on Environmental Mutagens in 1993. Several working groups addressed the methodology of test systems in widespread use for scientific research and regulatory practice. The consensus recommendations from those working groups were published in Mutation Research in 1994 (volume 312, pages 195–318), and had a significant impact on revisions to OECD guidelines for genotoxicity testing that were on-going at that time [1]. These IWGT discussions and conclusions were also recognised as valid contributions for the ICH guidances for pharmaceuticals for human use that were finalised in 1995 and 1997 [2,3].

At the second IWGT Workshop, held in Washington, DC in 1999, recommendations for the mouse lymphoma *tk* mutation assay and the *in vivo* micronucleus test were updated and some recommendations were made that were different from or extended beyond the published OECD guidelines for these test systems. In addition, several assays for which no OECD guidelines existed were discussed in detail and recommendations for their conduct made. These reports were published in Environmental and Molecular Mutagenesis in 2000 (volume 35, pages 159–263).

Recommendations for the conduct of the mouse lymphoma *tk* mutation assay, the *in vitro* micronucleus test and for *in vivo* transgenic mutation assays were updated at the third IWGT Workshop in Plymouth, UK in 2002 and published in Mutation Research in 2003 (volume 540, pages 119–181). Guidance was also introduced on the usefulness and techniques for molecular analysis of tumour and non-tumour tissues from carcinogenicity studies with the transgenic haploinsufficient p53 and *Ras*H2 mouse strains [4]. An important aspect of this meeting was the initiation of discussions and recommendations on strategic approaches to genotoxicity testing [5].

The published IWGT reports that have extended recommendations for testing beyond existing OECD guidelines are:

- Advances in conduct of the *in vivo* micronucleus test [6].
- Advances in conduct of the mouse lymphoma *tk* mutation assay [7–9].

For the *in vivo* micronucleus test (OECD Guideline 474) the key new recommendations are:

- Repeat dose studies (e.g. for 28 days) can be conducted in the rat, where the bone marrow is still the principal tissue but additional data can be obtained from scoring micronuclei (MN) in peripheral blood.
- Modern automated scoring methods such as flow cytometry and image analysis, allowing much greater numbers of cells to be scored, are acceptable, if the utility of the analytical system has been demonstrated in the performing laboratory.
- It is not necessary to treat a concurrent positive control group with every study, particularly in laboratories that run the test frequently, and when test chemical dosing solutions have been well characterised and systemic exposure has been demonstrated. Control of staining and scoring can be achieved by coding positive control slides from a previous GLP study into the current study. However, until more experience is gained, it is recommended to include positive controls in repeat dose studies (e.g. for 28 days).
- CREST antibodies and FISH with pancentromeric or chromosome-specific probes can be used under certain circumstances to distinguish clastogens from aneugens. Positive controls should be included with these types of investigations.
- Micronuclei can be measured in liver, colonic epithelium, skin, spleen, lung, spermatids and in foetal/neonatal tissue, although concurrent positive control treatments will be required with these tissues, at least until the methods become more widely established and published.

For the mouse lymphoma assay (OECD Guideline 476), which has become the preferred mammalian cell mutation test in many regulatory documents [3,10–12] the key new recommendations are:

- The IWGT Working Group confirmed as useful the recommendations made by ICH [3] that a 24 h continuous treatment in the absence of S9 should be conducted when short treatments give negative results.
- Relative total growth (RTG) is the recommended measure of toxicity.
- Alternative suggestions for positive control chemicals were made, and a recommendation made that positive control treatments should demonstrate adequate detection of small colony mutants.
- Acceptable negative (solvent) control mutant frequencies for the agar and microwell methods have been defined.
- No single statistical test was identified as appropriate for evaluation of responses as positive or negative, and a Global Evaluation Factor has been defined for both agar and microwell versions of the test to be used in conjunction with other statistical methods (e.g. dose-response) in evaluation of results.

In addition to these state-of-the-art updates on existing OECD methods, the following reports have been published for test systems for which no OECD guidelines currently exist:

- Photochemical genotoxicity [13].
- Single cell gel electrophoresis (Comet) assay *in vitro* and *in vivo* [14 and this issue].
- DNA adduct determination [15].
- *In vitro* micronucleus test [16,17].
- *In vivo* transgenic mutation assays [18,19].

3. The Fourth IWGT Workshop

The Fourth IWGT Workshop was recently held in San Francisco, USA as a satellite to the 2005 International Conference on Environmental Mutagens (ICEM). The majority of discussions and recommendations were in the area of strategic use of genotoxicity tests, but some new recommendations for methods were also made. The Working Group reports from this Workshop are published elsewhere in this issue.

Since OECD and ICH guidances (guidelines) constitute the two major sets of internationally harmonised genotoxicity guidelines in regulatory use, it is hoped that recommendations made by IWGT working groups are of particular help in supplementing test design and interpretation of genotoxicity test packages that are based on these guidelines. They may serve as a basis to open discussions for the revision of OECD test guidelines and the maintenance of ICH S2 guidance.

References

[1] OECD, OECD Guidelines for the Testing of Chemicals: Genotoxicity. Organisation for Economic Co-operation and Development, Paris. Revised and new guidelines adopted (1997).

[2] ICH, Topic S2A. Genotoxicity: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Harmonised tripartite guideline CPMP/ICH/141/95, approved September 1995.

[3] ICH, Topic S2B. Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Step 4 guideline, July 16, 1997.

[4] R.D. Storer, J.E. French, L.A. Donehower, D. Gulezian, K. Mitsumori, L. Recio, R.H. Schiestl, F.D. Sistare, N. Tamaoki, T. Usui, H. van Steeg, Transgenic tumor models for carcinogen identification: the heterozygous Trp53-deficient and *Ras*H2 mouse lines, Mutat. Res. 540 (2003) 165–176.

[5] L. Müller, D. Blakey, K.L. Dearfield, S. Galloway, P. Guzzie, M. Hayashi, P. Kasper, D. Kirkland, J.T. MacGregor, J.M. Parry, L. Schechtman, A. Smith, N. Tanaka, D. Tweats, H. Yamasaki, Strategy for genotoxicity testing and stratification of genotoxicity test results—report on initial activities of the IWGT Expert Group, Mutat. Res. 540 (2003) 177–181.

[6] M. Hayashi, J.T. MacGregor, D.G. Gatehouse, I.-D. Adler, D.H. Blakey, S.D. Dertinger, G. Krishna, T. Morita, A. Russo, S. Sutou, *In vivo* rodent erythrocyte micronucleus assay. II. Some aspects of protocol design including repeated treatments, integration with toxicity testing, and automated scoring, Environ. Mol. Mutagen. 35 (2000) 234–252.

[7] M.M. Moore, M. Honma, J. Clements, T. Awogi, G. Bolcsfoldi, J. Cole, B. Gollapudi, K. Harrington-Brock, A. Mitchell, W. Muster, B. Myhr, M. O'Donovan, M.-C. Ouldelhkim, R. San, H. Shimada, L.F. Stankowski Jr., Mouse lymphoma thymidine kinase locus gene mutation assay: International Workshop on Genotoxicity Test Procedures Workgroup report, Environ. Mol. Mutagen. 35 (2000) 185–190.

[8] M.M. Moore, M. Honma, J. Clements, G. Bolcsfoldi, M. Cifone, R. Delongchamp, M. Fellows, B. Gollapudi, P. Jenkinson, P. Kirby, S. Kirchner, W. Muster, B. Myhr, M. O'Donovan, J. Oliver, T. Omori, M.-C. Ouldelhkim, K. Pant, R. Preston, C. Riach, R. San, L.F. Stankowski Jr., A.K. Thakur, F. Van Goethem, S. Wakuri, I. Yoshimura, Mouse lymphoma thymidine kinase gene mutation assay: International Workshop on Genotoxicity Tests Workgroup report—Plymouth, UK, 2002, Mutat. Res. 540 (2003) 127–140.

[9] M.M. Moore, M. Honma, J. Clements, G. Bolcsfoldi, B. Burlinson, M. Cifone, J. Clarke, R. Delongchamp, R. Durward, M. Fellows, B. Gollapudi, S. Hou, P. Jenkinson, M. Lloyd, J. Majeska, B. Myhr, M. O'Donovan, T. Omori, C. Riach, R. San, L.F. Stankowski Jr., A.K. Thakur, F. Van Goethem, S. Wakuri, I. Yoshimura, Mouse lymphoma thymidine kinase gene mutation assay: Follow-up Meeting of the International Workshop on Genotoxicity Tests—Aberdeen, Scotland, 2003, Environ. Mol. Mutagen. 47 (2006) 1–5.

[10] COM, Guidance on a Strategy for Testing of Chemicals for Mutagenicity. Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment. December 2000. http://www.doh.gov.uk/com/guidance.pdf.

[11] FDA, Office of Food Additive Safety. Redbook 2000. Toxicological Principles for the Safety Assessment of Food Ingredients. U.S. Food and Drug Administration. http://vm.cfsan.fda.gov/~redbook/red-toca.html (2000).

[12] TGD, Technical Guidance Document, Edition 2, December 2003. European Chemicals Bureau, Joint Research Centre, Italy. http://ecb.jrc.it/new-chemicals/documents (2003).

[13] E. Gocke, L. Müller, P.J. Guzzie, S. Brendler-Schwaab, S. Bulera, C.F. Chignell, L.M. Henderson, A. Jacobs, H. Murli, R.D. Snyder, N. Tanaka, Considerations on photochemical genotoxicity: Report of the International Workshop on Genotoxicity Test Procedures Working Group, Environ. Mol. Mutagen. 35 (2000) 173–184.

[14] R.R. Tice, E. Agurell, D. Anderson, B. Burlinson, A. Hartmann, H. Kobayashi, Y. Miyamae, E. Rojas, J.-C. Ryu, Y.F. Sasaki, Single cell gel/comet assay: guidelines for *in vitro* and *in vivo* genetic toxicology testing, Environ. Mol. Mutagen. 35 (2000) 206–221.

[15] D. Phillips, P.B. Farmer, F.A. Beland, R.G. Nath, M.C. Poirier, M.V. Reddy, K.W. Turteltaub, Methods of DNA adduct determination and their application to testing compounds for genotoxicity, Environ. Mol. Mutagen. 35 (2000) 222–233.

[16] M. Kirsch-Volders, T. Sofuni, M. Aardema, S. Albertini, D. Eastmond, M. Fenech, M. Ishidate Jr., E. Lorge, H. Norppa, J. Surrallés, W. von der Hude, A. Wakata, Report from the *in vitro*

micronucleus assay working group, Environ. Mol. Mutagen. 35 (2000) 167–172.
[17] M. Kirsch-Volders, T. Sofuni, M. Aardema, S. Albertini, D. Eastmond, M. Fenech, M. Ishidate Jr., S. Kirchner, E. Lorge, T. Morita, H. Norppa, J. Surrallés, A. Vanhauwaert, A. Wakata, Report from the *in vitro* micronucleus assay working group, Mutat. Res. 540 (2003) 153–163.
[18] J.A. Heddle, S. Dean, T. Nohmi, M. Boerrigter, D. Casciano, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.C. Mirsalis, H.-J. Martus, T.R. Skopek, V. Thybaud, K.R. Tindall, N. Yajima, *In vivo* transgenic mutation assays, Environ. Mol. Mutagen. 35 (2000) 253–259.
[19] V. Thybaud, S. Dean, T. Nohmi, J. de Boer, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.A. Heddle, R.H. Heflich, I. Lambert, H.-J. Martus, J.C. Mirsalis, T. Suzuki, N. Yajima, *In vivo* transgenic mutation assays, Mutat. Res. 540 (2003) 141–151.

D.J. Kirkland *
Covance Laboratories Ltd., Otley Road, Harrogate, HG3 1PY, UK

M. Hayashi
National Institute for Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

D. Jacobson-Kram
Office of New Drugs, Center for Drugs Evaluation and Research, US FDA, Silver Spring, MD, USA

P. Kasper
BfArM, Kurt-Georg-Kiesinger Allee 3, Bonn D-53175, Germany

J.T. MacGregor
Toxicology Consulting Services, 201 Nomini Drive, Arnold, MD 21012, USA

L. Müller
Safety & Technical Sciences, PRBN-T, Building 73/311b, F Hoffmann-La Roche AG, CH 4070 Basel, Switzerland

Y. Uno
Mitsubishi Pharma Co., Toxicology Laboratory, 1-1-1 Kazusakamatari, Kisarazu, Chiba 292-0818, Japan

* Corresponding author. Tel.: +44 1423 848401; fax: +44 1423 848983.

E-mail address: david.kirkland@covance.com (D.J. Kirkland)

28 February 2006
Available online 28 November 2006

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 0I2, 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.

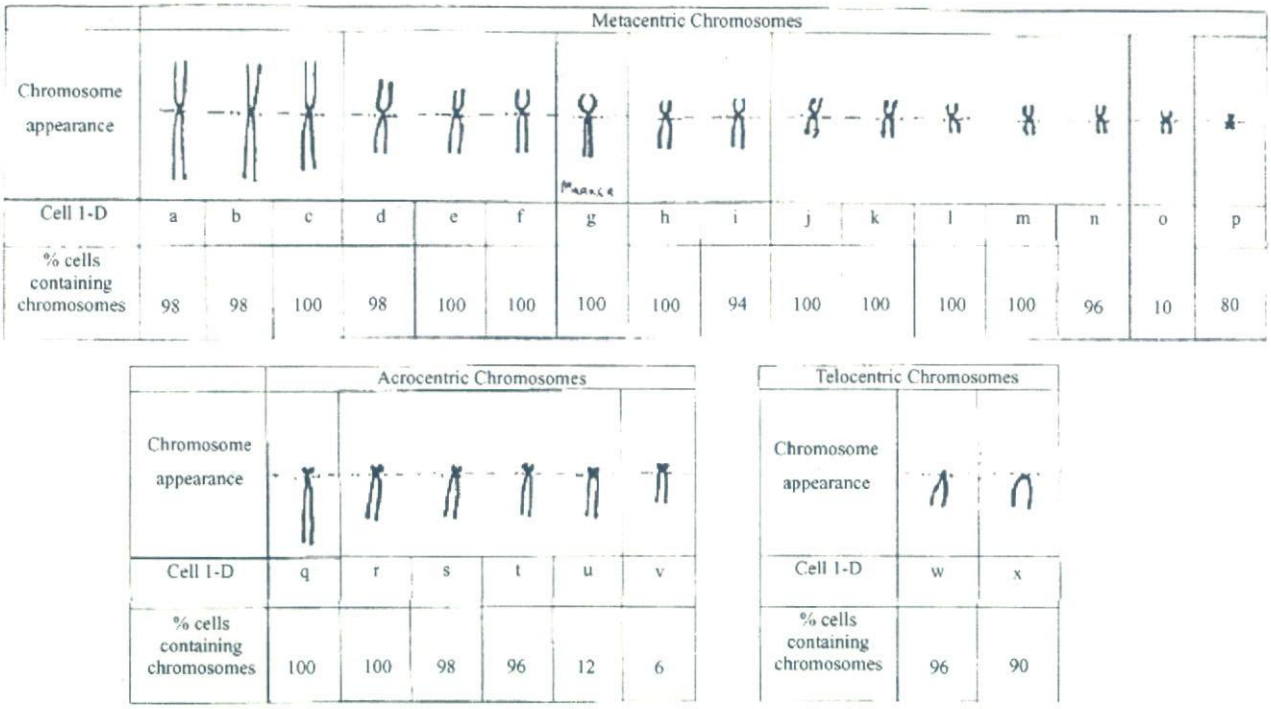
2.1. Cell culture

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



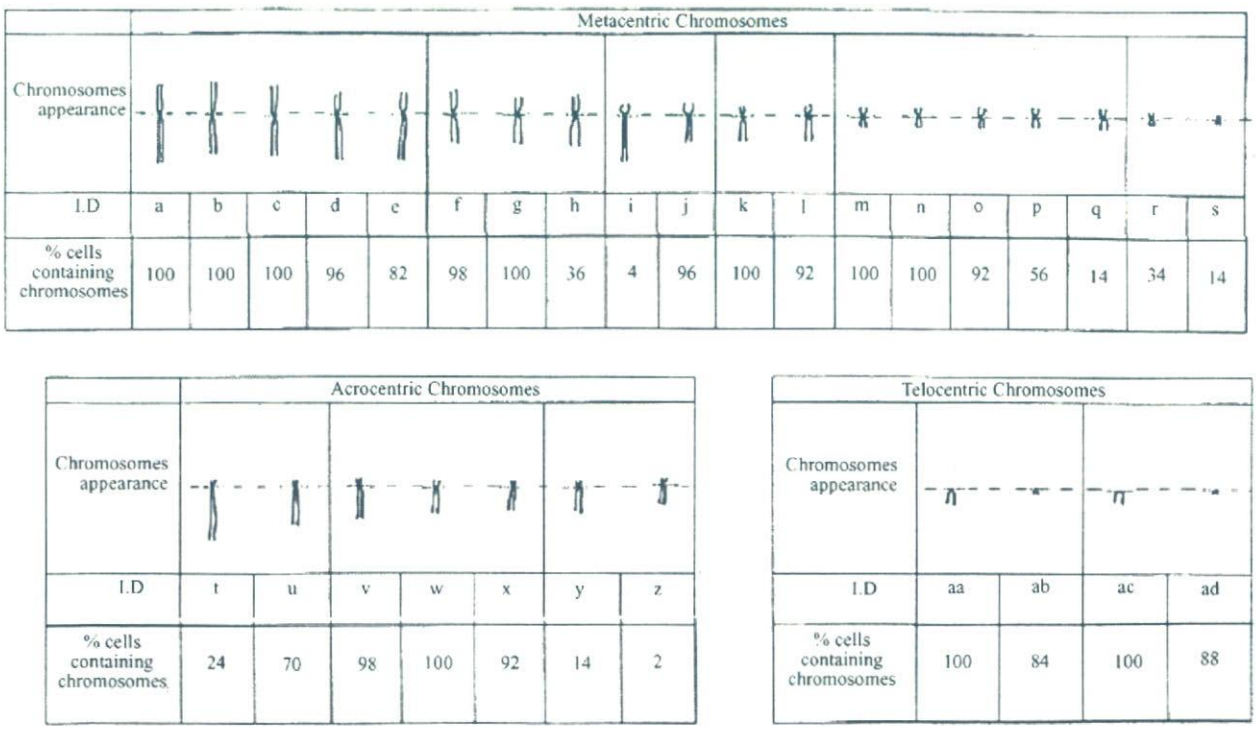
Metacentric Chromosomes															
Chromosome appearance															
Cell I-D	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
% cells containing chromosomes	98	98	100	98	100	100	100	100	94	100	100	100	100	96	10	80

Acrocentric Chromosomes						
Chromosome appearance						
Cell I-D	q	r	s	t	u	v
% cells containing chromosomes	100	100	98	96	12	6

Telocentric Chromosomes		
Chromosome appearance		
Cell I-D	w	x
% cells containing chromosomes	96	90

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



Metacentric Chromosomes																			
Chromosomes appearance																			
I.D	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
% cells containing chromosomes	100	100	100	96	82	98	100	36	4	96	100	92	100	100	92	56	14	34	14

Acrocentric Chromosomes							
Chromosomes appearance							
I.D	t	u	v	w	x	y	z
% cells containing chromosomes	24	70	98	100	92	14	2

Telocentric Chromosomes				
Chromosomes appearance				
I.D	aa	ab	ac	ad
% cells containing chromosomes	100	84	100	88

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

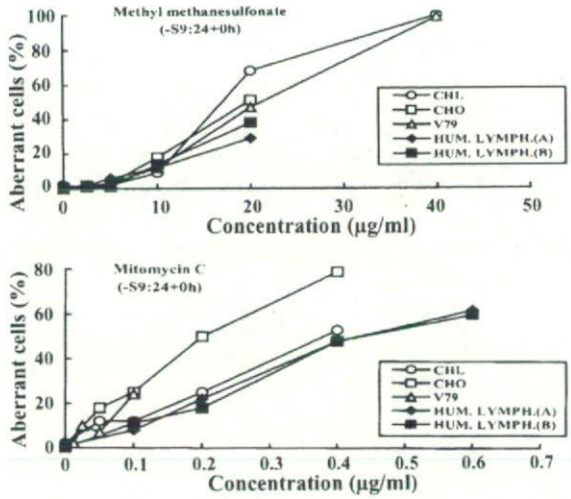


Fig. 3. Chromosomal aberration data for two chemicals tested in three different cell lines and peripheral blood lymphocyte cultures from two different donors at the same time.

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-

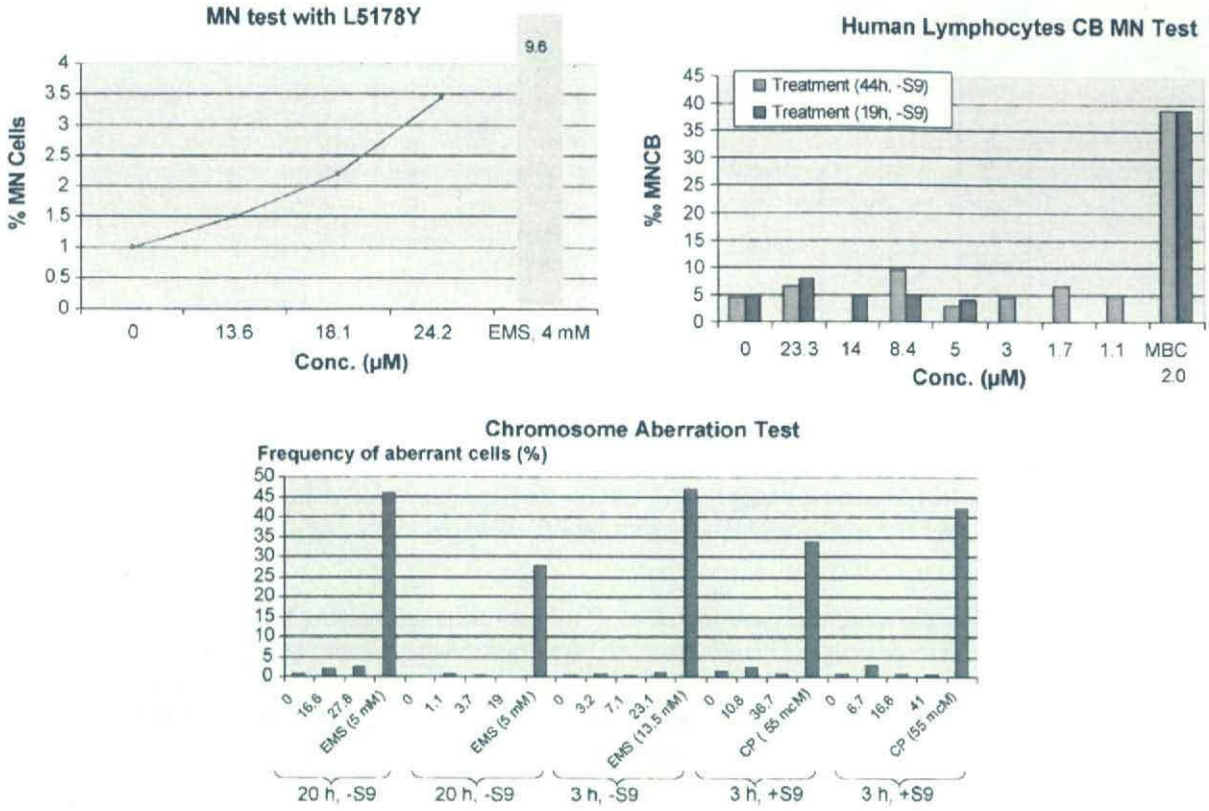


Fig. 4. Example from Novartis of a compound giving a positive micronucleus response when screened in L5178Y cells, but failing to induce either micronuclei or chromosomal aberrations in human lymphocytes at similar or higher concentrations. EMS, ethyl methanesulphonate; CP, cyclophosphamide; MBC, carbendazim.

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

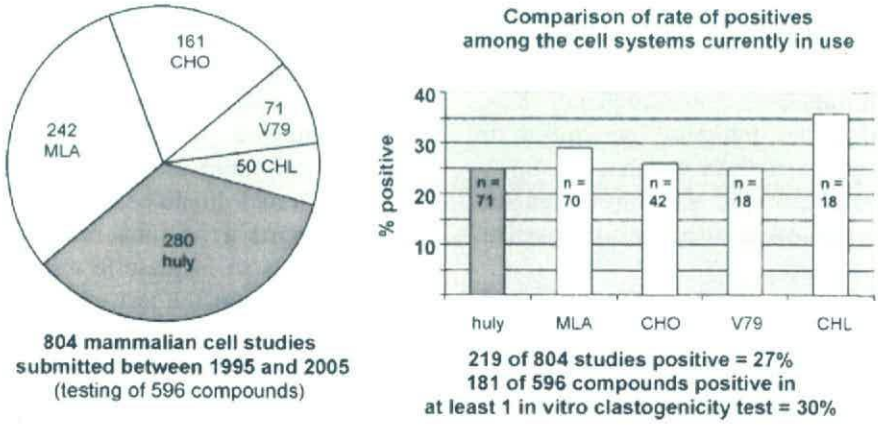


Fig. 5. Use of mammalian cell assays for regulatory submissions to the German Federal Institute for Drugs and Medical Devices (BfArM) between 1995 and 2005, and the frequency of positive results in human lymphocytes (huly), the mouse lymphoma assay (MLA) and the Chinese hamster cell lines V79, CHO and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

+ / -	MLA 11	Huly 7	V79 6	CHO 9	CHL 8
MLA 9		2	-	5	2
Huly 19	7		6	3	3
V79 4	-	2		-	2
CHO 8	4	3	-		1
CHL 1	-	-	-	1	

(4 compounds were tested in 3 different assays)

	pos : neg - ratio
MLA	11 : 9 = 1.22
Huly	7 : 19 = 0.37
V79	6 : 4 = 1.50
CHO	9 : 8 = 1.13
CHL	8 : 1 = 8.00
Hamster cell lines	19 : 9 = 2.11

Human lymphocyte assay less prone to oversensitivity?

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.

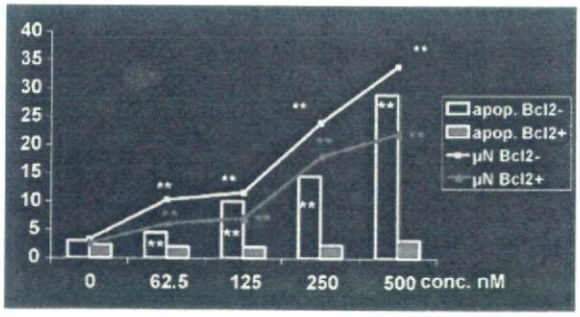


Fig. 7. Induction of micronuclei (μN) in the absence of apoptosis by etoposide in CTLL-2 cells with and without the human *bcl2* gene.

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

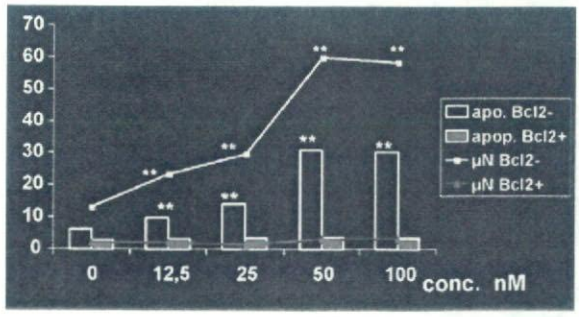


Fig. 8. Induction of micronuclei (μN) only in the presence of apoptosis by dexamethasone in CTLL-2 cells not expressing the human *bcl2* gene.

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.

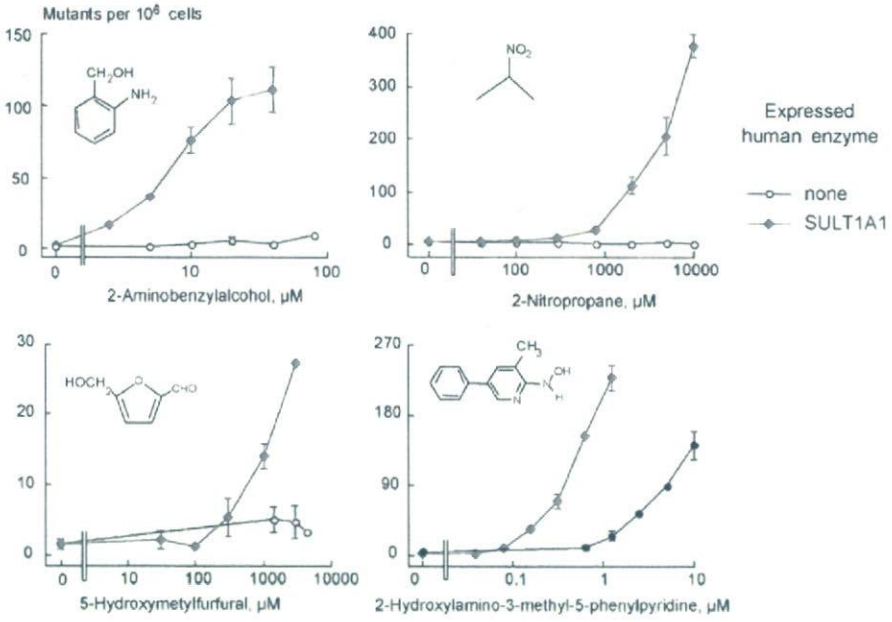


Fig. 9. Promutagens activated by human sulphotransferase SULT1A1 in recombinant V79 cells.

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

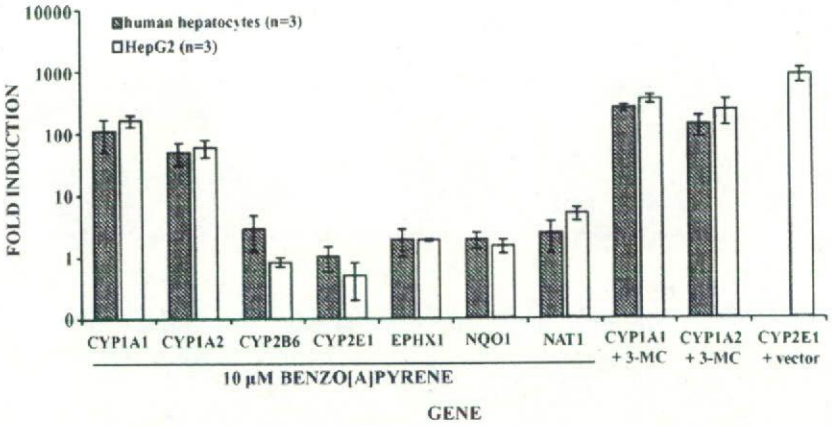


Fig. 10. Quantification of gene expression profiles in HepG2 cells and/or human hepatocytes treated with benzo[*a*]pyrene.

from HepG2 cells (Table 4). S9 prepared from HepG2 cells can be particularly useful for detecting carcinogenic heterocyclic amines like IQ and MeIQx because they require acetylation, and this is poor in Chinese hamster cell lines such as CHO. However, even when activated by S9 from HepG2 cells, the lack of penetration of the acetic acid ester metabolites into cells can be a problem. The HepG2 cell system has also been used to elucidate the genotoxic potential of a series of mycotoxins known to be carcinogens *in vivo* but so far (except for aflatoxin B_1) always reported to be negative in all *in vitro* assays studied. For the first time a ranking order could be established for the genotoxic potential of different mycotoxins, such as fumonisin B_1, citrinin and ochratoxin A [18]. Interestingly ochratoxin B, which is structurally related to ochratoxin A but is a non-carcinogen, revealed no genotoxic potential, and aflatoxin B_1 ranked as the most genotoxic. Recently, modulation of gene expression and DNA adduct formation in HepG2 cells following treatment with different classes of polycyclic aromatic hydrocarbons (PAHs) was assessed. Data indicate that discrimination of high and low potency carcinogenic PAHs by gene expression profiling is feasible [19]. Darroudi also presented data indicating that the anti-oxidants ascorbic acid and β-carotene do give genotoxic responses in HepG2 cells at high concentrations, much as in established rodent cell lines, and therefore the ability of anti-oxidants to become pro-oxidant at high concentrations may be just as likely in HepG2 cells.

The HepG2 cell system proved to be a useful *in vitro* model for detecting environmental and human dietary genotoxicants, anti- and co-genotoxicants [20,21]. Furthermore, this *in vitro* cell system has the potential to discriminate between structurally related carcinogens and non-carcinogens (Table 4) as well as between genotoxic and non-genotoxic carcinogens using gene expression profiling [22]. Consequently, it appears that

Table 4
A comparative study between genotoxicity data *in vitro* using human HepG2 cell system and (non)carcinogenicity data *in vivo*

Chemicals	Carcinogen (*in vivo*)	HepG2 (*in vitro*)	CHO cells (*in vitro*)		Ames test	
			With S9-fraction derived from		HepG2	Rat liver
			HepG2	Rat liver		
2-AAF	+	+	+	−		
4-AAF	−	−	−	−		
B(*a*)P	+	+	+	+		
Pyrene	−	−	−	−		
CP	+	+	+	+		
DMN	+	+	+	+		
HMPA	+	+	+	−	+	−
Safrole	+	+	+	−	+	−

Abbreviations: CHO, Chinese hamster ovary; 2-AAF, 2-acetylaminofluorene (*in vivo*, carcinogen); structurally related chemical 4-AAF, 4-acetylaminofluorene (*in vivo*, non-carcinogen); B(*a*)P, benzo(*a*)pyrene (*in vivo*, carcinogen) structurally related chemical pyrene is non-carcinogen (*in vivo*); CP, cyclophosphamide; DMN, dimethylnitrosamine; HMPA, hexamethylphosphoramide.

Table 5
Induction of micronuclei in AHH-1 and MCL-5 cells by trichloroethylene (University of Swansea data)

Concentration of trichloroethylene (μM)	AHH-1 cells		MCL-5 cells	
	% binucleated (BN) cells	Micronucleated cells/1000 BN	% binucleated (BN) cells	Micronucleated cells/1000 BN
0	55.9	0.90	56.3	1.00
3.8	50.0	1.05	45.1	3.60
9.5	38.7	2.20	35.3	5.10
19.0	37.2	1.60	21.9	5.10
38.0	33.2	1.65	9.3	13.79

the possibility of getting false positive results is low in HepG2 cells, but more experiments on this line of work are required. Furthermore, robust and important endpoints, such as gene mutations, are difficult to study using current protocols in HepG2 and other highly differentiated cells. Method development in this direction will be important if these cells are to be used more widely.

Parry from the University of Swansea presented genotoxicity (mainly *in vitro* micronucleus) data on the MCL-5 cell line, and the human lymphoblastoid AHH-1 cell line from which it was derived. The AHH-1 cell has a high level of expression of CYP1A1 [23]. The MCL-5 cells contain cDNAs for four human CYPs plus microsomal epoxide hydrolase [24]. Both cell lines are heterozygous for *p53*, but undergo a normal repair response and apoptosis following DNA damage. Both cell lines have a modal number of 46 chromosomes although there is some variability and they are probably not euploid. The metabolic competence of these cells avoids the general need for exogenous S9 and, for example, allows for prolonged treatments with pro-mutagens, which could not occur with S9 because of the toxicity induced. The additional CYPs in the MCL-5 cells is seen as an advantage over the AHH-1 for a number of reasons. As can be seen in Table 5, trichloroethylene is more clearly detected as a genotoxin in MCL-5 than in AHH-1 cells. However, as can be seen in Table 6, chloral hydrate is much less active in MCL-5 than in AHH-1 cells, and would have produced negative results at 50–60% reduction in binucleated cells. The underlying mechanism is unknown and probably not related to CYP expression. It has been the conclusion of the Swansea group that the use of both AHH-1 and MCL-5 provides a more comprehensive screen when testing chemicals of unknown genotoxicity. The cells have been particularly useful at discriminating genotoxic from non-genotoxic chlorinated compounds and studying non-disjunction by synthetic and natural oestrogens. However, very few chemicals have been tested that are neither *in vivo* genotoxins nor DNA-reactive, mutagenic carcinogens. Thus, the susceptibility of these cell lines to false positive results is not currently known. Furthermore, the cells are proprietary, and their routine use for regulatory testing would require appropriate supply and costing arrangements with the supplier.

2.6. New cell systems

White from Health Canada reported on the use of cells derived from the *LacZ* transgenic mouse, Muta™Mouse [25]. FE1 cells isolated from Muta™Mouse lung have a modal chromosome number of 78 (i.e. subtetraploid), but the range is from 62 to 82 chromosomes per cell. The cells appear stable, and have retained important cytogenetic, genetic, biochemical and structural features for 50 generations. The cells can be used to measure gene

Table 6
Induction of micronuclei in AHH-1 and MCL-5 cells by chloral hydrate (University of Swansea data)

Concentration of chloral hydrate (μM)	AHH-1 cells		MCL-5 cells	
	% binucleated (BN) cells	Micronucleated cells/1000 BN	% binucleated (BN) cells	Micronucleated cells/1000 BN
0	58.9	0.85	56.8	1.05
600	48.2	2.25	51.0	0.95
1510	29.7	2.75	40.0	1.25
3020	9.7	6.25	17.9	1.75
6040	2.7	12.39	15.11	2.38

mutations in the *LacZ* transgene (e.g. via the positive selection system described by Gossen and Vijg [26]) or to measure induction of micronuclei in binucleate cells using the cytokinesis block method. They have relatively high levels of inducible CYP1A1 and, with some substrates, can achieve higher glutathione-*S*-transferase activity than HepG2 cells, several cell lines derived from the Big Blue rat, and several mouse fibroblast lines, and have levels comparable to H4IIEC3 rat hepatoma cells. Good mutant responses have been obtained with several reference mutagens (ethylnitrosourea, ICR-191, benzo[*a*]pyrene), but heterocyclic amines such as PhIP (2-nitro-1-methyl-6-phenylimidazo[4,5-*b*]pyridine) are only mutagenic when a low concentration of Aroclor-1254 induced rat liver S9 is added to the medium, presumably indicating that the FE1 cells do not express the appropriate Phase 2 enzymes and sufficient levels of CYP1A2 to compensate for this deficiency. Several hepatic cell lines have recently been isolated and some have been cultured for 6 months. Whilst the modal chromosome numbers have been closer to diploid, the spontaneous mutant frequency has been very variable from <10 to >140 $\times 10^{-6}$. The metabolic activity of the Muta^{TM}Mouse cell lines appears to depend on attachment to a solid surface. Cells forced to grow in suspension appear to have lost the capacity for metabolic activation. In an effort to miniaturise the assay system, FE1 lung cells have been adapted to grow in small volumes on microcarrier beads coated with Porcine gelatin (Fig. 11). The mutagenic response to benzo[*a*]pyrene for cells attached to microcarrier beads was very similar to that seen in monolayer cultures (Fig. 12). This bead suspension approach has the distinct advantage of being able to screen novel test articles that are only available in minute amounts. As with other systems, very few chemicals have been tested that are neither *in vivo* genotoxins nor DNA-reactive, mutagenic carcinogens. Therefore, the potential for false positive results is not known. Nevertheless, a recent investigation of PAHs did reveal that some *Salmonella* positives that are classified as "inadequate evidence of carcinogenic activity" by IARC (e.g. benzo[*ghi*]perylene) are negative for induction of *LacZ* mutations in the FE1 Muta^{TM}Mouse *in vitro* system.

Hastwell from GlaxoSmithKline reported on the evaluation of a *GADD45a-GFP* reporter assay developed by Gentronix in the UK. In the *GADD45a-GFP* assay, TK6 cells have been transfected with a novel green fluorescent protein reporter based on the human *GADD45A* gene. The tests are performed in microwell plates, which

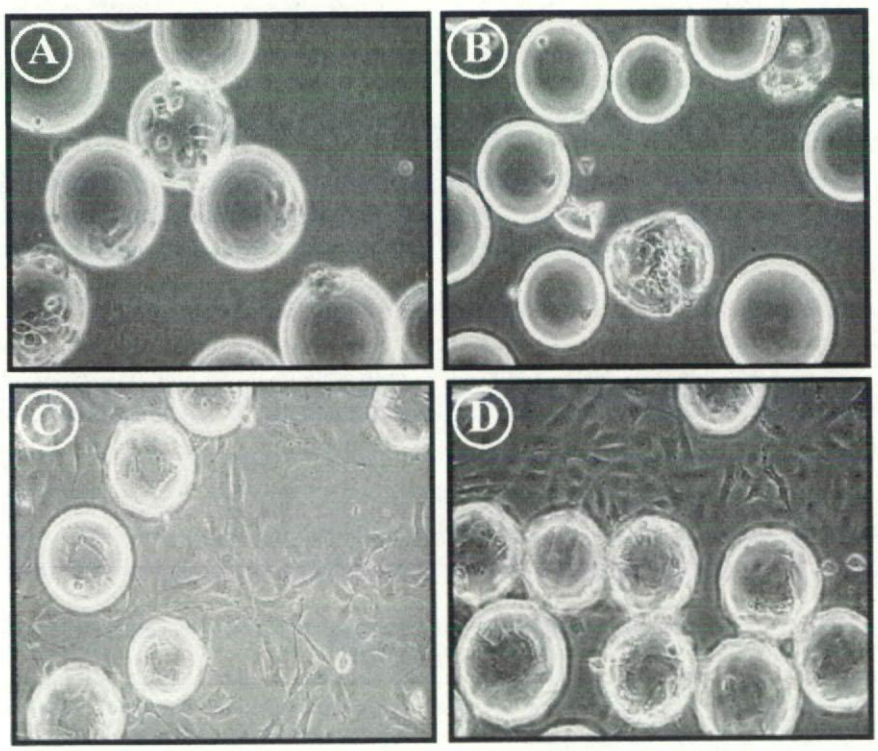


Fig. 11. Muta^{TM}Mouse FE1 cells growing on Cytodex® microcarrier beads (average diameter 175–190 μm, dextran coated with Porcine skin gelatin). Upper panels (A and B) show beads and cells after a 2.5 h initial attachment period followed by a 6 h exposure to benzo[*a*]pyrene. The lower panels (C and D) show the culture after 72 h expansion and growth on 100 mm polystyrene culture plates.

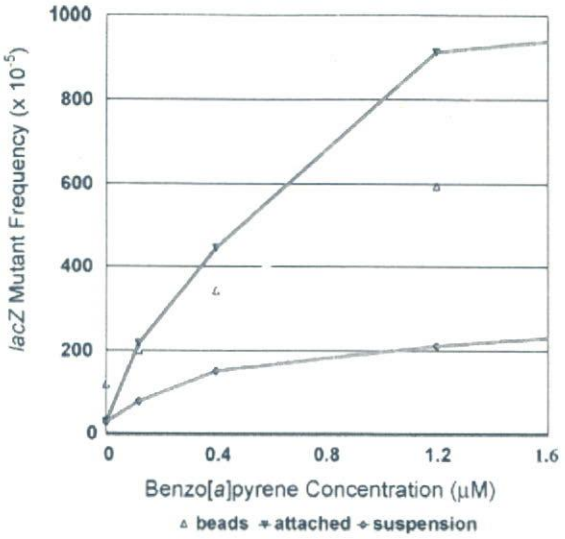


Fig. 12. *LacZ* mutant frequency ($\times 10^{-5}$) induced by benzo[*a*]pyrene (6 h exposure in serum-free medium followed by 72 h fixation period) in Muta™Mouse FE1 cells as an attached monolayer, in suspension culture, and attached to Cytodex® microcarrier beads.

means that only small quantities of chemical are needed. Generally <2 mg of a non-toxic chemical, tested up to 10 mM is required. The endpoint of the test is fluorescence, and based on historical data a positive result is concluded when the fluorescence increases 1.5-fold above control levels (Figs. 13 and 14). For the initial evaluation [27], the following sets of chemicals were chosen, and all tested at least four times:

- Thirty-four agents that are genotoxic in the absence of rat liver S9, with known mechanisms of action, that were positive in at least one test from the ICH battery [28], namely
  - 10 direct acting genotoxins,
  - 10 aneugens,
  - 7 nucleotide synthesis inhibitors,
  - 4 topoisomerase inhibitors,
  - 3 reactive oxygen species generators.
- Eleven cytotoxic positives, i.e. positive *in vitro* chromosome aberration data associated with cytotoxicity [29].
- Twenty-nine non-genotoxic agents with no positive *in vitro* genotoxicity data.

The 29 non-genotoxic agents and the 11 cytotoxic clastogens all gave negative results in the *GADD45a-GFP* assay, thus giving no false positives. Of the 34 expected genotoxins, 31 gave robust positive responses. Didanosine (a nucleoside analogue), thiabendazole (an aneugen) and methyl viologen (a reactive oxygen inducer) were negative. The evaluation has been extended by looking at marketed pharmaceuticals for which good *in vitro* and *in vivo* data exist (Hastwell et al., manuscript in preparation). From a total of 74 compounds (where rodent carcinogenicity data are available) the sensitivity of the *GADD45a-GFP* assay in relation to rodent carcinogenicity was 81% and the specificity was 94%. This is a much lower false positive rate than experienced with the conventional rodent cell lines used for chromosomal aberration and mutation experiments. Although a large number of chemicals has been evaluated, including a significant proportion of chemicals expected to be negative, none of these

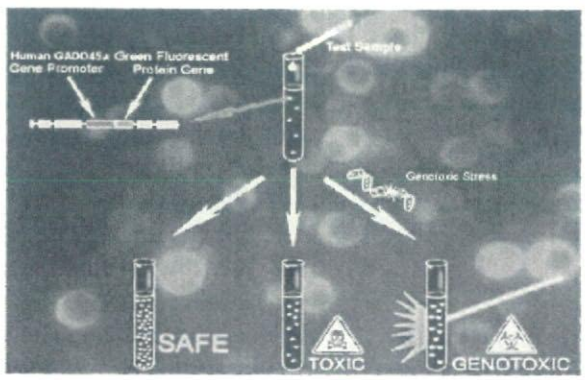


Fig. 13. Schematic of the 96-well plate method for determining genotoxicity in the *GADD45A-GFP* reporter system in TK6 cells (GreenScreen HC).

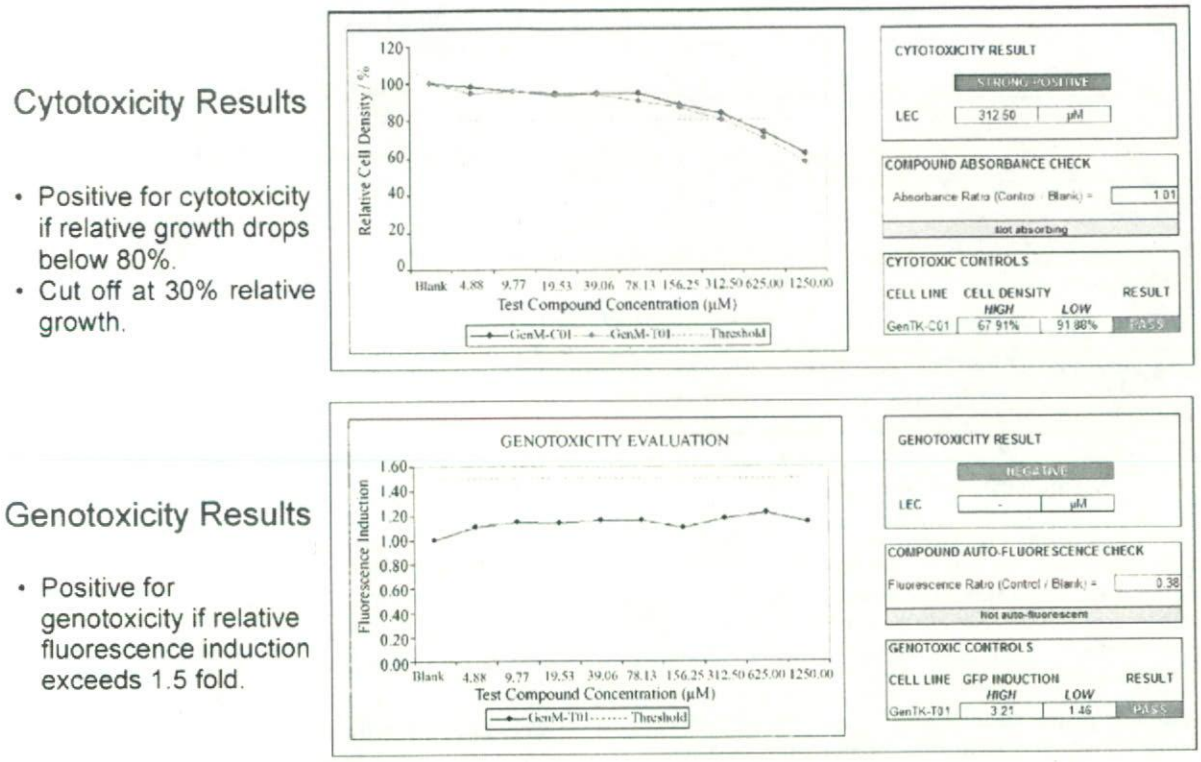


Fig. 14. Criteria for cytotoxicity, genotoxic response and graphical output from the GreenScreen HC assay system.

compounds required metabolic activation. S9 is fluorescent and absorbing and therefore could interfere with the current assay. Cyclophosphamide has been tested in a modified microwell plate. After a 24-h treatment in the presence of S9, significant fluorescence was observed in treated cultures. In principle, therefore, the *GADD45a-GFP* assay in TK6 cells may be adaptable to testing in the presence of S9. A modification of the assay using flow cytometry at 24 h after the start of treatment (4 h treatment, 20 h recovery period) has successfully detected responses with cyclophosphamide, 7,12-dimethylbenz[*a*]anthracene and benzo[*a*]pyrene in the presence of S9, but a better option may be to use HepG2 cells with the same target (*GADD45A*) and reporter (green fluorescent protein) genes. This system will also be proprietary when finally developed.

2.7. 3D skin models

False positive results in mammalian cell genotoxicity tests cause particular problems for the cosmetics industry because, according to the 7th Amendment to the EU Cosmetics directive, from 2009 onwards in the EU it will not be possible to follow up these findings with *in vivo* genotoxicity tests. As an alternative, 3D models of human skin are being evaluated for the possibility of measuring genotoxic endpoints.

Aardema from Procter and Gamble presented data on development of a MN assay in the Epiderm™ 3D human skin model [30]. Normal human-derived epidermal keratinocytes are grown on a membrane placed at an air–liquid (medium) interface inside a 9 mm culture insert. The cells differentiate and within 3 weeks tissue closely resembling human epidermis develops (Fig. 15). Methods were developed to reproducibly isolate individual cells from Epiderm™ cultures so as to prepare high quality slides for MN analysis. Gentle trypsinisation produces around 300,000 cells per tissue, and by using cytochalasin B to collect binucleate cells, around 40–50% of collected cells were found to be dividing. The background MN frequency has been found to be low (mean 0.05%, range 0–2/1000 cells) and clear induction of MN has been seen after treatment with mitomycin C, vinblastine sulfate, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine and methyl methanesulfonate (e.g. Fig. 16). More importantly, the rodent skin non-carcinogens trichloroethylene, 2-ethyl-1,3-hexanediol, 4-nitrophenol and 1,2-epoxydedocane were negative. Some metabolic characterisation has been performed, and Epiderm™ expresses numerous xenobiotic metabolism related genes observed in normal human skin. Further studies are in progress including experiments with chemicals requiring metabolic activation.

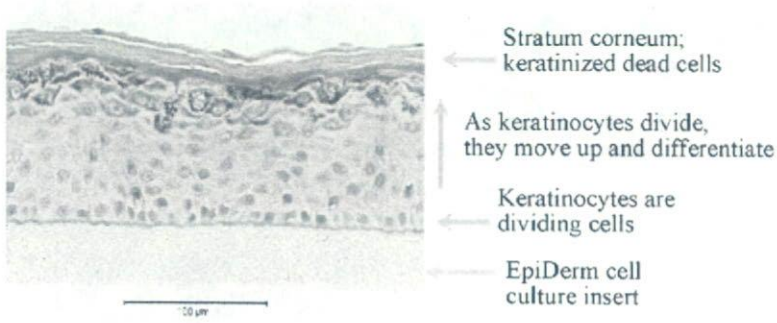


Fig. 15. Histology of Epiderm™ after *in vitro* culture of normal human keratinocytes.

Meunier from L'Oreal presented data from a similar 3D skin model, Episkin™. DNA damage was assessed by induction of comets in isolated keratinocytes. The induction of DNA damage by UV-A light was clearly enhanced by the photogenotoxic fluoroquinolone lomefloxacin (Fig. 17). 4-Nitroquinoline-*N*-oxide also induced comets in cells of Episkin™. However, no data were presented for chemicals that are accepted as neither genotoxic *in vivo* nor DNA-reactive, mutagenic carcinogens.

3. Discussion and recommendations

Discussion of the issues and presentations took place in four break-out groups, and then in a final plenary session. There was general agreement that the false positive rate with the current mammalian cell systems, in particular with the rodent cell lines, is not acceptable. There are several actions that need to be taken in order to improve the situation, and the consensus recommendations are summarised in the following paragraphs. Some of the actions are for the shorter term, i.e. to reduce the risk of false positives with the existing systems. Other actions are medium or long term requiring the development and evaluation of modified or new systems.

3.1. Interpretation of positive results as relevant or irrelevant

Until mammalian cell tests with higher specificity are identified or developed, it is necessary to obtain evidence on the relevance for humans of positive results in genotoxicity tests, in particular in cultured mammalian cells. It was acknowledged that the level of understanding of the mechanisms that might lead to false positive results is poor amongst many scientists involved in the safety of chemicals and drugs, and that education is needed. It is understood that several reviews of some of the accepted threshold and non-relevant mechanisms of genotoxicity, and approaches to obtain evidence, are in preparation. However, a detailed "trouble-shooting" manual on approaches to investigate whether a positive result is relevant could be very helpful. In addition to avoiding cells with unstable karyotypes and excessive cytotoxicity (see below) some of the approaches that can be taken are:

- Look for (lack of) suspicious activity in microarrays and quantitative structure activity databases.
- Determine whether reactive oxygen species were generated by reaction of the test chemical with culture medium.

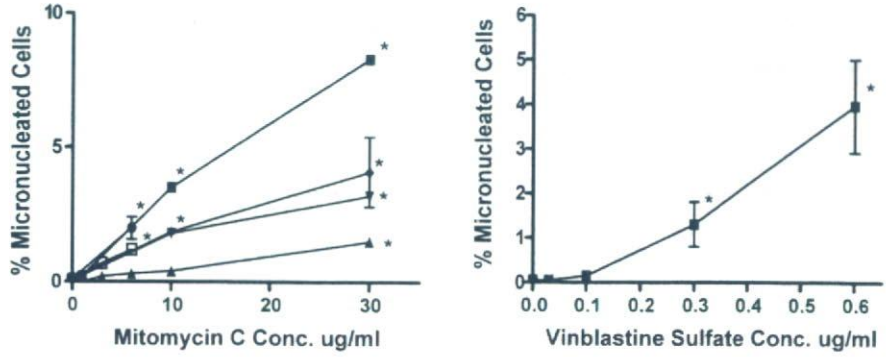


Fig. 16. Induction of micronuclei in Epiderm™ by mitomycin C and vinblastine sulfate.