Table 1. GHS classification on germ cell mutagens (1,2)

Category	Classification	Criteria
Category 1A	Chemicals known to induce heritable mutations in germ cells of humans	Positive evidence from human epidemiological studies.
Category 1B	Chemicals which should be regarded as if they induce heritable mutations in the germ cells of humans	—Positive result(s) from *in vivo* heritable germ cell mutagenicity tests in mammals; or —Positive result(s) from *in vivo* somatic cell mutagenicity tests in mammals, in combination with some evidence that the substance has potential to cause mutations to germ cells. This supporting evidence may, for example, be derived from mutagenicity/genotoxic tests in germ cells *in vivo*, or by demonstrating the ability of the substance or its metabolite(s) to interact with the genetic material of germ cells; or —Positive results from tests showing mutagenic effects in the germ cells of humans, without demonstration of transmission to progeny; for example, an increase in the frequency of aneuploidy in sperm cells of exposed people. Examples of *in vivo* heritable germ cell mutagenicity tests are: — Rodent dominant lethal mutation test — Mouse heritable translocation assay — Mouse specific locus test Examples of *in vivo* somatic cell mutagenicity test are: — Mammalian bone marrow chromosome aberration test — Mouse spot test — Mammalian erythrocyte micronucleus test Examples of mutagenicity/genotoxicity tests in germ cells are: (a) Mutagenicity tests: — Mammalian spermatogonial chromosome aberration test — Spermatid micronucleus assay (b) Genotoxicity tests: — Sister chromatid exchange analysis in spermatogonia — Unscheduled DNA synthesis test (UDS) in testicular cells
Category 2	Chemicals which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans	—Positive evidence obtained from experiments in mammals and/or in some cases from *in vitro* experiments, obtained from: —Somatic cell mutagenicity tests *in vivo*, in mammals; or —Other *in vivo* somatic cell genotoxicity tests which are supported by positive results from *in vitro* mutagenicity assays. Examples of genotoxicity tests in somatic cells are: — Liver UDS *in vivo* — Mammalian bone marrow sister chromatid exchanges Examples of *in vitro* mutagenicity tests are: — *In vitro* mammalian chromosome aberration test — *In vitro* mammalian cell gene mutation test — Bacterial reverse mutation tests Note: Chemicals which are positive in *in vitro* mammalian mutagenicity assays, and which also show chemical structure activity relationship to known germ cell mutagens, should be considered for classification as Category 2 mutagens.

established in analogy to the categories for carcinogenic chemicals. Category 4 carcinogenic substances are those with non-genotoxic mechanisms of action. By definition, germ cell mutagens are genotoxic. MAK categories and criteria for classification of germ cell mutagens are summarized in Table 3.

The MAK Commission describes germ cell mutagenicity as follows (9): 'Germ cell mutagens produce heritable gene mutations, and heritable structural and numerical chromosome aberrations in germ cells. The consequences of germ cell mutations in subsequent generations include genetically determined phenotypic alterations without signs of illness, reduction in fertility, embryonic or perinatal death, more or less severe congenital malformations, and genetic diseases with various degrees of health impairment. The term "germ cell mutagenicity" refers specifically to mutagenicity in male and female germ cells and is distinguished from mutagenicity in somatic cells, which can initiate cancer. Epidemiological studies, however, have been unable to

Table 2. EU criteria for classification of chemicals as mutagenic (6,7)

Category	Classification	Criteria
Category 1	Substances known to be mutagenic to human	Positive evidence from human mutation epidemiology studies will be needed. Examples of such substances are not known to date.
Category 2	Substances which should be regarded as if they are mutagenic to human	Positive results from assays showing (a) mutagenic effects, or (b) other cellular interactions relevant to mutagenicity, in germ cells of mammals *in vivo*, or (c) mutagenic effects in somatic cells of mammals *in vivo* in combination with clear evidence that the substance or a relevant metabolite reaches the germ cells. With respect to placement in category 2, at present the following methods are appropriate:2(a) *In vivo* germ cell mutagenicity assays: — specific locus mutation test, — heritable translocation test, — dominant lethal mutation test. These assay actually demonstrate the appearance of affected progeny or a defect in the developing embryo. 2(b) *In vivo* assays showing relevant interaction with germ cells (usually DNA): — assays for chromosomal abnormalities, as detected by cytogenetic analysis, including aneuploidy, caused by malsegregation of chromosomes, — test for sister chromatid exchanges (SCEs), — test for unscheduled DNA synthesis (UDS), — assay of (covalent) binding of mutagen to germ cell DNA, — assaying other kinds of DNA damage. These assays provide evidence of a more or less indirect nature. Positive results in these assays would normally be supported by positive results from *in vivo* somatic cell mutagenicity assays, in mammals or in man. 2(c) *In vivo* assays showing mutagenic effects in somatic cells of mammals, in combination with toxicokinetic methods, or other methodologies capable of demonstrating that the compound or a relevant metabolite reaches the germ cells. For 2(b) and 2(c), positive results from host-mediated assays or the demonstration of unequivocal effects in *in vitro* assays can be considered as supporting evidence.
Category 3	Substances which cause concern for human owing to possible mutagenic effects.	There is evidence from appropriate mutagenicity studies, but this is insufficient to place the substance in category 2. Assays showing (a) mutagenic effects or (b) other cellular interaction relevant to mutagenicity, in somatic cells in mammals *in vivo*. The latter would be supported by positive results from *in vitro* mutagenicity assays. For effects in somatic cells *in vivo* at present the following methods are appropriate: 3(a) *In vivo* somatic cell mutagenicity assays: — bone marrow micronucleus test or metaphase analysis, — metaphase analysis of peripheral lymphocytes, — mouse coat color spot test. 3(b) *In vivo* somatic cell DNA interaction assays: — test for SCEs in somatic cells, — test for UDS in somatic cells, — assay for the (covalent) binding of mutagen to somatic cell DNA, — assay for DNA damage, e.g, by alkaline elution, in somatic cells. Substances showing positive results only in one or more *in vitro* mutagenicity assays should normally not be classified. Their further investigation using *in vivo* assays, however, is strongly indicated. In exceptional cases, e.g., for a substance showing pronounced responses in several *in vitro* assays, for which no relevant *in vivo* data are available, and which shows resemblance to known mutagens/carcinogens, classification in category 3 could be considered.

provide any evidence as yet that exposure to chemicals or to radiation results in hereditary diseases in man. Although structural changes have been demonstrated in the chromosomes of the germ cells of men exposed to radiation, even this finding can only provide indirect evidence that such exposures could lead to hereditary disorders in the offspring. The proof that an increased frequency of hereditary diseases is related to a particular exposure would be associated with great methodological difficulties. In the human population there are a large number of hereditary diseases of unknown origin with frequencies that differ widely in different subpopulations. Since mutational events occur largely randomly in the genome, it is not to be expected that one particular substance would induce one characteristic genetic disease. Therefore, it is most unlikely that proof of a

Table 3. Categories for classification of germ cell mutagens by MAK commission (8,9)

Category	Classification	Criteria
Category 1	Substances shown to increase the mutant frequency in the progeny of exposed humans	In the section 'Epidemiological methods and their limitations' it is explained why epidemiological studies to date have not been able to prove that the exposure of a particular human population to a particular substance has resulted in an increase in the incidence of inherited mutations. This is true both for ionizing radiation and chemical mutagens. Even if epidemiological methods are improved further, it is unlikely that such proof will be available in the foreseeable future. Category 1 will therefore probably remain without any entries.
Category 2	Substances shown to increase the mutant frequency in the progeny of exposed mammals	Classified as category 2 are substances that increase the incidence of genetically modified live progeny in animal studies, for example in the specific locus test or in the test for heritable translocations. Likewise, substances that should be classified as category 2 are those that increase the incidence of embryos that die in utero, for example in the dominant lethal test.
Category 3A	Substances shown to induce genetic damage in germ cells of humans or animals, or which produce mutagenic effects in somatic cells of mammals *in vivo* and shown to reach the germ cells in an active form	The methods include tests for genotoxicity in germ cells of experimental animals, such as tests for induction of structural chromosomal changes in spermatogonia or spermatocytes, for sister chromatid exchange in spermatogonia, for micronuclei in round spermatids, for numerical chromosome changes in secondary spermatocytes or in spermatozoa, for DNA single strand breaks and for repair synthesis or for covalent binding to the DNA. Also relevant are the observations obtained from exposed human populations which provide evidence for structural or numerical chromosome changes in spermatozoa of exposed persons. The development of new methods, especially molecular genetic methods for the detection of gene mutations in germ cells is to be expected. Substances that yield positive results in tests with germ cells are classified as category 3A. Also taken into account are clearly positive results from *in vivo* tests for mutagenicity in somatic cells, for example chromosomal aberrations or micronuclei in bone marrow cells, somatic mutations in the mammalian spot test or transgenic animals, provided that it has been demonstrated that the active substance or an active metabolite reaches the germ cells after relevant exposure of the experimental animals. Such substances are also suspected of being mutagenic in germ cells. Therefore they are classified as category 3A.
Category 3B	Substances suspected of germ cell mutagens because of their genotoxic effects in mammalian somatic cells *in vivo*; in exceptional cases, substances without *in vivo* data but with clearly mutagenic *in vitro* and structurally related to known *in vivo* mutagens	If the available data are not sufficient for classification in category 3A but the substance is clearly genotoxic in somatic cells of exposed animals or humans, the substance is also suspected of being mutagenic in germ cells. Substances that have yielded positive results in one or several *in vitro* mutagenicity tests generally not classified as category 3B. An exception is made for substances for which there are no relevant *in vivo* data but which are clearly genotoxic *in vitro* and also structurally related to substances known to be genotoxic *in vivo*. Such substances raise concern and are classified as category 3B.
Category 4	Not applicable	Category 4 carcinogenic substances are those with non-genotoxic mode of action. By definition, germ cell mutagens are genotoxic. Therefore, a category 4 for germ cell mutagens cannot exist. Depending on future research results, a category 4 could be defined at a later time for genotoxic substances with targets other than DNA (*i.e.*, pure aneugens).
Category 5	Substances considered the potency is considered so low, their contribution to genetic risk for man is expected not to be significant	Substances classified as category 5 are not expected to contribute significantly to the genetic risk for humans provided the MAK value is observed. For classification in this category, information on the spectrum of effects and their dose-dependence, and toxicokinetic data for species comparison are required. Biochemical and biological end-points can be used to characterize the contribution to genetic risk. The contribution to genetic risk is considered not to be significant after exposure at the workplace if the internal exposure level of the substance or its biomarkers is in the range of the background levels in a not specifically exposed reference population: @ Under workplace conditions the levels of biochemical effect markers such as DNA and protein adducts are not significantly increased above the background levels. @ Physiological-toxicokinetic model calculations based on animal data do not reveal a significant genetic risk for humans.

causal relationship between exposure to a chemical and occurrence of heritable diseases will become available in the foreseeable future. In this situation, for the identification of germ cell mutagens the results of animal experiments must be given particular attention. The mutagenic effect of chemicals on the germ cells of exposed parent animals can be demonstrated by observing an increased mutant frequency among the progeny. In addition, the demonstration of genotoxic effects of a substance in germ cells or somatic cells provides evidence of a potential hazard for subsequent generations.'

United States of America (US): *US EPA*: A classification using the following three categories of germ cell mutagens was proposed by the US Environmental Protection Agency (EPA) in 1984 for the evaluation of chemicals with respect to their ability to induce mutations in mammalian germ cells (10,11). Category I is based on sufficient evidence obtained from at least one *in vivo* mammalian germ cell mutation test or from at least two *in vivo* somatic cell mutation tests (point mutation and/or chromosomal aberrations), plus sufficient *in vivo* evidence that the chemical interacts with mammalian germ cells. Category II is based on suggestive evidence provided from positive results of *in vivo* somatic cell mutation tests plus evidence for interaction of the chemical with mammalian germ cells, but the evidence is insufficient to place the chemical in Category I. Category III is based on limited evidence of *in vivo* mutagenic activity or interaction of the chemical with mammalian germ cell DNA or other chromatin constituents. In 1986, guidelines for a "weight-of-evidence" approach to human germ cell mutagenicity were established, leading to eight categories with a decreasing order of strength of evidence (10). In addition, a "non-mutagen" category and a category for substances with inadequate evidence were described. The eight categories of evidence are as follows: (i) positive data derived from human germ cell mutagenicity studies; (ii) valid positive results from studies on heritable mutational events (of any kind) in mammalian germ cells; (iii) valid positive results from mammalian germ cell chromosome aberration studies that do not involve transmission from one generation to the next; (iv) sufficient evidence for a chemical's interaction with mammalian germ cells, together with valid positive mutagenicity test results from two assay systems, at least one of which is mammalian (*in vivo* or *in vitro*). The positive results may be both for gene mutation and chromosomal aberrations in mammalian systems; (v) suggestive evidence for a chemical's interaction with mammalian germ cells, together with valid positive mutagenicity evidence from two assay systems as described above under #iv. Alternatively, positive mutagenicity evidence of less strength than defined under #iv, when combined with sufficient evidence for a chemical's interaction with mammalian germ cells; (vi) positive mutagenicity test results of less strength than defined under #iv, combined with suggestive evidence for a chemical's interaction with mammalian germ cells; (vii) although definitive proof of non-mutagenicity is not possible, a chemical could be operationally classified as a non-mutagen for human germ cells if it gives valid negative test results for all endpoints of concern; and (viii) inadequate evidence bearing on either mutagenicity or chemical interaction with mammalian germ cells.

This system is for the classification for transmissible germ cell genetic risk. The Guideline (10) describes 'Evidence that an agent induces heritable mutations in human beings could be derived from epidemiologic data indicating a strong association between chemical exposure and heritable effects. It is difficult to obtain such data because any specific mutation is a rare event, and only a small fraction of the estimated thousands of human genes and conditions are currently useful as markers in estimating mutation rates. Human genetic variability, small numbers of offspring per individual and long generation times further complicate such studies. In addition, only disorders caused by dominant mutations, some sex-linked recessive mutations, and certain chromosome aberrations can be detected in the first generation after their occurrence. Conditions caused by autosomal recessive disorders (which appear to occur more frequently than dominant disorders) or by polygenic traits may go unrecognized for many generations. Therefore, in the absence of human epidemiological data, it is appropriate to rely on data from experimental animal systems as long as the limitations of using surrogate and model systems are clearly stated. Despite species differences in metabolism, DNA repair, and other physiological processes affecting chemical mutagenesis, the virtual universality of DNA as the genetic material and of the genetic code provides a rationale for using various nonhuman test systems to predict the intrinsic mutagenicity of test chemicals. Additional support for the use of nonhuman systems is provided by the observation that chemicals causing genetic effects in one species or test system frequently cause similar effects in other species or systems. Evidence also exists that chemicals can induce genetic damage in somatic cells of exposed humans. Furthermore, a wide variety of different types of mutations have been observed in humans, including numerical chromosome aberrations, translocations, base-pair substitutions, and frameshift mutations. Although the cause of these mutations is uncertain, it is clear from these observations that human germ-cell DNA is subject to the same types of mutational events that are observed in other species and test systems.'

***Recent US EPA proposals*:** Dearfield *et al.* from

Table 4. Proposed mutagenicity classification categories by US EPA researchers (12)

Classification	Criteria in Somatic cells	Criteria in Germ cells
Human mutagen	Positive in human somatic cell mutagenicity studies as a result of human *in vivo* exposure. May be human carcinogens; unless the risk characterization suggest not as likely. This can include cytogenetic endpoints in tissues (such as lymphocytes) from exposed persons.	Positive in human *in vivo* germ cell mutagenicity studies. Human germ cell mutagens. This is based on positive *in vivo* findings from appropriate germ cell targets in exposed humans. It is recognized that a human germ cell mutagen is not currently identified.
Probable human mutagen	Clear evidence for genotoxic activity *in vivo* mammalian test(s), usually supported by *in vitro* test(s). Usually animal carcinogens and may be human carcinogens. Classification at this level usually means that some *in vivo* testing has been performed as follow-up to positive results from *in vitro* testing. Therefore, there is usually some supporting positive evidence from *in vitro* testing.	Sufficient evidence of interaction with mammalian germ cells with clear evidence for genotoxic activity. Includes valid positive results from studies on heritable mutational events in mammalian germ cells; or, valid positive results from mammalian germ cell chromosomal aberration studies that do not include an intergeneration test. Putative human germ cell mutagens if they reach target cells. Alternatively, it should produce positive results for DNA strand breaks (e.g. comet test), UDS, SCE and/or chromosome aberrations in germinal cells.
Possible human mutagen	Some evidence for genotoxic activity. May be carcinogenic through genotoxic mechanisms; possibly in humans. Confirmed positive results may be seen in the *in vitro* test systems without supporting evidence from the *in vivo* assays. Agents falling into this category are considered to have intrinsic mutagenic potential which is not detected *in vivo*.	Suggestive evidence of interaction with mammalian germ cells with some evidence for genotoxic activity. May be putative human germ cell mutagens if they reach target cells. For a test agent to be considered to present suggestive evidence of germ cell interaction, data are needed to demonstrate that: (i) the test agent shows some evidence of somatic cell mutagenicity and/or genotoxicity; (ii) the test agent reaches the gonads (e.g. data from pharmacokinetic/tissue distribution studies of the test agent and/or metabolites); (iii) the test agent interacts with germinal cells; these data come from subchronic or chronic toxicity tests showing gonadal pathology (e.g. sperm abnormalities); and (iv) the test agent causes adverse effects on reproductive parameters (e.g. decreased fertility, increased dead implants, reduced litter sizes).
Equivocal evidence	Results from acceptable tests that cannot be convincingly called negative or positive.	
Negative evidence	Negative results in acceptable tests.	

the US EPA proposed 6 categories of mutagenicity classification that are divided to 3 categories each for somatic cells and germ cells (12). This includes (i) human somatic cell mutagens in which positive data in human somatic cells are derived from studies with exposed humans that gives positive data in human; (ii) human germ cell mutagen in which positive data from human *in vivo* germ cell studies are obtained; (iii) probable human somatic cell mutagen in which clear evidence for genotoxic activity from positive *in vivo* mammalian test(s); (iv) probable human germ cell mutagen in which sufficient evidence of interaction with mammalian germ cells with clear evidence for genotoxic activity; (v) possible human somatic cell mutagen in which confirmed positive results may be seen in the *in vitro* test systems without supporting evidence from the *in vivo* assays; and (vi) possible human germ cell mutagen in which suggestive evidence of interaction with mammalian germ cells is seen. The mutagenicity classification categories proposed by Dearfield *et al.* (12) are summarized in Table 4.

The mutagenicity/genotoxicity data organized into "clear" and "some" evidence for mutagenicity and into "suggestive" and "sufficient" evidence germ cell interaction. The classification of mutagenicity results falls into the broad categories of inadequate, negative, equivocal, and positive data. With respect to the effects of mutagens, Dearfield *et al.* stated that 'In addition to cancer, adverse health effects from somatic cell mutations and/or germ cell mutations include sickle cell anemia, cardiovascular disease, reproductive/developmental effects, and neurobehavioral effects among many specific and general endpoints, as well as having impact on the aging process.' (12).

Canada: *Health Canada*: Health Canada proposed 6 categories of classification of chemicals with respect to

their mutagenic potential for germ cells in the Canadian Environmental Protection Act (CEPA), Human Health Risk Assessment for Priority Substances (13). Four categories have several subgroups based on the degree of evidence (Table 5). These classifications are as follows: human germ cell mutagen (group I), for which data from adequate epidemiological studies indicate that there is a causal relationship between exposure of humans to a chemical and an increased incidence of inherited mutations in live or dead offspring; probable human germ cell mutagen (group II), for which data from epidemiological studies to assess germ cell mutagenicity are inadequate: however, there is sufficient evidence of germ cell mutagenicity in animal species (*i.e.*, there is an increased incidence of gene mutations, structural or numerical chromosomal aberrations, or inherited congenital malformations in the live offspring of exposed animals; or an increase in dominant lethal mutations in the potential offspring of exposed animals); possible human germ cell mutagen (group III); unlikely to be a human germ cell mutagen (group IV); probably not a human germ cell mutagen (group V); unclassifiable with respect to germ cell mutagenicity in humans (group VI). Groups III, IV, V and VI have four, two, three and three subgroups, respectively (see Table 5).

The Canadian Environmental Protection Act (13) mentions 'These effects including mutagenic are manifested at the biochemical, cellular, histopathological and morphological levels.', and 'Chemical substances are classified, therefore, with respect to their potential carcinogenicity and mutagenicity to humans; this is accomplished on the basis of rigorous examination of the quantity, quality and nature of the results of available toxicological and epidemiological studies. The criteria by which Priority Substances are classified based on their weight of evidence of carcinogenicity and mutagenicity.'

Health Protection Branch: The Health Protection Branch defined toxicological findings that would be regarded as germ cell genotoxicity (14). The toxicological indication of germ cell genotoxicity are (i) *in vitro* test results and positive evidence for mutagenicity in somatic cells *in vivo*, and (ii) evidence from pharmacokinetic/tissue distribution studies that the test material and/or metabolites reaches the gonads; or (iii) evidence from subchronic or chronic treatment studies that gonadal pathology indicates germ cell damage; or (iv) evidence for reproductive/developmental effects showing reduced numbers of pregnancies, reduced litter sizes or increased time to mating following treatment in some cases. A significant proportion of agents that cause *in vivo* somatic cell mutation might also possess the ability to lead to mutation in germ cells that may be transmitted to offspring. When evidence for *in vivo* somatic genotoxicity is demonstrated, along with tissue distribution, metabolic and/or pathologic evidence that the genotoxic chemical (or metabolites) reaches the germ lines (whether or not overt effects on fertility are found), the possibility of induced genetic damage to germ cells leading to heritable effects should be investigated (14).

Canadian Centre for Occupational Health and Safety: Controlled Products Regulations (CPR) has two mutagenicity criteria (11,15). According to CPR paragraph 57 in the classification of subdivision A (very toxic material), a chemical or chemical mixture is considered to present a hazard to man if (a) there is epidemiological evidence that shows a causal connection between exposure of persons to the substance or mixture and heritable genetic effects; or (b) there is evidence of mutagenicity in mammalian germ cells *in vivo* as shown by (i) positive results in a study that measures mutations transmitted to offspring, or (ii) positive results in an *in vivo* study showing chemical interaction with the genetic material of mammalian germ cells and positive results in an *in vivo* study assessing either gene mutation or chromosomal aberration in somatic cells. CPR paragraph 62 in the classification of subdivision B (toxic material) explains that a pure chemical or a chemical mixture is considered to present a hazard if evidence of mutagenicity in mammalian somatic cells is obtained in a test to assess either gene mutations or chromosomal aberrations.

Japan: The Japanese Industrial Safety and Health Law addresses the mutagenicity classification of substances in the workplace (11). Under this legislation, Japan does not classify mutagenic substances according to a weight of evidence approach or whether a compound may be a somatic or germ cell mutagen but rather recognizes one category: "mutagenic". A substance is classified as mutagenic when the results of a "bacterial reverse mutation assay" are positive. Workers' health impairment by the exposure to the 'strong' mutagenic chemicals or chemical mixtures (excluding those containing 1% or less by weight of mutagenic chemicals), which induce more than 1000 revertants/mg/plate in a bacterial reverse mutation assay, should be prevented in the work for manufacture or handling of these chemicals (16).

Future of Classification of Germ Cell Mutagens

Chemical evaluation of mutagenicity/genotoxicity has three major directions that are (i) screening of carcinogens, (ii) mechanistic investigation of carcinogenesis, and (iii) investigation of heritable adverse effects in germ cells including those in humans. Mutagenicity in GHS is focused on the last topic in terms of classification of germ cell mutagens. GHS and the other existing classification systems have different

Table 5. Criteria for classification of mutagenicity in germ cells in CEPA by Health Canada (13)

Category	Classification	Criteria
Group I	Human Germ Cell Mutagen	Data from adequate epidemiological studies indicate that there is a causal relationship between exposure of humans to a substance and an increased incidence of inherited mutations in live or dead offspring.
Group II	Probable Human Germ Cell Mutagen	Data from epidemiological studies to assess germ cell mutagenicity are inadequate: however, there is sufficient evidence of germ cell mutagenicity in animal species (i.e., there is an increased incidence of gene mutations, structural or numerical chromosomal aberrations, or inherited congenital malformations in the live offspring of exposed animals; or an increase in dominant lethal mutations in the potential offspring of exposed animals).
Group III. A	Possible Human Germ Cell Mutagen	Data from epidemiological studies indicate an association between exposure and human germ cell mutagenicity, but alternative explanations such as chance, bias, or confounding cannot be excluded.
Group III. B	ditto	Data from epidemiological studies to assess germ cell mutagenicity are inadequate: however, there is sufficient evidence of somatic cell mutagenicity (*in vivo* gene mutations or chromosomal aberrations) in humans or animal species, and sufficient evidence of exposure to germ cells in humans or animal species.
Group III. C	ditto	Data from epidemiological studies to assess germ cell mutagenicity in humans are inadequate or lacking. There is sufficient data in animals to indicate that the chemical is a germ cell mutagen, but available data indicate that the induction of mutations occurs through an epigenetic threshold-based mechanism.
Group III. D	ditto	Data from epidemiological studies to assess germ cell mutagenicity in humans are inadequate. There is sufficient evidence of mutagenicity of somatic cells in humans or animal species (*in vivo* gene mutations or chromosomal aberrations), but evidence of exposure to germ cells is inadequate or lacking.
Group IV. A	Unlikely to Be a Human Germ Cell Mutagen	There is no evidence of human germ cell mutagenicity in sufficiently powerful and well-designed epidemiological studies. There is evidence of mutagenicity of somatic cells in well-designed and well-conducted studies in humans or animals, but there is no evidence of exposure of human or animal germ cells in well-designed studies.
Group IV. B	ditto	Data on germ cell mutagenicity in epidemiological studies in humans are inadequate; there is no evidence of mutagenicity *in vivo* in germ or somatic cells in well-designed and properly conducted studies in animals.
Group V. A	Probably Not a Human Germ Cell Mutagen	There is no evidence of germ cell mutagenicity in sufficiently powerful and well-designed epidemiological studies; there is no evidence of germ cell mutagenicity in animal species.
Group V. B	ditto	There is no evidence of germ cell mutagenicity in sufficiently powerful and well-designed epidemiological studies; data in animal species are inadequate.
Group V. C	ditto	Data from epidemiological studies to assess germ cell mutagenicity in humans are inadequate, but evidence of the lack of germ cell mutagenicity in animal species is strongly supported by other data on mutagenicity *in vivo*.
Group VI. A	Unclassifiable with Respect to Germ Cell Mutagenicity in Humans	Data from epidemiological and/or animal studies are inadequate (i.e., because of major qualitative limitations, the studies cannot be interpreted as showing either the presence or absence of germ cell mutagenicity).
Group VI. B	ditto	There are no *in vivo* mutagenicity data available for evaluation.
Group VI. C	ditto	Results of epidemiological studies in human populations and experimental studies in animal species are conflicting, without an identifiable mechanistic basis.

objectives, target audiences and criteria. For example, the primary objective of the systems in GHS, EU and Germany MAK is for hazard classification, on the other hand, the systems of US EPA and Health Canada are for risk assessment. Target audiences are workers in the systems of Germany MAK, Canadian CPR and Japan; consumers and workers in EU system; consumers, workers, transport workers, and emergency responders in GHS. As for criteria, the report from EPA researchers includes adverse effects on reproductive

parameters.

The fundamental purposes are different in these regulations; therefore, it is difficult to discuss the significance of the different approaches. Simple classification category and criteria will be useful for hazard classification. The GHS and EU systems on (germ cell) mutagenicity meet to this point. Application of the criteria needs expert judgment in a weight of evidence approach in GHS. However, the expert judgment leads sometimes different conclusion from expert to expert. This will be a critical issue in the classification of germ cell mutagens by GHS. Furthermore, hazard classification/evaluation is often confused as risk evaluation, especially in Japan. Further efforts including risk evaluation and communication on germ cell mutagenicity will be needed to make understanding of regulations global.

Recently, the European Commission proposed a new regulatory framework for chemicals called REACH for the Registration, Evaluation and Authorisation of Chemicals in October 2003 (17). The authorisation process pays particular attention to the risks that the substance poses due to any carcinogenic, mutagenic, and/or reproductive toxicity (CMR) properties. CMR's category 1 and 2 in the EU criteria, which correspond to the category 1A and 1B, respectively in GHS criteria, are subject to authorization (18,19). All substances imported in quantities over 1 tonne that contain more than 0.1 per cent CMR material must be authorised before gaining access to the EU market. Authorisation provides a permit for specific uses and can be requested by 'producer' or 'user'. The GHS itself is not legally binding, however, some national or regional laws including REACH may be legally binding. Now, classification of germ cell mutagens becomes an important issue.

As GHS criteria have been adopted in worldwide, it will become standard for hazard classification. After implementation of GHS in each country, the classification of chemicals on germ cell mutagenicity will be performed by chemical suppliers (manufacturers or importers). Understanding classification systems for germ cell mutagens will be helpful for scientifically sound classification of chemicals in the GHS.

Acknowledgement: We are grateful to Dr. D.J. Tweats for his review of this manuscript. This work was supported in part by the Health and Labor Sciences Research Grants (H18-Chemistry-General-009).

References

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

2 United Nations (UN). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First revised edition, New York and Geneva, 2005. (http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html)

3 Japanese inter-ministerial committee for GHS. Japanese translation version of the first revised edition of GHS documents, 2005. (in Japanese). (http://www.meti.go.jp/policy/chemical_management/kokusai/GHS/GHStexts/kariyaku.htm)

4 Japanese inter-ministerial committee for GHS. Publication of classification of chemicals under each relevant domestic law (in progress), 2006. (in Japanese). (http://www.safe.nite.go.jp/ghs/index.html)

5 European Communities (EC). Commission directive 2001/59/EC of 6 August 2001, Official J. of Europe Commun. L225, 21.8.2001.

6 Pratt IS, Barron T. Regulatory recognition of indirect genotoxicity mechanisms in the European Union. Toxicol Lett. 2003; 140–141: 53–62.

7 European Communities (EC). Classification, packaging, and labeling of dangerous substances in the EU, Part I, Office for Official Publications of the European Communities, Luxembourg, 1997.

8 Adler ID, Anrae U, Kreis P, Neumann HG, Their R, Wild D. Recommendation for the categorization of germ cell mutagens. Int Arch Occup Environ Health. 2000; 73: 428–32.

9 Deutsche Forschungsgemeinschaft (DFG). List of MAK and BAT values, Commission for the investigation of health hazards of chemical compounds in the work area, Report No. 39, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany; 2003.

10 US Environmental Protection Agency (USEPA). Guidelines for Mutagenicity Risk Assessment, Federal Register 51(185):34006-34012, EPA/630/R-98/003; September 1986.

11 Organisation for Economic Co-operation and Development (OECD). OECD Series on testing and assessment, Number 12, Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries, ENV/JM/MONO(99)2, Environment Directorate, OECD, Paris, 1999.

12 Dearfield KL, Cimino MC, McCarroll NE, Mauer I, Valcovic LR. Genotoxicity risk assessment: a proposed classification strategy. Mutat. Res. 2002; 26: 121–35.

13 Health Canada. Canadian Environmental Protection Act (CEPA). Human Health Risk Assessment for Priority Substances, Minister of Supply and Services Canada, Ottawa, Canada, Cat. No. En40-215/41E; 1994.

14 Health Protection Branch. The assessment of mutagenicity, Mutagenicity Guidelines, Ministry of Supply and Services Canada, Cat. H49-85/1993E; 1993.

15 Canadian Centre for Occupational Health and Safety (CCOHS). Controlled Products Regulations. SOR/88-66; 1988.

16 Guidelines for Preventing Health Impairment by Chemical Substances with Mutagenicity Recognizes, LSB Notification No. 312, Industrial Safety and Health Law, Japan, dated May 17, 1993.

17 European Commission. Proposal for a Regulation of the European parliament and of the council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency and amending Directive 1999/45/EC and Regulation (EC), COM(2003) 644 final, Brussels; 2003.

18 European Commission Services. Comparison between EU and GHS criteria, Human health and environment (Draft), Version June 08, 2005.

19 Morita T, Ishimitsu S, Morikawa K. Genotoxicity in risk assessment, Global perspectives. Environ. Mutagen Res. 2005; 27, 47–56. (in Japanese)

Available online at www.sciencedirect.com
ScienceDirect
Mutation Research 628 (2007) 31–55
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop☆

David Kirkland[a,*], Stefan Pfuhler[b], David Tweats[c], Marilyn Aardema[d], Raffaella Corvi[e], Firouz Darroudi[f], Azeddine Elhajouji[g], Hansruedi Glatt[h], Paul Hastwell[i], Makoto Hayashi[j], Peter Kasper[k], Stephan Kirchner[l], Anthony Lynch[i], Daniel Marzin[m], Daniela Maurici[e], Jean-Roc Meunier[n], Lutz Müller[l], Gerhard Nohynek[o], James Parry[c], Elizabeth Parry[c], Veronique Thybaud[p], Ray Tice[q], Jan van Benthem[r], Philippe Vanparys[s], Paul White[t]

[a] *Covance Laboratories Ltd., Otley Road, Harrogate HG3 1PY, UK*
[b] *Procter & Gamble, Cosmital SA, Rte de Chesalles 26, CH-1723 Marly, Switzerland*
[c] *Medical School, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK*
[d] *Procter & Gamble Co., Miami Valley Innovation Center, 11810 East Miami River Road, Cincinnati, OH 45239-8707, USA*
[e] *European Centre for the Validation of Alternative Methods (ECVAM), Institute for Health and Consumer Protection (IHCP), European Commission-Joint Research Centre, TP 580, Via E. Fermi 1, I-21020 Ispra (Va), Italy*
[f] *Department of Toxicogenetics, Leiden University Medical Centre, Einthovenweg 20, 2300-RC Leiden, The Netherlands*
[g] *Novartis Pharma AG, MUT-2881.5.38, CH-4002 Basel, Switzerland*
[h] *German Institute of Human Nutrition (DIfE) Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany*
[i] *GlaxoSmithKline R&D, Park Road, Ware, Herts SG12 0DP, UK*
[j] *Division of Genetics & Mutagenesis, NIHS, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan*
[k] *BfArM, Genetic Toxicology, Kurt-George-Kiesinger Allee 3, D-53175 Bonn, Germany*
[l] *Non-clinical Drug Safety, F. Hoffmann-La Roche, PRBN-T, Bldg. 73-311b, CH-4070 Basel, Switzerland*
[m] *Institut Pasteur de Lille, 1 rue de Profeseur Calmette, BP 245, 59019 Lille Cedex, France*
[n] *L'Oréal, Avenue Eugène Schuller, F-93600 Aulnay sous Bois Cedex, France*
[o] *L'Oréal, Worldwide Safety Department, 25-29 Quai Aulagnier, Asnières-sur-Seine Cedex 92600, France*
[p] *Sanofi-Aventis, 13 Quai Jules Guesde, BP 14, Vitry sur Seine Cedex 94403, France*
[q] *NICEATM, NIEHS, Mail Code EC-17, P.O. Box 12233, Research Triangle Park, NC 27709, USA*
[r] *National Institute of Public Health and the Environment, Antonie van Leeuwenhoeklaan 9, P.O. Box 1, 3720 BA Bilthoven, The Netherlands*
[s] *Johnson & Johnson Pharmaceutical Research and Development, Turnhoutseweg 30, B-2340 Beerse, Belgium*
[t] *Mutagenesis Section, Safe Environments Programme, Health Canada, Tunney's Pasture 0803A, Ottawa, Ontario K1A 0L2, Canada*

Received 18 September 2006; accepted 22 November 2006
Available online 13 January 2007

Abstract

Workshop participants agreed that genotoxicity tests in mammalian cells *in vitro* produce a remarkably high and unacceptable occurrence of irrelevant positive results (e.g. when compared with rodent carcinogenicity). As reported in several recent reviews,

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

* Corresponding author.
E-mail address: david.kirkland@covance.com (D. Kirkland).

1383-5718/$ – see front matter
doi:10.1016/j.mrgentox.2006.11.008

the rate of irrelevant positives (i.e. low specificity) for some studies using *in vitro* methods (when compared to this "gold standard") means that an increased number of test articles are subjected to additional *in vivo* genotoxicity testing, in many cases before, e.g. the efficacy (in the case of pharmaceuticals) of the compound has been evaluated. If *in vitro* tests were more predictive for *in vivo* genotoxicity and carcinogenicity (i.e. fewer false positives) then there would be a significant reduction in the number of animals used. Beyond animal (or human) carcinogenicity as the "gold standard", it is acknowledged that genotoxicity tests provide much information about cellular behaviour, cell division processes and cellular fate to a (geno)toxic insult. Since the disease impact of these effects is seldom known, and a verification of relevant toxicity is normally also the subject of (sub)chronic animal studies, the prediction of *in vivo* relevant results from *in vitro* genotoxicity tests is also important for aspects that may not have a direct impact on carcinogenesis as the ultimate endpoint of concern.

In order to address the high rate of *in vitro* false positive results, a 2-day workshop was held at the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy in April 2006. More than 20 genotoxicity experts from academia, government and industry were invited to review data from the currently available cell systems, to discuss whether there exist cells and test systems that have a reduced tendency to false positive results, to review potential modifications to existing protocols and cell systems that might result in improved specificity, and to review the performance of some new test systems that show promise of improved specificity without sacrificing sensitivity.

It was concluded that better guidance on the likely mechanisms resulting in positive results that are not biologically relevant for human health, and how to obtain evidence for those mechanisms, is needed both for practitioners and regulatory reviewers.

Participants discussed the fact that cell lines commonly used for genotoxicity testing have a number of deficiencies that may contribute to the high false positive rate. These include, amongst others, lack of normal metabolism leading to reliance on exogenous metabolic activation systems (e.g. Aroclor-induced rat S9), impaired *p53* function and altered DNA repair capability.

The high concentrations of test chemicals (i.e. 10 mM or 5000 μg/ml, unless precluded by solubility or excessive toxicity) and the high levels of cytotoxicity currently required in mammalian cell genotoxicity tests were discussed as further potential sources of false positive results. Even if the goal is to detect carcinogens with short *in vitro* tests under more or less acute conditions, it does not seem logical to exceed the capabilities of cellular metabolic turnover, activation and defence processes. The concept of "promiscuous activation" was discussed. For numerous mutagens, the decisive *in vivo* enzymes are missing *in vitro*. However, if the substrate concentration is increased sufficiently, some other enzymes (that are unimportant *in vivo*) may take over the activation—leading to the same or a different active metabolite. Since we often do not use the right enzyme systems for positive controls *in vitro*, we have to rely on their promiscuous activation, i.e. to use excessive concentrations to get an empirical correlation between genotoxicity and carcinogenicity. A thorough review of published and industry data is urgently needed to determine whether the currently required limit concentration of 10 mM or 5000 μg/ml, and high levels of cytotoxicity, are necessary for the detection of *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens.

In addition, various measures of cytotoxicity are currently allowable under OECD test guidelines, but there are few comparative data on whether different measures would result in different maximum concentrations for testing. A detailed comparison of cytotoxicity assessment strategies is needed. An assessment of whether test endpoints can be selected that are not intrinsically associated with cytotoxicity, and therefore are less susceptible to artefacts produced by cytotoxicity, should also be undertaken.

There was agreement amongst the workshop participants that cell systems which are *p53* and DNA-repair proficient, and have defined Phase 1 and Phase 2 metabolism, covering a broad set of enzyme forms, and used within the context of appropriately set limits of concentration and cytotoxicity, offer the best hope for reduced false positives. Whilst there is some evidence that human lymphocytes are less susceptible to false positives than the current rodent cell lines, other cell systems based on HepG2, TK6 and MCL-5 cells, as well as 3D skin models based on primary human keratinocytes also show some promise. Other human cell lines such as HepaRG, and human stem cells (the target for carcinogenicity) have not been used for genotoxicity investigations and should be considered for evaluation. Genetic engineering is also a valuable tool to incorporate missing enzyme systems into target cells. A collaborative research programme is needed to identify, further develop and evaluate new cell systems with appropriate sensitivity but improved specificity.

In order to review current data for selection of appropriate top concentrations, measures and levels of cytotoxicity, metabolism, and to be able to improve existing or validate new assay systems, the participants called for the establishment of an expert group to identify the *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens that we expect our *in vitro* genotoxicity assays to detect as well as the non-genotoxins and non-carcinogens we expect them not to detect.

Keywords: Genotoxicity *in vitro*; False positives; Animal tests

1. Introduction

The *in vitro* genetic toxicology tests used for regulatory purposes measure formation of gene mutations and chromosomal changes following DNA damage induced by the compounds under test, and are used to predict the carcinogenic potential of pharmaceuticals, industrial chemicals, food additives and cosmetic ingredients. If a compound is positive in one or more of these basic tests, further *in vitro* mechanistic studies will most likely be performed, but in addition *in vivo* genotoxicity (and possibly also carcinogenicity) studies may be undertaken to assess the health risk for humans. A recent analysis of nearly 1000 chemicals for which data have been published [1] has highlighted the strikingly imprecise nature of *in vitro* genetic toxicology tests in discriminating non-carcinogens from carcinogens. When the standard battery of two or three *in vitro* genotoxicity tests was performed, at least 80% of the 177 non-carcinogenic compounds tested gave a false positive result in at least one test. The false positive rate was highest in mammalian cell tests such as those to detect chromosomal aberrations or micronuclei in Chinese hamster cells, or mutations in the mouse lymphoma assay. A similar outcome was obtained in an analysis by the U.S. FDA of an even larger database of chemicals [2]. These findings highlight the urgent need for more meaningful *in vitro* genotoxicity tests or practical interpretation of current positives.

The high false positive rate (low specificity) of the established *in vitro* mammalian cell tests means that an increased number of compounds are subjected to earlier and additional *in vivo* genotoxicity testing. This lack of specificity is a problem when one considers the many thousands of industrial chemicals to be evaluated in the REACH programme where it is estimated that genotoxicity testing will be the 3rd most animal-consuming area of testing. In practice this will also inhibit, or even preclude, development of new cosmetic ingredients for which the 7th Amendment to the EU Cosmetics Directive foresees a complete ban on animal testing for the genotoxicity endpoint by 2009.

For pharmaceuticals, many compounds will fail in later development, and so the extra *in vivo* testing will have been in vain. If in the European Union (EU) only 200–400 pharmaceuticals/year were progressed after giving false positive results, the additional animal testing would be estimated to require around 5000–10,000 rodents/year. In some cases the positive *in vitro* genotoxicity results may trigger the conduct of rat and/or mouse carcinogenicity studies (at least 500 rodents/species study) on compounds that would not otherwise be subjected to carcinogenicity testing, causing delays in development of up to 3 years. Alternatively, the positive results may trigger the conduct of carcinogenicity tests much earlier than normal in the development process, or trigger further chronic *in vivo* tests that are believed to specifically detect genotoxic carcinogens (e.g. using transgenic tumour models), thus resulting in even more animal usage. More accurate, predictive *in vitro* tests for genotoxicity (i.e. less false positives) could significantly reduce the number of animals used. In addition, more accurate and reliable *in vitro* tests may ultimately mean less reliance on or need for data from *in vivo* genotoxicity and carcinogenicity tests.

In order to address the high rate of false positive results (particularly in mammalian cells) a 2-day workshop was hosted and sponsored by the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy from 26–28 April 2006. More than 20 genotoxicity experts from academia, government and industry were invited to contribute their experiences. The objectives of the workshop were:

- To discuss data from the currently available mammalian cell genotoxicity test systems, to see whether it is possible to select cells and systems that would likely give fewer false positive results (i.e. show the highest specificity).
- To review modifications to existing protocols and mammalian cell genotoxicity test systems, as well as established but less widely used models, to identify changes that could likely reduce the frequency of false positives, and to define the experimental needs to implement these modifications.
- To discuss the performance of some new test systems that show promise of acceptable sensitivity but with improved specificity, to define which new test systems show sufficient promise for further development, and to define the experimental needs for that development.

Several participants were concerned that the term "false positive" was related to the ability of a chemical to induce tumours in rodent carcinogenicity studies, and that the relevance of rodent carcinogenicity for human health is in many cases questionable. It was acknowledged that we should not necessarily expect non-genotoxic carcinogens to give positive results in *in vitro* and/or *in vivo* genotoxicity tests, and if they do, then it may not be representative of their mechanism of action. In addition, some carcinogens may have been misclassified as genotoxic carcinogens on the basis of false positive results from *in vitro* genotoxicity tests. It was also acknowledged that assessment of the

performance of genotoxicity tests in detecting human carcinogens would be preferable. However, "false positives" in genotoxicity tests would be judged against non-carcinogens in humans, and whilst there is a recognised list of human carcinogens there is no such list of human non-carcinogens. It was therefore recommended that, for purposes of judging the performance of *in vitro* genotoxicity tests, the following groups of chemicals need to be identified:

- Chemicals that are *in vivo* genotoxins and DNA-reactive, mutagenic rodent carcinogens.
- Chemicals that are not genotoxic in at least two *in vivo* tests, and induce tumours via a non-DNA-reactive, non-mutagenic mechanism.
- Chemicals that are *in vivo* genotoxins but not carcinogenic, yet whose genotoxicity may be a relevant risk for human health.
- Chemicals that are neither rodent carcinogens nor genotoxic in at least two *in vivo* tests.

2. Summaries of presented material

Relevant information from the presentations given by various participants, that is pertinent to any decision-making, is summarised below under convenient headings.

2.1. Cell culture

Halliwell's group in Singapore has published a number of papers indicating the potential of cell culture media to oxidise a wide range of chemicals (including flavonoids and thiols) to produce hydrogen peroxide [3,4]. Hydrogen peroxide is a clastogen, and therefore certain levels of peroxide produced as a result of oxidation by media could lead to chromosomal aberrations and small colony mouse lymphoma mutants, particularly in the absence of exogenous metabolic activation (rat liver S9). Kirkland from Covance presented some of Halliwell's data, including some from a recent investigation of a wide range of media. Oxidation of ascorbic acid by most commonly used tissue culture media (e.g. DMEM, RPMI, McCoy's, William's E medium, etc.) was seen although the levels were much lower with Ham's F10 and F12 media [5]. However, the levels of hydrogen peroxide produced by oxidation of epigallocatechin gallate (EGCG) rapidly exceeded levels reported by Santoro et al. [6] to be clearly clastogenic in CHO cells *in vitro*. Again the peroxide levels with Ham's F10 and F12 were much lower. Thus, there is a risk of clastogenic damage arising with some compounds, not as a result of being DNA reactive, either directly or after metabolic activation, but as a result of the generation of reactive oxygen species through oxidation of the test compound by the culture medium. A comparison of the genotoxicity of various chemicals that are oxidised in different media would be useful to explore the consequences of these observations. The participants agreed that it would be wise to check for the ability of the proposed culture medium to oxidise the test chemical (e.g. by measuring production of hydrogen peroxide) and, if this is occurring, select another cell type or another medium to minimise or eliminate this effect. The latter may require considerable work in order to establish that growth conditions, appropriate controls, etc. are acceptable in the new media, and it may be necessary to conclude that it is not technically possible to perform the test without the complication of uncontrolled oxidative stress.

2.2. Stability of cell lines

The need for good scientific practice in the handling of cell lines was emphasised. Several participants were concerned that a cell line in one laboratory may be phenotypically different from cells with the same "name" in another laboratory, and may not respond in the same way to mutagens. There has been widespread distribution of rodent cell lines (e.g. CHO, V79, CHL, L5178Y), and different phenotypes may be due to changes in the karyotype. The instability of rodent cell lines and its contribution to "spurious" positive results has been questioned before [7], but no definitive data exist to address this possible source of false positives. Kirkland presented data generated at Covance showing that the sensitivity of CHO cells to two established clastogens (azidothymidine and 4-NQO) *decreased* dramatically after long-term culturing (i.e. 51 passages) as might occur if a clone was transferred from one laboratory to another. During this time the modal chromosome number increased from 21 to 22, and the range of identifiable chromosomes increased from 24 (Fig. 1) to 30 (Fig. 2). Although this was not a study to investigate sources of false positive results, this demonstrates the instability of some of the rodent cells and the potential contribution of this process to misleading results. In terms of future developments, it will be important to know whether human cell lines (particularly the more highly differentiated lines such as HepG2) are phenotypically and functionally more stable than rodent cell lines over long-term culturing.

To minimise the chances of "genetic drift" the participants agreed that cell lines should only be obtained from reputable sources at known early passage, and should be grown for a minimal number of passages before a

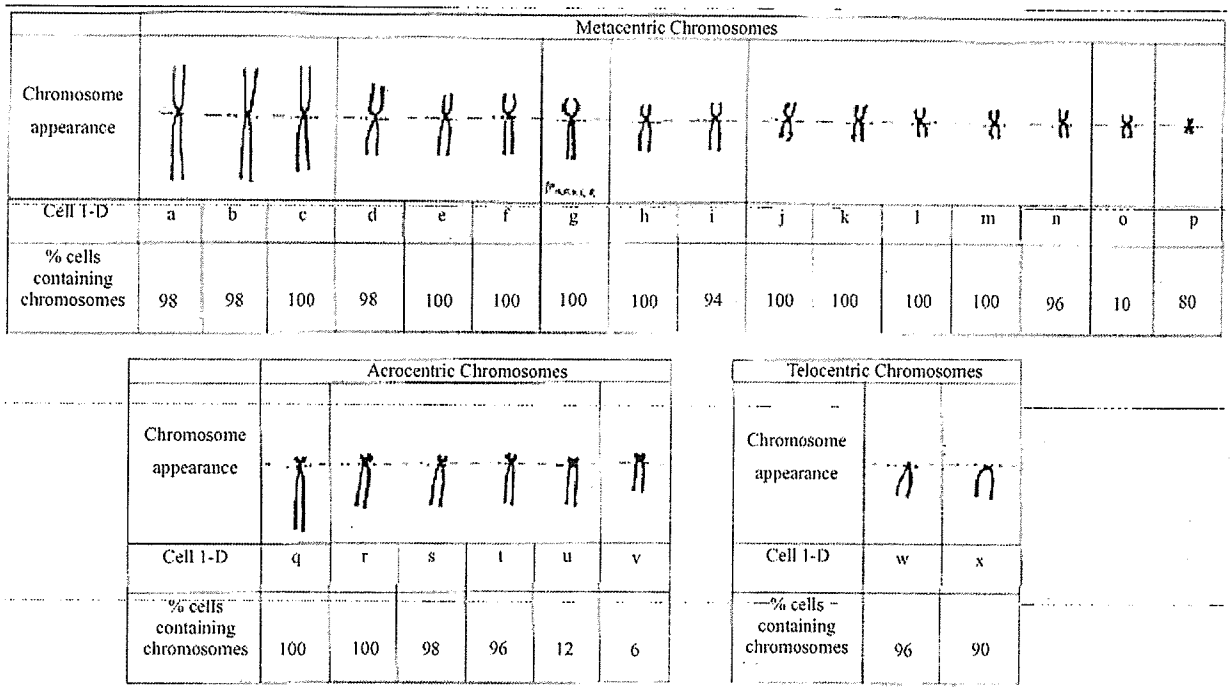

Fig. 1. Twenty-four chromosomes of different morphology identifiable in CHO-WBL clone at passage 11 (modal number 21).

quality-controlled "master stock" is frozen. A vial would be thawed, grown up for a small number of passages, and a "working stock" frozen. For each experiment a vial would be thawed from the working stock. When all the working stock is used up, another vial from the master stock would be thawed and grown to provide a replacement working stock of the same age (in terms of passages in culture) as the previous one. This process should allow

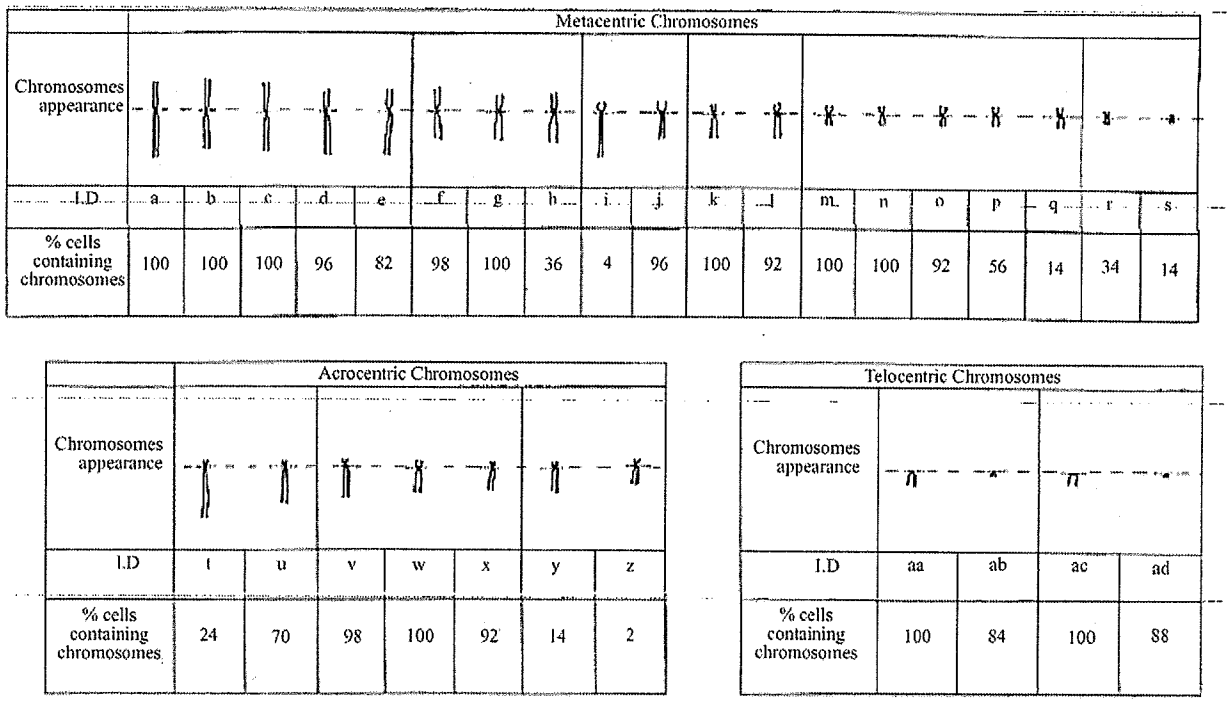

Fig. 2. Thirty chromosomes of different morphology identifiable in CHO-WBL clone at passage 51 (modal number 22).

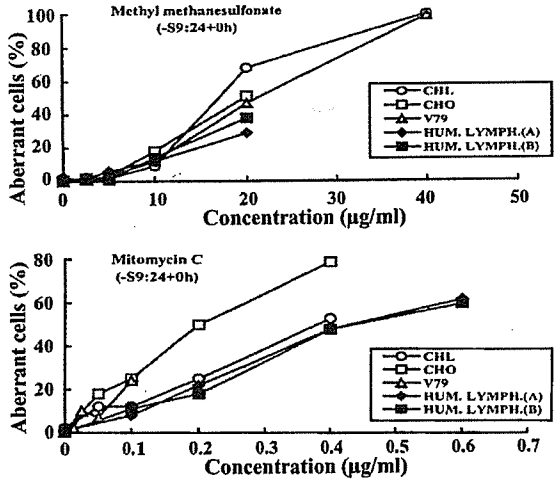

Fig. 3. Chromosomal aberration data for two chemicals tested in three different cell lines and peripheral blood lymphocyte cultures from two different donors at the same time.

for the master stock to support experiments for decades with cells of the same "age" in culture. This process works well for new cells, but we cannot be sure this was followed for existing cell stocks. It may be necessary to discard some cell stocks if the history is unclear.

ECVAM has already published recommendations for good cell culture practice [8] and has been heavily involved in drafting the OECD Advisory Document on the application of the principles of GLP to *in vitro* studies [9]. It was agreed that these recommendations should be re-examined in light of the experiences with genotoxicity tests, and the ECVAM recommendations updated if necessary.

2.3. Comparison of different cell types

2.3.1. CHO and CHL

A comparison of CHO and CHL cells on 25 compounds for which differing results had been obtained was conducted and published several years ago [10] and was summarised by Hayashi. The study had concluded that CHL cells might be more sensitive to the detection of clastogens than CHO cells, but that most differences in previously published responses to these chemicals were due to protocol differences, in particular to the length of treatment and sampling time. It is not possible to conclude from this data set whether CHL or CHO cells might be more susceptible to false positive results. Hayashi also showed (Fig. 3) that CHO, V79, CHL and human lymphocytes from two different donors, when experiments were conducted under identical conditions, gave very similar chromosomal aberration responses with MMS and Mitomycin C.

2.3.2. V79, L5178Y, TK6, human lymphocytes

Elhajouji presented Novartis data comparing screening test results for *in vitro* micronucleus (MN) induction in V79, L5178Y and TK6 cells, and induction of MN or chromosomal aberrations in full regulatory studies in human lymphocytes. Studies with V79 and L5178Y cells predicted all of the positive results that were subsequently obtained in human lymphocytes. In contrast to the data of Hayashi, 30–40% of chemicals tested (20/51 and 17/42, respectively) gave positive responses in V79 or L5178Y cells that were subsequently found to be negative when tested in human lymphocytes. It may be important that the human lymphocyte studies were not performed at the same time as the V79 and L5178Y studies. Detailed data from two chemicals were presented and one of these is shown in Fig. 4. The TK6 screening test, by contrast, failed to detect 1 of 10 chemicals that were subsequently positive when tested in human lymphocytes, but gave positive results that turned out to be negative in the human lymphocyte assay in only 9 of 52 chemicals tested. In most of the cases where the screening MN test over-predicted the human lymphocyte response, the treatment was a continuous (e.g. 20 h) exposure in the absence of S9 followed by a 24 h recovery. It was speculated that the *p53* deficient status of V79 and L5178Y cells might explain their inability to tolerate the toxic conditions imparted by long, continuous exposures, and could explain the lower frequency of positive results in TK6 cells (although they are deficient at DNA double strand break rejoining) and primary cultures of human lymphocytes.

2.3.3. L5178Y, human lymphocytes

By contrast, Kirchner from Roche presented data to show that screening for MN induction in L5178Y cells tended to underpredict (i.e. gave negative results in 6/27 cases) compounds that were subsequently positive for chromosomal aberrations in regulatory tests in human lymphocytes. However, the positives in the chromosomal aberration test mainly occurred at cytotoxic concentrations, and therefore the implications were that the screening MN test had correctly predicted lack of clastogenicity, and the chromosomal aberration test had, on some occasions, given a false positive response due to cytotoxicity. This is discussed further below.

2.3.4. BfArM submissions

A comparison of the positive and negative chromosomal aberration results in various cell types amongst data submitted to the German Federal Institute for Drugs and Medical Devices (BfArM) between 1995 and 2005 was made by Kasper. Data from 804 chromosomal aberra-

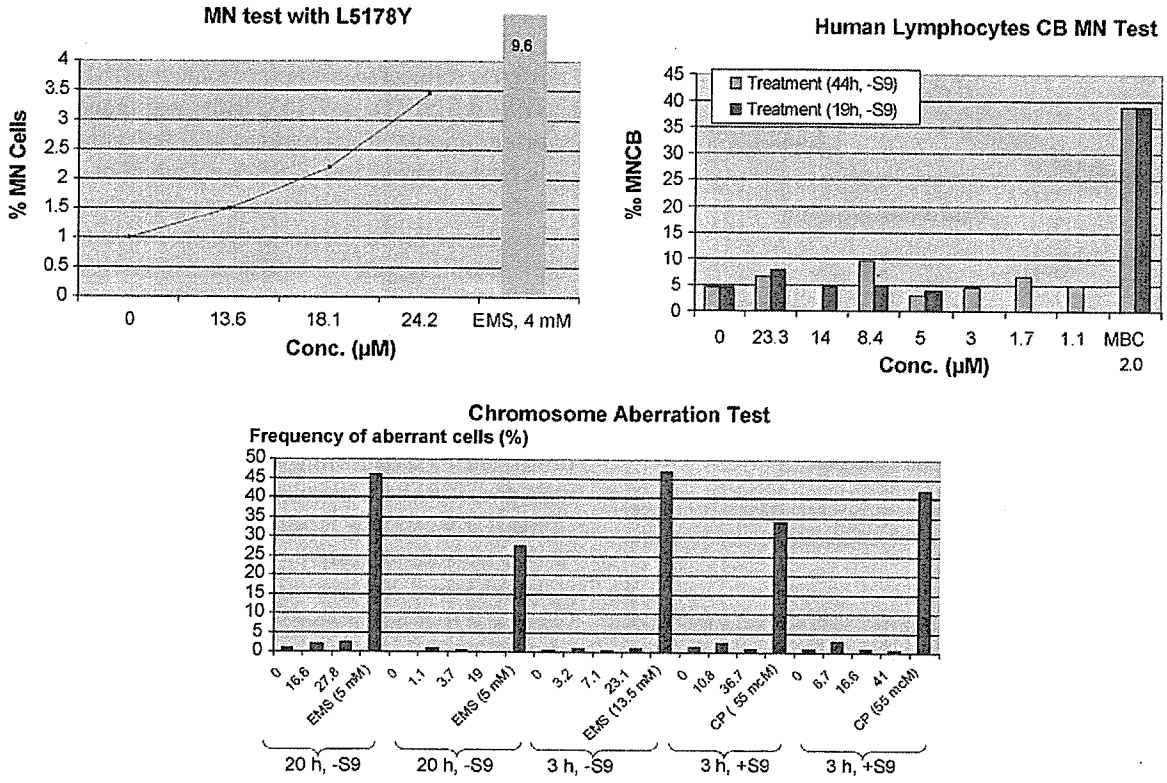

Fig. 4. Example from Novartis of a compound giving a positive micronucleus response when screened in L5178Y cells, but failing to induce either micronuclei or chromosomal aberrations in human lymphocytes at similar or higher concentrations. EMS, ethyl methanesulphonate; CP, cyclophosphamide; MBC, carbendazim.

tion studies on nearly 600 pharmaceuticals submitted to BfArM were reviewed. As shown in Fig. 5, the frequency of positive results in four different cell types studied for chromosomal aberrations and in the mouse lymphoma assay (detecting gene mutations as well as chromosomal damage) was very similar and averaged about 30%. It is interesting that such a high percentage of positive mammalian cell results is seen after companies have already screened out compounds that are not considered suitable for development. Although no significant differences in the frequency of positive results were seen amongst the five mammalian cell systems reviewed (Fig. 5), some differences were seen when the same compounds were tested in more than one mammalian cell system. Fig. 6

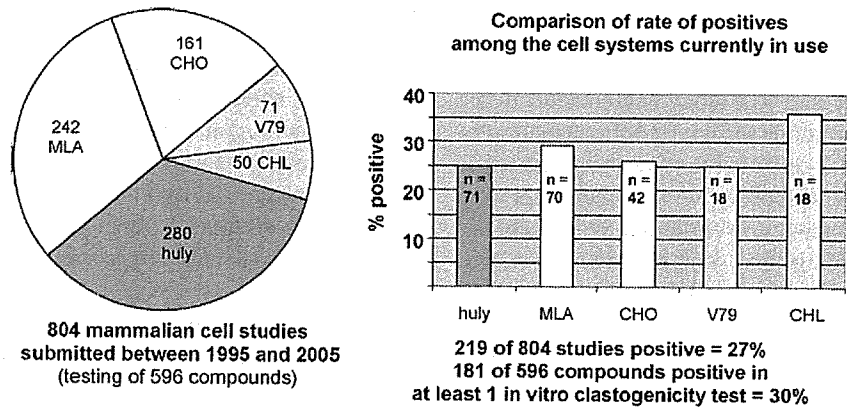

Fig. 5. Use of mammalian cell assays for regulatory submissions to the German Federal Institute for Drugs and Medical Devices (BfArM) between 1995 and 2005, and the frequency of positive results in human lymphocytes (huly), the mouse lymphoma assay (MLA) and the Chinese hamster cell lines V79, CHO and CHL.

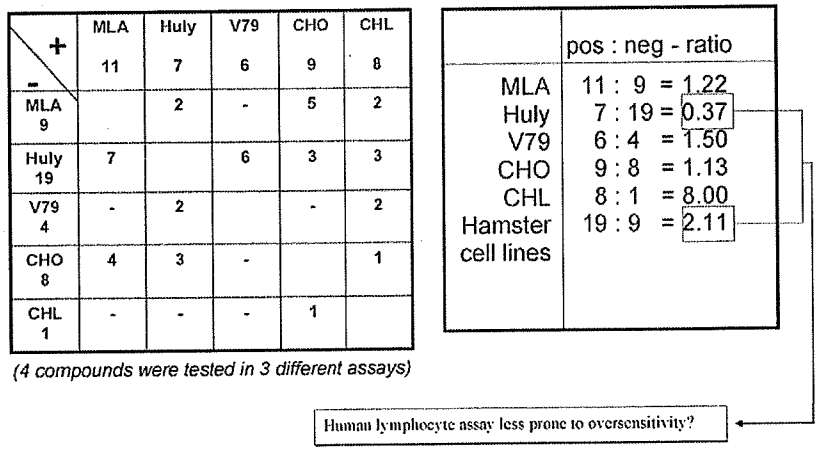

Fig. 6. Results for 37 pharmaceuticals submitted to BfArM between 1995 and 2005 with contradictory results within the mammalian cell assays. Mouse lymphoma assay (MLA) data were only included between 2000 and 2005. Huly, human lymphocytes. V79, CHO and CHL are Chinese hamster cell lines.

shows the responses for 37 compounds that gave contradictory results when tested in more than one mammalian cell system. Clearly, such compounds did not induce reproducible results across different mammalian cell systems and therefore the biological significance of the positive results might be questionable. Fig. 6 shows that human lymphocytes were most likely to give negative results when the mouse or Chinese hamster cell lines gave positive results. These were most likely to be clastogenicity results, but no distinction was made between large or small colony responses in the mouse lymphoma assay. Thus, it might be concluded that human lymphocytes are less prone to oversensitivity and therefore less prone than standard rodent cell lines to false positive results. Kasper noted that most (127/181 or 70%) of the compounds that were positive in one or more of the mammalian cell tests (cell lines and primary human lymphocytes) were uniquely positive, i.e. there were no supporting positive findings from the Ames test or rodent bone marrow micronucleus or chromosomal aberration tests. As such the majority of mammalian cell positives were considered non-relevant for *in vivo* genotoxicity. Kasper therefore raised the intriguing question as to whether it is the endpoint of "clastogenicity *in vitro*", particularly when associated with extensive cytotoxicity, which is prone to non-relevant positives irrespective of which cell model is used.

As a result of the above presentations there was some discussion as to the need for *in vitro* mammalian cell assays to be included in regulatory submissions. One suggestion was to focus on Ames-negative carcinogens, to determine their mechanism of action, decide if these are "important carcinogens to detect" (i.e. expected to represent a human risk), and then decide what genotoxicity tests are needed to detect them. The question was raised whether any DNA-reactive, mutagenic carcinogens give false negative results in the Ames test and therefore additional mammalian cell tests would be needed. If the requirement for *in vitro* mammalian cell assays continues, then there is a need to select better test systems, be more critical of the test conditions, and understand better the relevance of the results.

2.4. Cytotoxicity and cytotoxic mechanisms

As mentioned above, Kirchner from Roche presented data that suggested cytotoxicity is a major contributor to false positive results in clastogenicity assays. Greenwood et al. [11] showed that, in cytogenetic assays, measurement of reduction in population doubling (PD) to identify the 50% toxic concentration could avoid some of the cytotoxic positives that could occur if the 50% toxic concentration was chosen by reduction in cell count or mitotic index. Data from Kirchner shown in Table 1 similarly reveal that it is possible for relative cell count, mitotic index and reduction in population doubling to give quite different concentrations for 50% toxicity, and therefore selection of the top concentration to be tested should be done more carefully. Mitotic index becomes a very inaccurate measure of toxicity when there is an increase in mitotic activity, possibly through effects of the test chemical on spindle structure and function. Cell growth reduction may result from a number of different mechanisms, e.g. apoptosis, necrosis, cell cycle delay, mitotic block, etc., and these need to be distinguished if the true impact of a cytotoxic mechanism on an *in vitro* clastogenicity result is to be appreciated. The participants agreed that a thorough

Table 1
Example from Roche of the different toxicity profiles seen when relative cell count (RCC) mitotic index and population doubling (PD) are measured with the same concentration series of a test chemical

Treatment	Concentration (μg/ml)	% MN cells	Relative cell count on day 1 (%)	Mitotic index (%)	Number of population doublings
MMS	15	4.80	82	5.10	1.008
1% DMSO	–	0.50	100	3.80	1.288
Compound X	0.10	0.30	107	5.60	1.381
	0.25	1.00	99	10.30	1.274
	0.5	1.60	75	10.50	0.878
	0.75	4.00	35	18.20	−0.236
	5.00	3.30	35	51.30	−0.207
	10.00	1.00	35	134.40	−0.236

comparison of different measures of toxicity is needed such that the most appropriate measures can be recommended, and additional observations on the impact of apoptosis and necrosis on the genotoxicity result can be made.

One approach to assessing (or excluding) the involvement of apoptosis in genotoxicity may be to use mouse CTLL-2 cells with and without transfection with the human *bcl2* gene. Marzin from Institute Pasteur, who has published on these cells [12], presented data to show how "true" genotoxins such as methyl methanesulphonate, ethyl methanesulphonate, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine, methylnitrosourea, benzo[*a*]pyrene, 7,12-dimethylbenz[*a*]anthracene and phytoestrogens such as genistein, topoisomerase inhibitors such as etoposide, and aneugens such as griseofulvin and nocodazole induce MN in the absence of apoptosis. An example is shown in Fig. 7. Cytotoxic compounds such as anisomycin C, curcumin and dexamethasone only induced MN in the presence of apoptosis. An example is shown in Fig. 8. Although the published data with these cells are impressive, they have been generated only in one laboratory. It is understood that attempts to demonstrate the reproducibility of these effects between laboratories have not yet been completed.

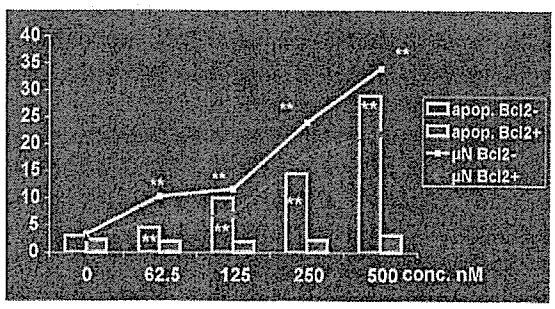

Fig. 7. Induction of micronuclei (μN) in the absence of apoptosis by etoposide in CTLL-2 cells with and without the human *bcl2* gene.

2.5. Metabolic considerations

The importance of metabolism in the activation of many *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens cannot be overstated. However, different carcinogens are activated by different CYP and non-CYP enzymes and yet there is almost universal use of a single metabolic activation system (Aroclor 1254-induced rat liver S9) for all *in vitro* genotoxicity tests. Metabolites produced by this S9 may be quite different from those produced by normal human liver metabolism. The induction by Aroclor-1254 leads to over-representation of the CYP 1A and 2B enzymes compared to other hepatic CYP forms, as shown in Table 2 (presented by Glatt from Potsdam). Phase 2 enzymes are essentially inactive in standard S9, as their cofactors are not added, unlike NADPH, the cofactor for CYPs.

Glatt presented results from *hprt* gene mutation tests using V79-derived cell lines engineered for various enzymes. Standard carcinogens (e.g. dimethylnitrosamine, benzo[*a*]pyrene, dibenzo[*a,l*]pyrene, aflatoxin B_1, 2-aminoanthracene, 2-acetylaminofluorene, IQ and PhIP) showed strong mutagenicity even at extremely low concentrations (0.05–500 nM, depending

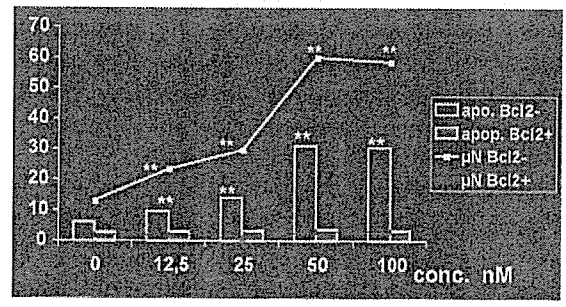

Fig. 8. Induction of micronuclei (μN) only in the presence of apoptosis by dexamethasone in CTLL-2 cells not expressing the human *bcl2* gene.

Table 2
The impact of Aroclor on the induction of various CYPs in comparison to normal rat and human liver[a]

Enzyme	CYP (nmol/mg microsomal protein)		Induction factor (rat)	Level in human liver
	Untreated rat	Aroclor-treated rat		
CYP1A1	0.04	1.45	36	0
CYP1A2	<0.03	1.23	>41	0/+[b]
CYP2B1	0.03	1.29	43	+
CYP2B2	0.07	1.46	21	
CYP2C6	0.36	0.36	1	++
CYP2C11	1.20	0.27	0.23	
CYP2D1	0.15	0.15	1	0/+[c]
CYP3A	0.39	0.77	2	+++

[a] From Guengerich et al. [13].
[b] Depending on induction state.
[c] Depending on genotype.

on the compound) when cells with appropriate enzyme systems were used (Table 3). In contrast, some of these carcinogens (e.g. 2-acetylaminofluorene) were not mutagenic in standard tests conducted in the parental cell lines in the presence of S9. Other carcinogens (dimethylnitrosamine, benzo[*a*]pyrene, dibenzo[*a*,*l*]pyrene, aflatoxin B_1, 2-aminoanthracene) required 700- to 25,000-fold higher substrate concentrations in the standard test compared to the metabolically engineered cell models. At least two mechanisms underlie the highly increased sensitivity of the engineered cells:

(a) For all mentioned compounds, except aflatoxin B_1, enzymes that were involved in the activation in the recombinant cells, are either inactive (sulfotransferases [SULTs] and acetyltransferases [NATs]) or very low (CYP1B1 and CYP2E1) in S9.
(b) A much smaller portion of the active metabolite may reach the target structure when it is generated by external enzyme systems as opposed to within the target cell. In general, membrane permeation will be particularly low, or even nil, with extremely short-lived and/or ionised (Phase 2) metabolites.

Glatt also noted that it is not sufficient to have any kind of CYP and/or NAT and/or SULT present for the activation of a given promutagen. In general, very specific forms of these enzymes, which vary depend-

Table 3
Concentrations of standard mutagens required to obtain a positive result in *hprt* gene mutation assays using standard liver S9 or cDNA expressed enzymes in target cells for the activation

Test compound	Engineered cell lines		Standard test using S9, concentration required (μM)[b]
	Expressed enzymes[a]	Concentration required (μM)[b]	
Dimethylnitrosamine	hCYP1E1 – hSULT1A1	0.5	3000[c]
Benzo[*a*]pyrene	hCYP1B1	0.01	7, 8.3[c]
Dibenzo[*a*,*l*]pyrene	hCYP1B1	0.00005	1
Aflatoxin B_1	hCYP1A2	0.002	3, 0.5[c]
2-Aminoanthracene	hCYP1A2 + chNAT[d]	0.002	50
2-Acetylaminofluorene	rCYP1A2 + rSULT1C1	0.1	−(600)[e]
IQ	hCYP1A2 + hNAT2	0.02	Not tested
PhIP	hCYP1A2 + hSULT1A1	0.5	Not tested

[a] h, human; r, rat; ch, Chinese hamster; mutagenicity was completely abolished in cell lines missing any of the indicated enzymes, except that hSULT1A1 was not required for the activation of dimethylnitrosamine (but it enhanced the expression of hCYP2E1 via an unknown mechanism).
[b] Concentration required for increasing the mutant frequency by $20/10^6$ cells above the spontaneous level (usually $1–10/10^6$ cells). Some values are calculated from effects observed at higher concentrations using linear extrapolation). Unless specified otherwise, data from Glatt laboratory.
[c] Data from review of Bradley et al. [14]—concentration leading to a 10-fold increase in mutant frequency [a criterion that is similar to that used by Glatt (footnote b)].
[d] Endogenous enzyme expressed in some sublines of V79 (e.g. V79-NH).
[e] Negative test result, highest concentration used in parenthesis.

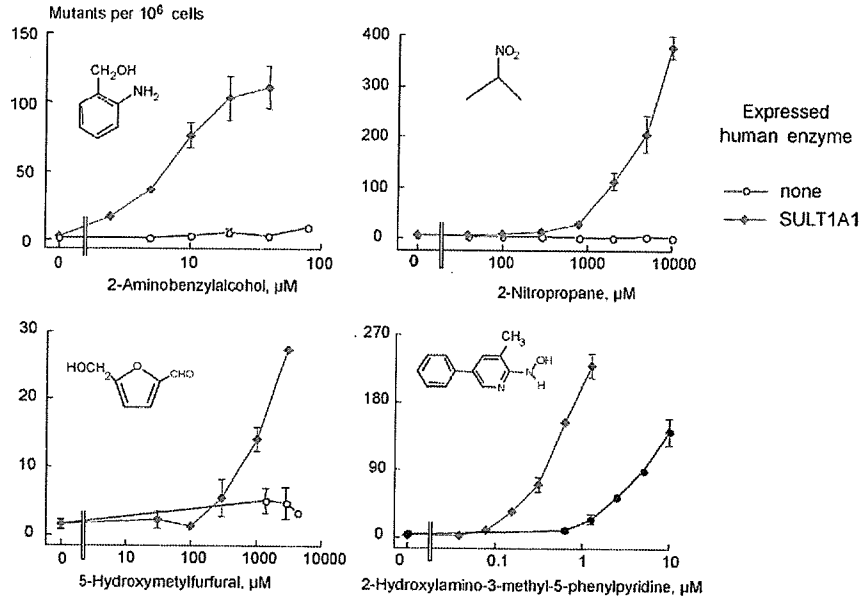

Fig. 9. Promutagens activated by human sulphotransferase SULT1A1 in recombinant V79 cells.

ing on the promutagen, are required for activation at low substrate concentrations. According to Glatt it is likely that these high-affinity activations are those that determine the bioactivation *in vivo*. Based on his findings, Glatt developed his concept of "promiscuous activation". For numerous mutagens, the decisive *in vivo* enzymes are missing *in vitro*. However, if the substrate concentration is increased sufficiently, some other enzymes (that are unimportant *in vivo*) may take over the activation—leading to the same or a different active metabolite. Since we often do not use the right enzyme systems for positive controls *in vitro*, we have to rely on their promiscuous activation, i.e. to use excessive concentrations to get an empirical correlation between genotoxicity and carcinogenicity. The situation is worsened by the low efficiency of external activation. However, if excessive concentrations are needed for the positive controls, such high concentrations have also to be used with test chemicals. In this case, promiscuous activation (or any other high concentration effects not requiring activation) that does not occur *in vivo* is less welcome, as it may lead to false positive results. Glatt suspects that in general, relevant *in vivo* genotoxicants would be detected positive *in vitro* at concentrations of less than 100 μM if the true activation mechanisms were taken into account.

Glatt also emphasised the importance of non-CYP enzymes in the activation of many genotoxicants. This activation is largely underestimated by many genetic toxicologists. For example, Glatt found more than 100 promutagens that are activated by SULTs (examples in Fig. 9) [15,16]. SULTs are not endogenously expressed in V79 cells or any other bacterial or mammalian target cells of standard *in vitro* tests. SULTs are inactive in S9 due to the lack of cofactor. Addition of the corresponding cofactor is not a reliable remedy, since sulfo conjugates are charged and therefore do not reliably penetrate into target cells, especially if they are short-lived. Unless an alternative (sometimes promiscuous) activation pathway exists, SULT-dependent mutagens are missed in standard *in vitro* test systems. This problem is not unique for SULTs, but may extend to other classes of non-CYP enzymes.

Darroudi from Leiden described the application and validation of human HepG2 cells *in vitro* for detecting different classes of human dietary mutagens and antimutagens. Attempts were made also to define the metabolic capabilities of the HepG2 cell line. Various housekeeping genes (porphobilinogen deaminase, *hprt*, ATP-synthetase, glyceraldehyde-3-phosphate dehydrogenase, elongation factor-1-alpha) are expressed equally in HepG2 cells and primary human hepatocytes. Various CYPs and some Phase 2 enzymes (e.g. UGT and NAT) are also constitutively expressed. Moreover, as shown in Fig. 10, they can be induced by similar factors in HepG2 cells as in human hepatocytes following treatment with benzo[*a*]pyrene [17]. This means that some carcinogens which are difficult or impossible to detect using induced rat liver S9 preparations (e.g. safrole, hexamethylphosphoramide) can be detected as inducing genotoxicity in HepG2 cells, or can induce genotoxicity in CHO cells and Ames bacteria when S9 is prepared