phenacetin	LD_{50}: 3600 mg/kg	100–900	○	○	△	○
$KBrO_3$	LD_{50}: 321 mg/kg	10–80	○	○	X	○
Urethane	LD_{50}: 1809 mg/kg	89–300	●	●	●	●
Ara-C	Mouse LD_{50}: >5 g/kg	40–320	●	●	●	●
Colchicine	Mouse LD_{50}: 5886 μg/kg	0.7–6	○	X	X	○
6-MP	Mouse LD_{50}: 260 mg/kg	3.13–12.5	●	●	●	●

Symbols relate to the detection of general toxicity and represent "excellent correlation" (●), "good correlation" (○), "borderline between detectable and undetectable" (△) and "not measured" (X).

Periodically performed positive controls (e.g. every few months) for MN and CA assays would still be needed for quality control and furthermore would provide slides/samples that can be evaluated for experiments performed without a concurrent positive control, provided that the slides/samples were stored and coded appropriately. Hence, such slides/samples would allow checking the critical aspects such as the quality of staining and the ability of the evaluator or instrument used (e.g. flow cytometer, image analyser) to distinguish positive from negative responses. The key criteria that any laboratory will have to fulfil before it can consider omitting a concurrent positive control with every experiment will be to demonstrate consistent and reproducible detection of positive control responses in multiple independent experiments, as well as the presence of a well-defined historical negative/solvent control range in the specific strain of animals used [11]. Examples of positive control chemicals are given in OECD 474 and 475. Doses should be chosen so that the effects are clear but not excessive.

3.1.6. *Toxicokinetics*

Demonstration of exposure of the target tissue is important for all studies, and those performed without knowledge of whether the test substance will reach the tissue investigated for the genotoxic effect may be a waste of animals. Toxicokinetic (TK) data are routinely generated for pharmaceuticals but they may not be available for other substance classes, e.g. industrial chemicals. In these cases, other factors such as clinical symptoms or discoloured urine may help to demonstrate that systemic exposure occurred. In order to avoid the use of extra satellite animals when TK data are needed, *in vivo* genotoxicity testing should be conducted in the species for which TK data are generated. Genotoxicity protocols involving intravenous administration of test compounds may be an alternative to carry out TK studies in cases where quantitative systemic exposure data are not readily obtained.

3.2. *Can the number of animals be reduced by applying smarter testing strategies?*

3.2.1. *Integration of the MN endpoint into repeat-dose toxicity studies*

Integration of genotoxicity endpoints into repeat-dose studies (14- to 90-days RDT studies, dose-range finding studies, teratogenicity, etc.) is possible providing the highest dose tested is appropriate (i.e. represents a limit dose, MTD, maximum feasible dose (MFD), saturation, accumulation or accounts for at least 50% of the top dose of the acute toxicity test). The preferred method for MN analyses of blood sampled during the study is flow cytometry, however, acridine-orange staining with manual scoring is also considered acceptable.

The workshop participants addressed the question of whether integration of genotoxicity endpoints into acute toxicity testing would be possible. Acute toxicity testing uses a very limited number of animals and would test the dose groups in a step-wise manner. Therefore, no concurrent negative or positive controls would be included, which would be neither scientifically acceptable nor compliant with the current guidelines.

The influence of additional blood sampling on other parameters of repeat-dose toxicity (RDT) should be considered, but initial experience from pharmaceutical companies indicates that it is most probably not an issue. The integration of genotoxicity endpoints offers the possibility to cross-reference the results with routine haematology parameters evaluated during the RDT study for effects such as haematotoxicity or accelerated erythropoiesis, confound-

Table 5
Reduction options and their compatibility with current guidelines/recommendations.

	OECD	Pharmaceuticals e.g. ICH	Chemicals e.g. REACH ITS	IWGT and others
Dose range finding	No barrier	No barrier	No barrier	No barrier
Single sex	Covered adequately	Covered adequately Usually males	Covered adequately	Covered adequately Usually males
Omission positive control	No provision	Mentioned as option	Accepted, see OECD text and Art 13 par 3 REACH	Recommended
Combination of endpoints	Compliant except for sampling times Applies to MN; UDS	Recommended where appropriate Applies to MN; Comet; UDS; TG; etc.	Scope available	Possible Applies to MN; Comet; UDS; TG; etc.
Integration into repeated dose	No barrier	Recommended	Scope available	Proposed

ing factors known to have the potential to generate artefacts in the MN test [10].

Integration of the MN endpoint into RDT studies is in compliance with the OECD guideline for the MN test [26], strongly encouraged by ICH and an option in the REACH ITS [6] and, most importantly, scientifically credible. The workshop concluded that integration of the MN test should be the standard if RDT studies are foreseen for the test compound. In these cases, the performance of a stand-alone acute study should be justified. Such an integrated MN study should include an early blood sample taken 3 to 4 days after the start of dosing for highly clastogenic/toxic compounds (Table 4) [11].

3.2.2. Integration of the Comet assay endpoint into repeat-dose toxicity studies

The participants of the workshop viewed the integration of the Comet assay into RDT studies differently. To date, no published data are available to support this approach and industry experience is limited. There are several aspects that raise concern whether this is an appropriate option. As the Comet assay measures primary DNA effects which may be quickly repaired, according to the expert recommendations [14,15] the inclusion of a dose 3–6 h before sacrifice is needed unless the TK profile of the compound indicates accumulation. However, this may challenge the acceptability of the RDT study by influencing the general toxicity parameters or organ toxicity. If accumulation of the test compound over time occurs, this may not present a problem and the additional treatment may not be needed. However, the addition of a positive control group will still be required and therefore the full animal savings potential cannot be realized (see Section 3.1.5). Therefore, at this point in time, the workshop participants were unable to recommend this option. New information from the previously described initiative of EFPIA/PhRMA is expected and may alter this conclusion.

3.2.3. Combination of MN and Comet assay endpoints and integration into RDT studies

Ideally both the Comet and MN assays could be integrated into RDT studies but there is only limited scientific evidence available to support this. Most uncertainties are related to the Comet assay (see point 2b), however, the EFPIA/PhRMA initiative is planning to investigate this topic.

3.2.4. Combination of acute MN and Comet assay studies

In order to obtain the maximum information from one *in vivo* acute genotoxicity study, a combination of the *in vivo* MNT and the *in vivo* Comet assay is proposed. Such a combination can cover systemic genotoxic effects as well as local effects (site of contact tissue and target organ for toxicity) and different genotoxic mechanisms [28]. The test performance, including evaluation and interpretation of the results, should be in accordance with MN OECD guideline 474 and should meet state-of-the-art criteria regarding the performance of the *in vivo* Comet assay [14,15]. A test design that matches these criteria would involve administration of the test substance three times to each animal, at 48, 24 and 3–6 h prior to sacrifice. However, the third administration can be modified according to TK information, if available. It is recognised, however, that the third administration is not in compliance with OECD 474 as it is stated therein that the last administration should be within 18–24 h before sacrifice. While this is a formal deviation from the guideline, from a scientific point of view this is acceptable: the last administration is done to comply with the requirements for the Comet assay and no unfavourable impact on the outcome of the micronucleus test is expected. The last application relevant for the evaluation of micronuclei is done 24 h before sacrifice, which is in compliance with OECD 474.

According to the MN OECD guideline, three dose levels should be tested, while for the Comet, two (high and mid-dose) might be sufficient, whereas only one dose level is needed for MN and Comet if the limit test is performed. In order to avoid the inclusion of separate positive control animals for each endpoint, it is highly recommended to use a positive control compound that is suitable for both endpoints. In cases where a positive control might not be needed for the MN endpoint (see respective chapter), the Comet standard positive control can be used. Further studies are currently ongoing to confirm the compatibility of a three administration approach. However, based on the data presented at the workshop, the participants were convinced that such a combination of tests can already be used in laboratories that can demonstrate sufficient experience in both assays.

Other approaches for combinations of assays such as Comet and MN tests integrated into a transgenic mouse study, as presented by Makoto Hayashi, were also discussed and were considered to be valid from a scientific point of view, but currently there is a lack of experience.

3.3. How can the reduction possibilities be promoted and implemented?

The reduction options described above differ with regard to their guideline compliance (see Table 5). Fortunately, the majority of options, namely the use of single sex, one administration and two sampling times *versus* two administrations and one sampling time, and integration of genotoxicity endpoints into RDT studies, are in compliance with the OECD guidelines. Furthermore, their use is supported by the draft ICH S2 (R) guideline, encouraged by IWGT and covered in the REACH ITS. So why are these options not widely used today? The key factors seem to be uncertainty about regulatory compliance/acceptance and a simple lack of awareness. In addition, the workshop recognised that there is uncertainty regarding the scientific acceptance of reduction options, which may not be justified in all cases. However, a considerable amount of information generated by CROs and industry has not yet been published. For reduction options that can be implemented without generating conflict with the existing guidelines, the workshop participants were of the opinion that broader implementation can be achieved by better communication and motivation. The workshop agreed that the following actions may help to raise awareness and build consensus in the scientific and regulatory community:

- Collect more data regarding feasibility and use of reduction options and present these at scientific meetings.
- Generate and communicate data for those options that currently lack scientific proof (e.g. Comet integration into RDT studies).
- Use data to convince the national OECD Coordinators and encourage them to urge OECD to speed up the acceptance of amendments to guidelines and/or revisions of guidelines.
- Encourage people who are already using reduction possibilities to present their results and publish data in peer-reviewed literature.
- Use scientific, industry and regulatory societies and networks to promote reduction options.
- Use the distribution list from the ECVAM survey [12] to follow-up questions from the workshop.

It was the opinion of the workshop participants that the implementation of reduction possibilities that are not in full compliance with the guidelines, such as the omission of the positive control group, will depend on the interest/willingness of the regulators and the quality of the supporting evidence provided. The European Parliament has repeatedly voiced concerns about animal use in general and more specifically about the increase in animal numbers expected as a result of the REACH regulation. The EU Commission,