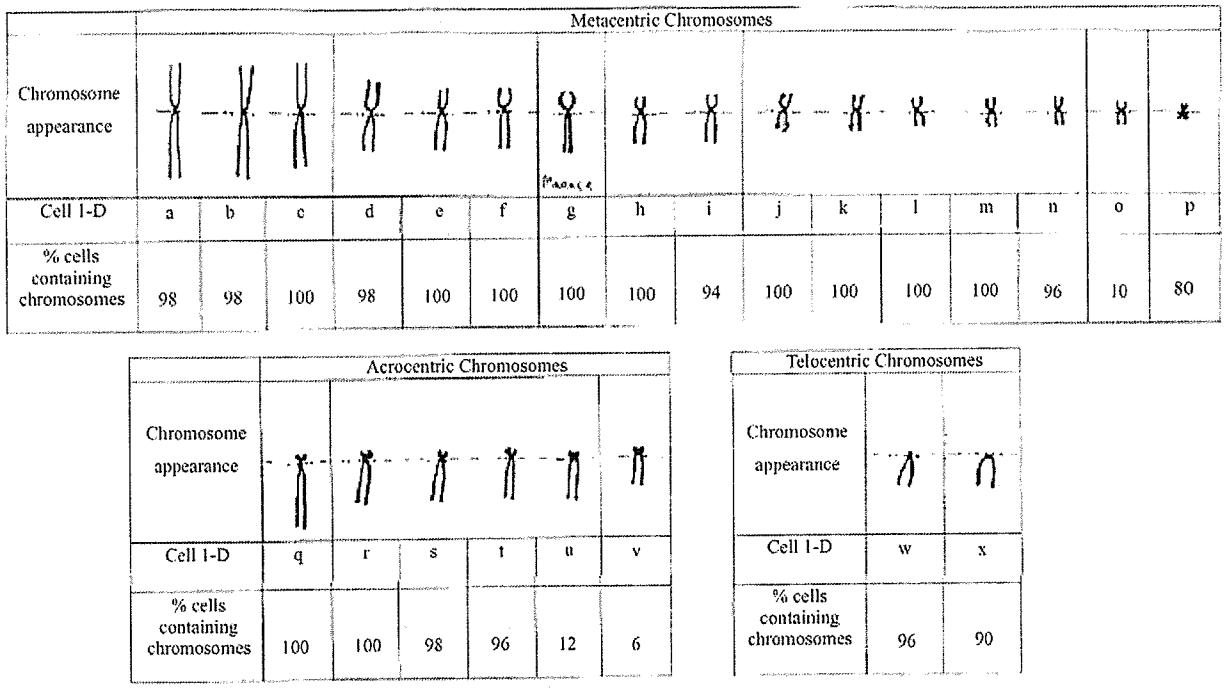


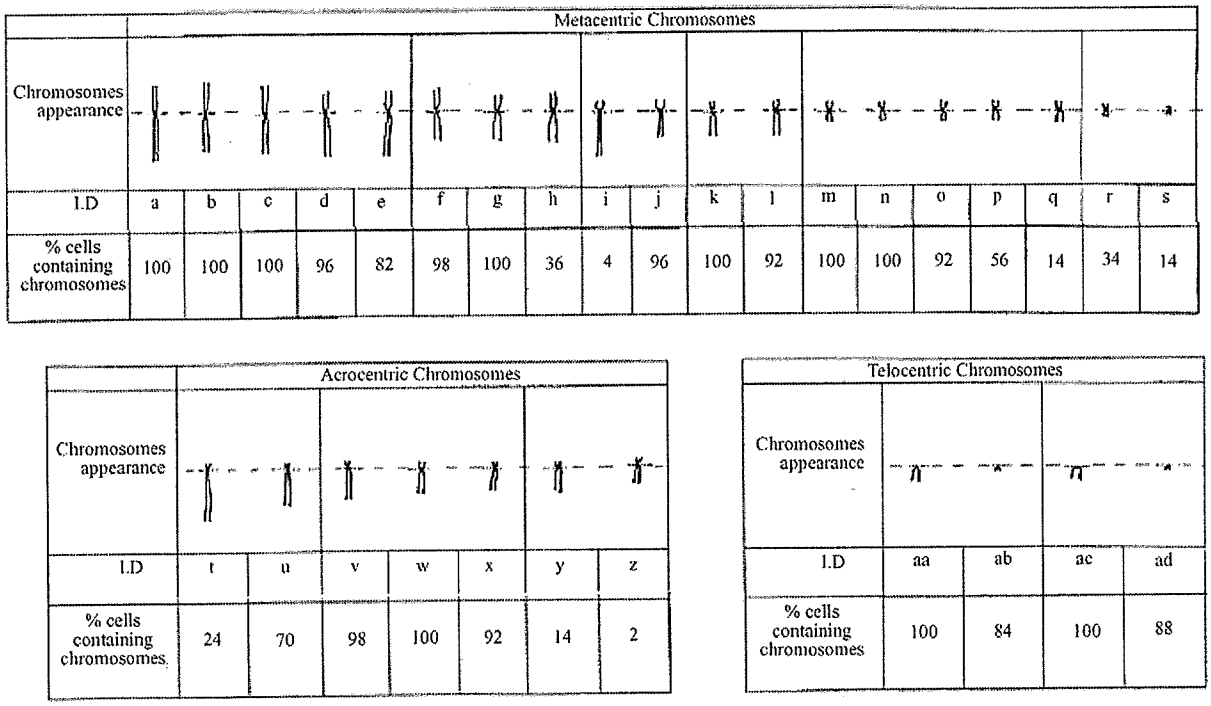
Metacentric Chromosomes																
Chromosome appearance																
Cell I-D	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
% cells containing chromosomes	98	98	100	98	100	100	100	100	94	100	100	100	100	96	10	80

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).

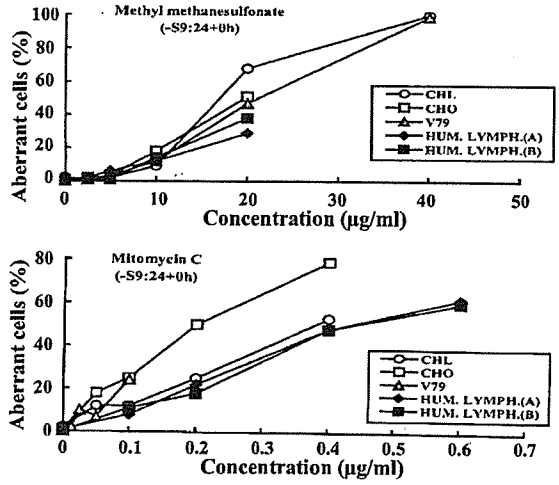


Fig. 3. Chromosomal aberration data for two chemicals tested in three different cell lines and peripheral blood lymphocyte cultures from two different donors at the same time.

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-

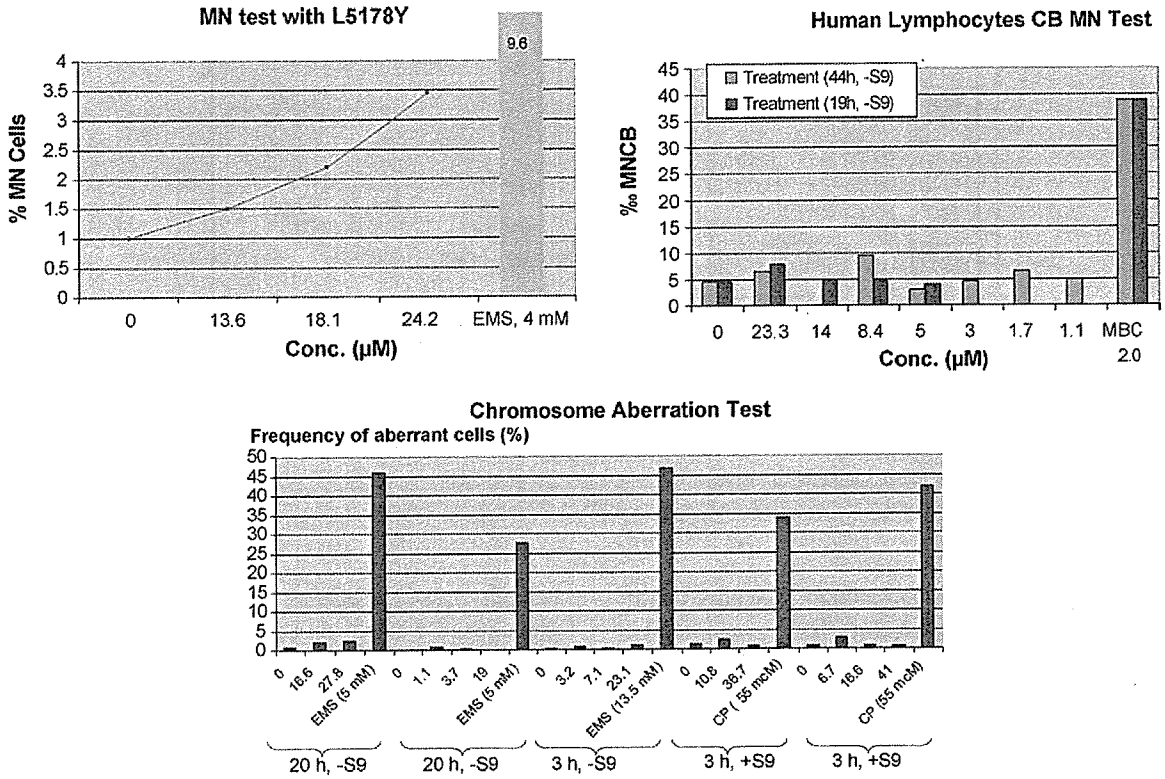


Fig. 4. Example from Novartis of a compound giving a positive micronucleus response when screened in L5178Y cells, but failing to induce either micronuclei or chromosomal aberrations in human lymphocytes at similar or higher concentrations. EMS, ethyl methanesulphonate; CP, cyclophosphamide; MBC, carbendazim.

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

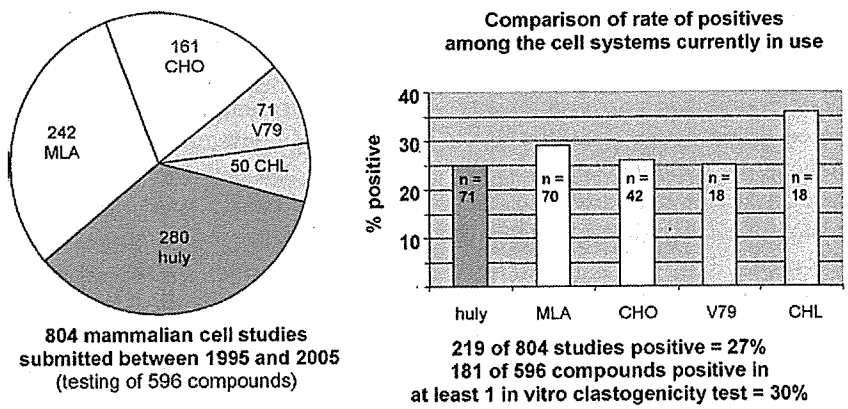


Fig. 5. Use of mammalian cell assays for regulatory submissions to the German Federal Institute for Drugs and Medical Devices (BfArM) between 1995 and 2005, and the frequency of positive results in human lymphocytes (huly), the mouse lymphoma assay (MLA) and the Chinese hamster cell lines V79, CHO and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

+ / -	MLA	Huly	V79	CHO	CHL
	11	7	6	9	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.

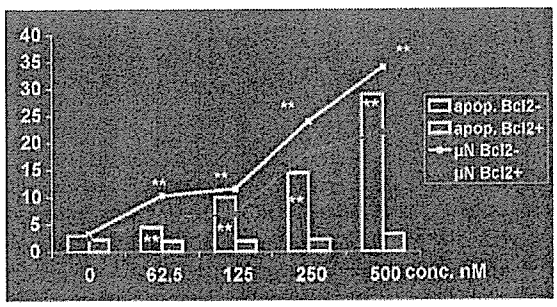


Fig. 7. Induction of micronuclei (μN) in the absence of apoptosis by etoposide in CTLL-2 cells with and without the human *bcl2* gene.

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

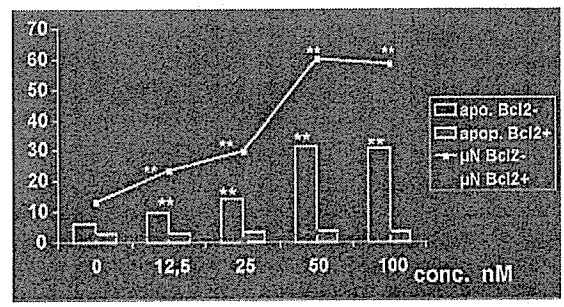


Fig. 8. Induction of micronuclei (μN) only in the presence of apoptosis by dexamethasone in CTLL-2 cells not expressing the human *bcl2* gene.

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.

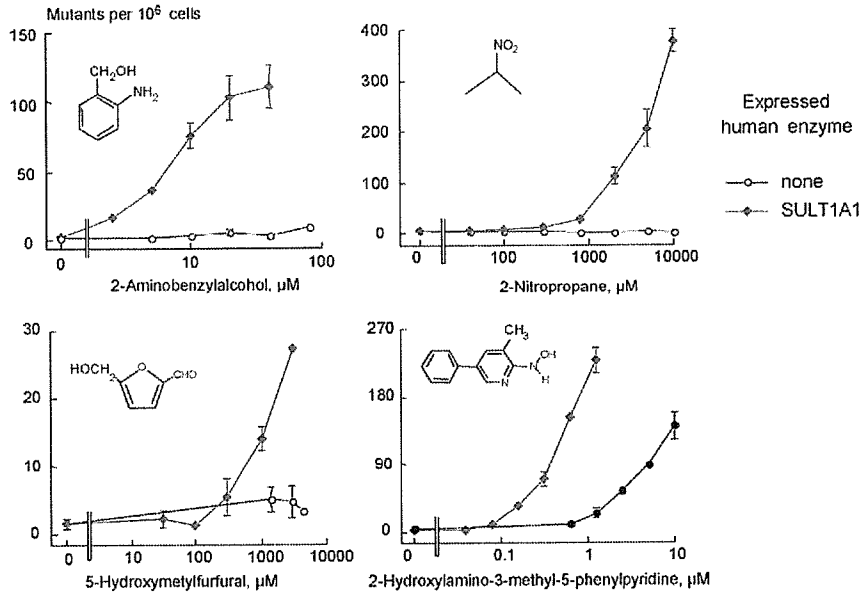


Fig. 9. Promutagens activated by human sulphotransferase SULT1A1 in recombinant V79 cells.

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

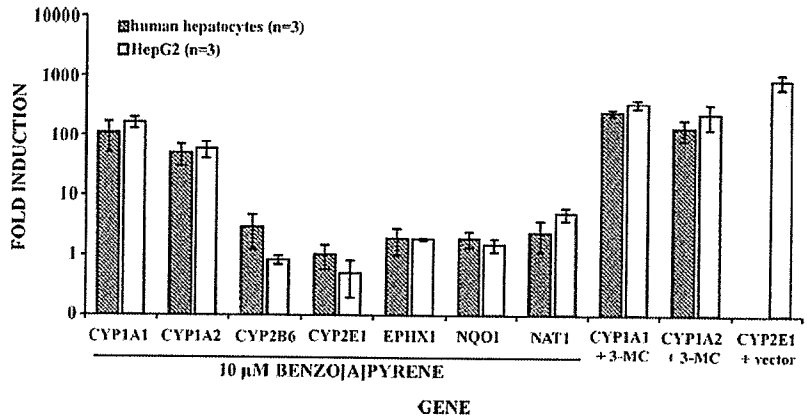


Fig. 10. Quantification of gene expression profiles in HepG2 cells and/or human hepatocytes treated with benzo[*a*]pyrene.

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

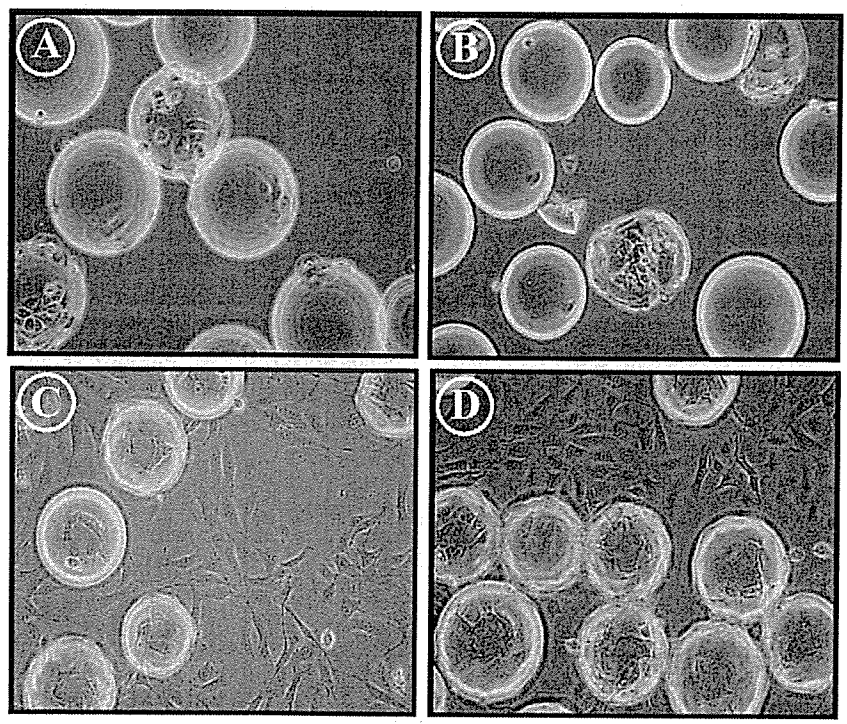


Fig. 11. Muta^{TM}Mouse FE1 cells growing on Cytodex® microcarrier beads (average diameter 175–190 μm, dextran coated with Porcine skin gelatin). Upper panels (A and B) show beads and cells after a 2.5 h initial attachment period followed by a 6 h exposure to benzo[*a*]pyrene. The lower panels (C and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

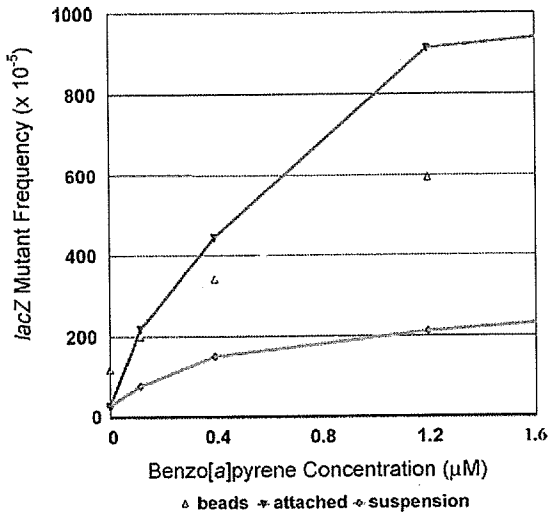


Fig. 12. *LacZ* mutant frequency ($\times 10^{-5}$) induced by benzo[*a*]pyrene (6 h exposure in serum-free medium followed by 72 h fixation period) in Muta™Mouse FE1 cells as an attached monolayer, in suspension culture, and attached to Cytodex® microcarrier beads.

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

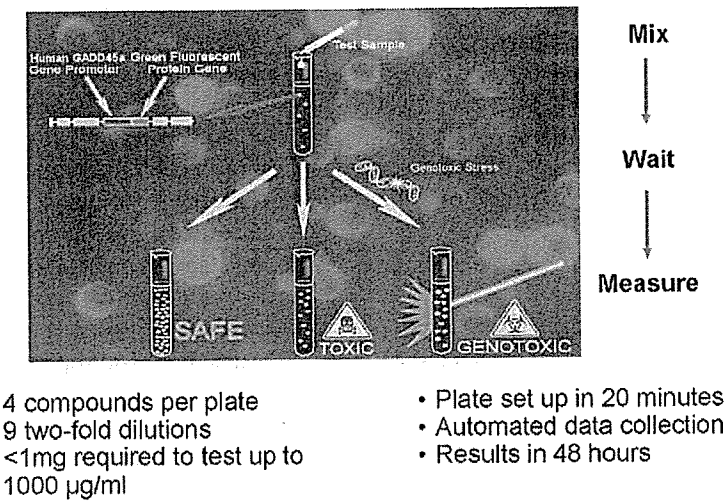


Fig. 13. Schematic of the 96-well plate method for determining genotoxicity in the *GADD45A-GFP* reporter system in TK6 cells (GreenScreen HC).

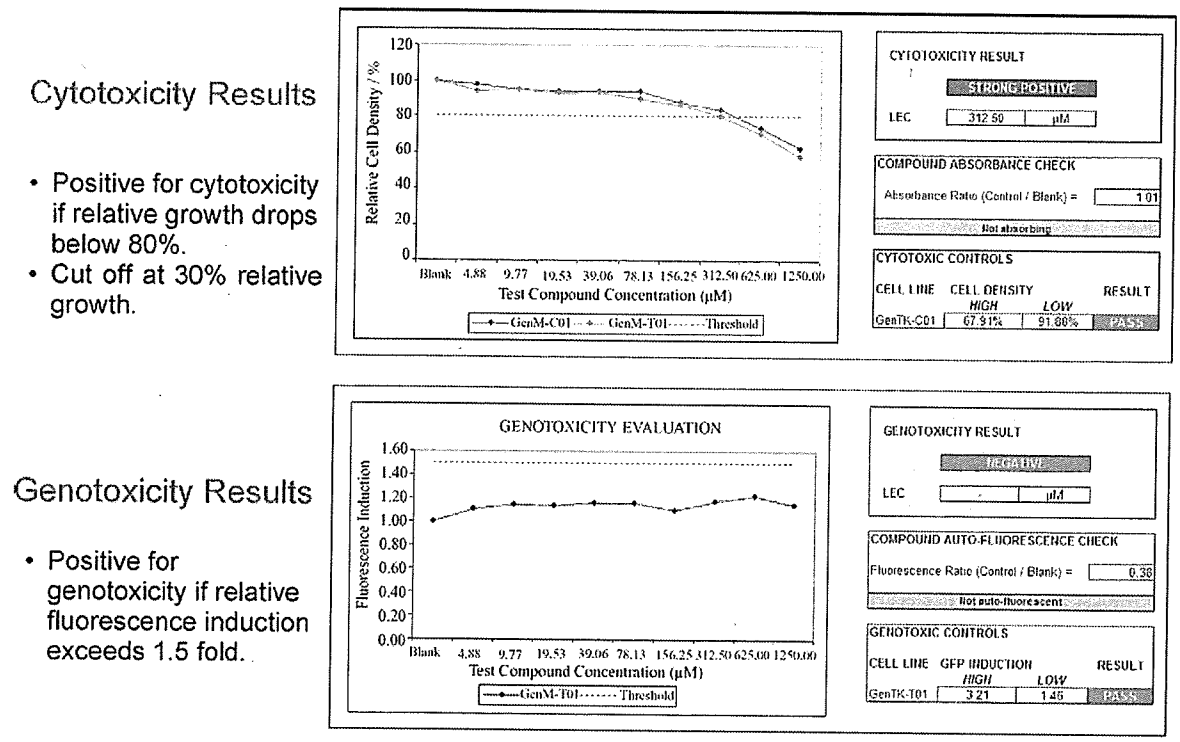


Fig. 14. Criteria for cytotoxicity, genotoxic response and graphical output from the GreenScreen HC assay system.

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.

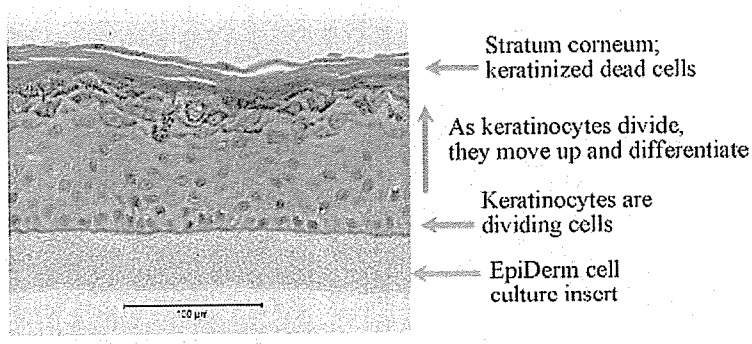


Fig. 15. Histology of Epiderm™ after *in vitro* culture of normal human keratinocytes.

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.

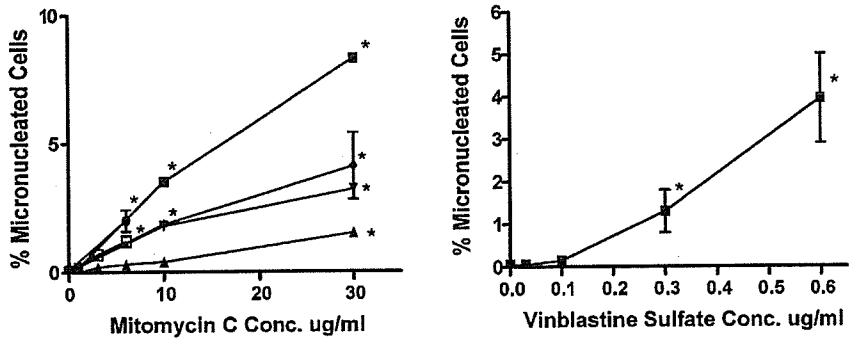


Fig. 16. Induction of micronuclei in Epiderm™ by mitomycin C and vinblastine sulfate.

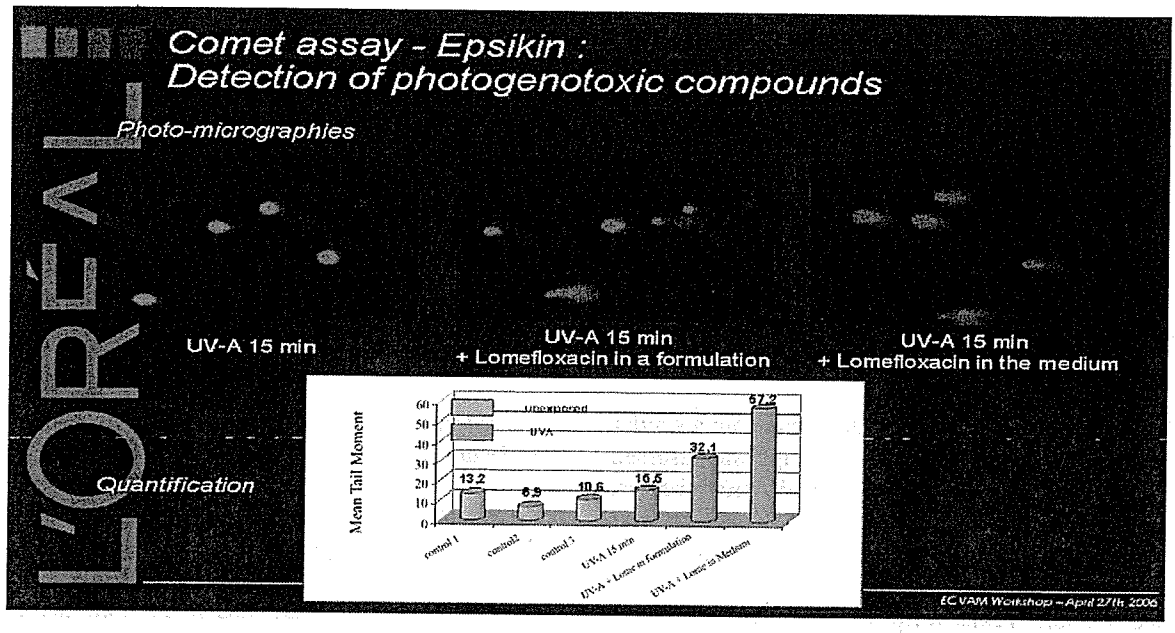


Fig. 17. Induction of comets by lomefloxacin + UV in Episkin. For the photosensitization experiments with lomefloxacin, the Episkin® epidermises were treated either by adding a 50 μm lomefloxacin solution in ethanol to the Episkin® culture medium "underneath the skin" (mimicking the systemic flow) or by topical application of 1 mg/cm^2 of a 4% lomefloxacin cream (kindly provided by L'Oreal Applied Research Laboratories at Chevilly LaRue, France). In those cases the Episkin® was treated for 1 h with lomefloxacin, rinsed three times with PBS and then irradiated 15 min with UVA light. The comet assay was performed immediately after the irradiation. All the unexposed controls (topically treated or culture-media treated epidermises, untreated epidermis) were included in the experiment as described in the figure.

- Check for degradation products in culture medium.
- Investigate possible confounding effects of apoptosis (e.g. by using CTLL-2 *bcl2* cells or Annexin-V analysis) and necrosis.
- Check for agonist and antagonist effects on kinases.
- Check for loss of cellular homeostasis (e.g. high osmolality, low pH).
- Check for metabolic poisoning and inhibition of DNA synthesis.
- Check for possible exposure to UV light.
- Check for nucleotide pool imbalances.
- Check for metabolic overload (e.g. glutathione depletion).
- Determine absence of DNA adducts under genotoxic conditions, preferably using radiolabeled chemical rather than ^{32}P-postlabeling.

In order that practising scientists and regulatory reviewers may become more familiar with the likely causes of genotoxic responses that are not relevant for humans, we encourage journals to publish data on coded compounds (from industrial in-house databases) that will help exemplify non-DNA or non-relevant mechanisms of genotoxicity. However, in such cases, as much information as possible should be provided on each test agent, and an independent review of the classification of positive and negative calls will probably be needed.

In addition to the above suggestions, it was acknowledged that 3D tissue models, such as those presented for skin, could provide valuable information on the relevance of *in vitro* positive results, and the further development of such models is to be encouraged. However, given that relatively few human or animal skin carcinogens (UV light, PAHs) exist, these models may be of limited value as replacements for whole animals.

3.2. Cell culture conditions and techniques

It was agreed that "good housekeeping" of cell cultures is a requirement for reliable and reproducible results. Practitioners should avoid working with high passage cells and should look at chromosomal content, karyotype and other characteristics such as metabolic capability and response to reference genotoxins for evidence of genetic drift. There are some suggestions that cell density and culture size can have an impact on the response of the cultures to chemical insult and this needs to be investigated. The ECVAM task force on good cell culture practice (GCCP) has previously published recommendations on GCCP [8]. These recommendations need to be reviewed in light of the current workshop to

see if they need to be up-dated and/or adapted to *in vitro* genotoxicity testing.

There was concern at the possibility of reactive oxygen species being formed by reaction between the culture medium and test chemical. More data are needed before advice can be given on whether certain cell/media systems are less likely to produce artefactual results through oxidative stress than others, and Halliwell is encouraged to continue his investigations in the hope that such recommendations can be made.

3.3. Biotransformation

The xenobiotic-metabolising system comprises several hundred enzymes, which are usually expressed with high selectivity in varying tissues, cell types and ontogenetic stages, and in rodents also often with high sex specificity. Some enzymes are involved in the biotransformation of many genotoxicants, others are important only for a small number of compounds, and some reaction types involve a higher risk of formation of reactive metabolites than others. Various enzymes are only present at significant levels after induction by specific endogenous or xenobiotic factors. Thus, no cell type *in vivo* reflects the full biotransformation capacity of the organism. Even hepatocytes, which are heavily involved in biotransformation, only express a limited selection of xenobiotic-metabolising enzymes. For example, there are now strong indications that the hepatocarcinogenicity of PAHs in rodents is due to bioactivation by CYP1B1 in extrahepatic tissues, and that hepatic CYP1A1 – which normally is used for "promiscuous" activation of PAHs *in vitro* – acts as a major PAH-detoxifying enzyme *in vivo*. Moreover, the expression of numerous enzymes ceases, or is drastically decreased, in cells in culture. In part, this simply reflects the propensity of the organism to avoid expression of risk-borne enzymes in proliferating cells. This is true even for hepatic cell lines (e.g. HepG2) that have retained much biotransformation activity in comparison to fibroblastoid or lymphocytic cell lines. Classical S9 is a rich source of selected CYPs, but otherwise the spectrum of enzymes present in active form is very low.

The risks resulting from the formation of reactive intermediates is reduced *in vivo* by the presence of detoxifying systems, which are often extremely efficient (but overlap with toxifying systems). Two processes of detoxification can be distinguished:

(i) Metabolic "sequestration" of a promutagen into pathways that avoid the formation of the ultimate mutagen.

(ii) Inactivation of an active metabolite after its formation.

These processes differ in the enzyme systems involved. Sequestration commonly occurs by CYPs, reductases/dehydrogenases, UGTs and SULTs. The individual members of these enzyme classes are often expressed with high selectivity in certain cell types of physiological stages. Inactivation of active (electrophilic) metabolites is normally conducted by epoxide hydrolases or GSTs. These enzymes – although not every single form – are widely expressed in many tissues and cells. Thus, microsomal epoxide hydrolase and substantial levels of GST activity towards some substrates have been detected in all mammalian cell lines studied [31]. However, high-efficiency (high V_{max}/K_m) enzymes may be most important for sequestration and inactivation, especially in systems with high-efficiency, rather than promiscuous, activation. Some proximate and ultimate genotoxicants equilibrate *in vivo*. In this case, efficient protection may even occur through enzymes located at sites different from the site of activation. However, when this equilibration is limited, the appropriate localisation of the detoxifying system may be critical. An example is aflatoxin B_1, which is a potent hepatocarcinogen in the rat but only weakly active in the mouse. Constitutive expression of Gst a5, an enzyme that efficiently inactivates aflatoxin B_1 8,9-oxide in mouse but not rat liver, appears to be an important mechanism underlying this difference. Although a rat Gst a5 is constitutively expressed in various extrahepatic tissues, this localisation appears to be inefficient for toxification. However, after hepatic induction of Gst a5 by certain chemicals, the rat becomes resistant towards the hepatocarcinogenicity of aflatoxin B_1 [32]. It is not possible to mimic such complex, varying pharmacokinetic processes in a simple *in vitro* screening model.

Detoxification may occur to the parent compound or downstream to its metabolites. The capacity of *in vitro* systems to detoxify is usually modest, often limited to the nanomolar to low micromolar concentration range over the entire exposure period. Thus, no significant competition between toxifying and detoxifying activities would occur unless the maximum concentration of parent compound tested was rigorously restricted to a very low range, and the relevant enzymes were present. The situation is very different for metabolites. Their concentrations are low, implying high-affinity (or more precisely high-efficiency) detoxification reactions can potentially occur.

The purported lack of enzymic detoxification *in vitro* is in contrast to observations that various chemicals,

when tested at high concentrations, are positive in direct genotoxicity assays, but negative in the presence of S9. Although the underlying mechanism(s) for the lack of genotoxicity in the presence of S9 have not been elucidated in most cases, it has been demonstrated for some cases that heat-inactivated S9 was also protective (H.R. Glatt, unpublished results). This suggests that physical trapping (e.g. of lipophilic compounds in microsomes) or chemical trapping (e.g. reaction of electrophiles with nucleophilic sites), rather than enzymatic activities, produced the effect. Therefore, it is difficult to assess the *in vivo* significance of this observation.

Currently the impact of metabolic differences between *in vitro* and *in vivo* test systems on the false positive rate in *in vitro* genotoxicity tests is not known. However, it is clear that variation of the metabolising system can have dramatic effects on the results of *in vitro* tests as well as for animal studies, where genetic knockout or inhibition of an individual enzyme can eliminate the ability of a carcinogen to induce tumours. In an ideal world the same metabolic modulation should have parallel consequences *in vivo* and *in vitro*.

A review of the important *in vivo* genotoxins and DNA-reactive mutagenic carcinogens is needed to determine whether metabolic differences between *in vitro* and *in vivo* test systems are, in fact, contributing to the high false positive rate, and to better define the relevant metabolic systems to include in *in vitro* tests. Perhaps we will have to include a much larger variety or different set of enzyme systems than have been traditionally used in *in vitro* tests to predict better what happens in animals. Genetic engineering may be used to help address this goal, but this approach will be very time- and cost-intensive. In addition, to achieve the proper balance of all of the relevant enzymes in a given engineered cell line to appropriately detect all *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens is likely to be an insurmountable task. Moreover, such an approach would not reflect the *in vivo* situation, where different enzymes are often compartmentalised in different cells. Alternatively, a panel of individual cell lines, each with a small number of expressed enzymes, could be used. Almost infinite possibilities for permutations of various enzymes exist with such an approach, generating ideal tools for research on mechanisms. However, the selection of a cell battery for broad scale genotoxicity screening would require some arbitrary, pragmatic decisions, and this arbitrary character, as well as large deviations from the balance of enzymes *in vivo*, would remain obvious. This is in contrast to alternative metabolising systems (S9, conventional "metabolically competent" cell lines) where the situation may be similar but less obvious and less open for remedies by systematic, hypothesis-driven research in a concrete situation.

From the point of view of genetic engineering, one may either start with a cell line that has retained some residual xenobiotic-metabolising activities (such as HepG2), or from a relatively "clean" cell line (V79 or a human equivalent, if available). The former model has the advantage that less engineering is required to achieve broad xenobiotic-metabolising capacities. The latter model is favoured when used as an analytical tool to specify critical host factors, as background activities are minimised. Further aspects that should be taken into account in the selection of the basic system(s) are:

(a) Genetic engineering intrinsically involves numerous culture passages and clonal selections; therefore, long-term stability of critical properties (obviously including biotransformation activities) is pivotal.
(b) The cell line selected should be suitable for efficient analysis of important and robust endpoints, such as gene mutations.

3.4. Top concentration for testing

Current OECD guidelines for genotoxicity testing in mammalian cells require that the top concentration with soluble and non-toxic substances should be 10 mM or 5000 μg/ml, whichever is the lower. There was some discussion about whether this may be appropriate for complex mixtures and technical grade (impure) industrial chemicals for example, where the objective is not only to test the genotoxicity of the main ingredient. However, given that mutagenic impurities are only detected in an Ames test carried out to 5 mg/plate (when spiked at a level of about 5% [33]), it was questioned whether detection of impurities is a sufficiently important role for genotoxicity testing to justify pushing to such high concentrations. However, the K_m for many biochemical reactions, whether involved in metabolic activation/inactivation, general cellular defence/balance, cellular transport or cellular turnover is less than 100 μM [34,35,36]. It is probable that low K_m reactions primarily determine the bioactivation pathway *in vivo*. These kinetic characteristics suggest that the high concentrations currently required for *in vitro* testing may not be informative for human risk assessment. The 10 mM and 5000 μg/ml requirements are seemingly based on a small number of carcinogens that needed high concentrations before giving positive responses in mammalian cell tests *in vitro*, sometimes using inappropriate metabolic conditions. It is not known whether the carcinogens that require these high concentrations for detection *in vitro*

are "important", how robust the *in vitro* mammalian cell findings are, or whether these chemicals are positive in other test systems (e.g. the Ames test) currently used in a standard battery. The fact that the published data on these chemicals are probably quite old could mean that under current chromosomal aberration and gene mutation protocols they could be detected at lower concentrations. It also has to be considered that simple detection of a carcinogen at high *in vitro* concentrations that are not relevant *in vivo* does not mean there is a mechanistic correlation between the *in vitro* genotoxicity and the *in vivo* carcinogenicity.

Further to the above considerations on the K_m of important biochemical processes, general considerations on *in vivo* exposure to (toxic) compounds have been put forward. In this context one can consider knowledge about intentional high dose or long-term exposure to pharmaceuticals as worst case examples. It is clear that even high dose pharmaceuticals such as antibiotics (e.g. penicillins, fluoroquinolones) or pain relief agents such as acetaminophen (also known as paracetamol) seldomly yield systemic or tissue levels >10 μM [37]. Thus, taking chronic intake, possible accumulation and overdosing scenarios into account, a lowering of the current maximum *in vitro* concentration, perhaps by 10-fold or more, may be justified, at least for certain types of chemicals (such as pharmaceuticals) from scientific and consumer protection viewpoints.

The participants therefore agreed that a new review of existing data is needed to determine whether such high concentrations as 10 mM or 5000 μg/ml are needed to detect *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens. The following actions are recommended:

- An expert panel should be assembled to determine which *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens need to be detected in *in vitro* mammalian cell tests. This subset of chemicals should be determined from both published and industry (confidential) data. A first suggestion of an important data set would be the IARC groups 1, 2A and 2B carcinogens, omitting those, such as hormones and immunosuppressants, which are acknowledged to be of a non-genotoxic mode of action. In view of the on-going initiative of the Health and Environmental Sciences Institute of the International Life Sciences Institute (ILSI-HESI) with regard to false positives in *in vitro* mammalian cell genotoxicity tests, it was suggested this action could be best achieved in conjunction with ILSI-HESI, and the outcome of their workshop (held in June 2006) will be reported elsewhere.
- The role of metabolism in the activity of the above *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens needs to be reviewed in order to define the appropriate metabolic systems to include in *in vitro* genotoxicity assays.
- The published and industry data should be reviewed to determine whether concentrations as high as 10 mM or 5000 μg/ml are needed to detect this important subset of chemicals, or whether a lower level could be justified.
- If high concentrations are needed in the mammalian cell tests, data from other tests such as the Ames test should be reviewed to see if the chemical(s) would be detected in other parts of the standard battery.
- If high concentrations are needed, and genotoxicity was not detected in other parts of the standard battery, an opinion should be formed as to whether a more modern protocol or modified metabolic conditions would be likely to detect genotoxic effects at lower concentrations. If necessary, new testing should be initiated.
- If high concentrations are needed (with appropriate metabolic conditions) a scientific effort should be mounted to elucidate whether the mechanism(s) of genotoxicity for these chemicals would trigger responses in other toxic endpoints.
- A thorough evaluation of various human exposure scenarios should be made to determine an upper limit of *in vitro* testing from the viewpoint of human consumer protection.

It is evident that the concentration of a pro-genotoxicant required for a positive test result can vary dramatically depending on the activating system used. General improvements in the activation systems, or in a chemical class-dependent manner (with corresponding positive control compounds), might be an important pre-requisite for a reduction in the top concentration. For the time being, the participants of the workshop concluded that it is prudent to challenge the current recommended upper concentration of *in vitro* testing (10 mM or 5000 μg/ml), and that a lower level appears to have scientific merit.

3.5. Measures and extent of cytotoxicity

Many different measures of cytotoxicity are used in mammalian cell tests, in particular the chromosomal aberration test. Reductions in cell count, confluency, mitotic index and population doubling are all widely used and equally accepted, but it is unlikely that all these measures would select the same top concentra-

tion for testing. Other indicators of toxicity such as ATP levels, mitochondrial function, LDH-leakage may also be appropriate. Greenwood et al. [11] showed that several non-DNA-reactive chemicals and metabolic poisons would not have given positive chromosomal aberration results if the 50% cytotoxic concentration had been chosen based on a reduction in population doubling rather than a reduction in cell count, and that, using this measure, no important DNA-reactive genotoxins would have been missed. These findings have not been independently verified and there are no other publications comparing different measures of cytotoxicity in relation to genotoxicity.

The participants therefore agreed there is a need for a thorough comparison of different measures of cytotoxicity in case some measures may select concentrations for testing that allow the detection of all important *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens but lowers the risk of false positives. In particular:

- A collaborative trial is needed on a selected set of chemicals. This could be the same subset of chemicals selected for evaluation of top concentration (above). Additional sets of chemicals, such as the non-DNA-reactive chemicals of the Greenwood et al. [11] publication, and agreed non-genotoxins also need to be included. If possible, this trial should be co-ordinated with any initiatives coming from the ILSI-HESI workshop (June 2006) which will be reported elsewhere.
- Multiple endpoints of toxicity need to be compared at the same time in each participating laboratory.
- Dose–response relationships for cytotoxicity and genotoxicity should be determined, and include the current required levels (i.e. at least 50% toxicity for chromosomal aberrations, at least 60% toxicity for the micronucleus assay and at least 80% toxicity in the mouse lymphoma assay).
- Human cells such as lymphocytes should be included as well as the rodent (and any other, e.g. TK6, HepG2) cell lines, and measures of cytotoxicity for lymphocytes other than mitotic index need to be identified.
- Robust endpoints of genotoxicity that are not sensitive to interference by cytotoxicity (e.g. gene mutations and DNA adducts) should be identified and included in the trials.

Although the only indication came from the data of Elhajouji, there was a belief amongst many participants that a majority of false positive results in chromosomal aberration and mouse lymphoma tests probably occur in the prolonged, continuous treatments in the absence of exogenous metabolic activation. Several factors such as extent of exposure, prolonged cytotoxicity, and lack of detoxification by S9 may be involved, but the exact reasons are not known. However, it was noted that the 50% and 80% toxicity requirements for the chromosomal aberration and mouse lymphoma assays were originally based on short (e.g. 3–6 h) treatments, and the need for these levels of toxicity (or even the appropriateness of the current measures) has not been independently justified. Therefore, this collaborative trial must include chemicals that are only positive after prolonged (e.g. 20–24 h) treatments.

3.6. Criteria for and evaluation of new mammalian cell test systems

Certain characteristics of the commonly used rodent cell lines (CHO, CHL, V79, L5178Y, etc.) such as their *p53* status, karyotypic instability, DNA repair deficiencies, etc. are recognised as possibly contributing to the high rate of false positives. The need for exogenous metabolism with the cell systems is also expected to contribute to the false positive rate. If these cell types are to be replaced in the future, any new systems should ideally:

- Be early passage.
- Be karyotypically stable and, if possible, normal.
- Be *p53* proficient.
- Be DNA repair proficient.
- Preferably consist of human cells.
- Be metabolically competent (at least Phase 1 and Phase 2 capacity should be defined), if necessary through genetic engineering. Expert advice is needed on what are the essential Phase 1 and Phase 2 enzymes that should be functional in any new system for the biotransformation of *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens. Several different cell lines with different enzyme profiles may cover a reasonable fraction of the complex biotransformation machinery present *in vivo*. However, the selective expression of a limited number of enzymes in a cell line may direct the biotransformation into pathways that are different from the major pathways occurring in animals and humans, where many different enzymes compete and interact with each other.
- Be able to detect the majority of genetic endpoints relevant to human somatic and inherited disease.
- Show improved specificity without reducing the ability to detect *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens.

The participants were aware that this wish list is far too ambitious to be realised in a single test system, at least within a reasonable time. Therefore, several systems may be required, which in combination will facilitate the distinction between correct and false positive results.

Some of the data presented at the workshop indicated that the human lymphocyte cell system might produce a lower level of false positives than the common rodent cell lines. If the use of the clastogenicity endpoint *in vitro* is to be continued, further work is needed to establish whether human lymphocytes do offer a lower false positive rate, and also to find alternative methods (other than mitotic index) for measuring cytotoxicity.

Although many of the new systems presented at the Workshop show promise, and fulfil some of the criteria given above, none fulfils all of the criteria. In many cases (e.g. MCL-5, HepG2, transgenic cell lines) it is the lack of data on specificity that is the problem. In the case of the *GADD45a-GFP* assay it is mainly the lack of data with compounds requiring metabolic activation. The new 3D skin models are also at an early stage of development. It was therefore agreed that a collaborative research program is needed to evaluate new mammalian cell-based methods and models for genotoxicity. In addition to the cell systems discussed at the workshop, other cells that have retained some xeno-metabolic activities (e.g. HepaRG [38] and AR42J-B13 rat pancreatic stem cells [39,40]) or cell lines genetically engineered to express appropriate Phase 1 and Phase 2 metabolism should be considered.

This collaborative research program will be a major exercise, but if it is not undertaken we will still be faced with an unacceptable level of false positives and the consequential unnecessary follow-up *in vivo* testing, for decades to come. A steering committee, consisting of genotoxicity, metabolism and chemistry experts from academia, industry and the regulatory agencies should be established to draft a plan for this program. The participants did not reach any conclusion as to how this trial should be organised, but it is hoped that ECVAM may be able to contribute, to liase with other related activities from organisations such as ILSI-HESI and to begin to identify sources of funding and support.

4. Conclusions

The workshop participants agreed that some of the commonly used cells for genotoxicity testing (in particular some of the rodent cell lines) have produced an unacceptably high level of false positive results when compared with known *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens. In most of the rodent cell lines used, deficiencies in metabolism, *p53* function and DNA repair capability almost certainly contribute to this high false positive rate. Better guidance on the likely mechanisms resulting in positive results that are not relevant for humans, and on how to obtain evidence for those mechanisms, is needed both for practitioners and regulatory reviewers.

Testing up to high concentrations and high levels of cytotoxicity as is currently required in mammalian cell genotoxicity tests are also likely to contribute to the high frequency of false positive results, and may not be justified. A thorough review of published and industry data to determine whether such levels are required for the detection of *in vivo* genotoxins and DNA-reactive, mutagenic carcinogens is urgently needed. Suggestions to lower the current upper limit (perhaps by 10-fold or more) may be justified in terms of metabolic and cellular processes, and human tissue exposures. This needs urgent but careful evaluation.

Various measures of cytotoxicity are currently allowable under OECD guidelines, but there is little comparative data on whether different measures would select different concentrations for test. A detailed comparison of multiple measures of cytotoxicity, in relation to endpoints such as clastogenicity, is needed. Also, genotoxicity endpoints that are not intrinsically linked with processes leading to cytotoxicity need to be developed.

There was agreement amongst the workshop participants that cell systems preferably of human origin, which are *p53* and DNA-repair proficient, and have defined Phase1 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 in the future. 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, and 3D skin models based on primary human keratinocytes also show some promise. However, much effort will be required to introduce a broader spectrum of metabolic capabilities into these or other target cells. Other human cell lines such as HepaRG have not been used for genotoxicity investigations and should be studied. A collaborative research programme is needed to identify and evaluate new cell systems with appropriate sensitivity but improved specificity.

Perhaps most importantly, the participants in this workshop felt that it is time for the scientific community to give careful consideration to the carcinogens and *in vivo* genotoxins we expect any new assay, or modified existing assay, to detect. Rodent bioassays have flaws in terms of detecting human carcinogens, and this is com-

pounded by trying to use genotoxicity assays to detect as many rodent carcinogens as possible. This has led to the continued expansion of genotoxicity test protocols (e.g. addition of new treatment and sampling regimens, new strains of bacteria, etc.), and the combination of these tests in batteries. The consensus of the group was that there is a need to refocus on the detection on human carcinogens and *in vivo* genotoxins that are DNA reactive. Without this change of focus, the development and validation of any new methods or assays will likely not lead to an overall improvement.

References

[1] D. Kirkland, M. Aardema, L. Henderson, L. Müller, Evaluation of the ability of a battery of 3 *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity, Mutat. Res. 584 (2005) 1–256.

[2] E.J. Matthews, N.L. Kruhlak, M.C. Cimino, R.D. Benz, J.F. Contrera, An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data. I. Identification of carcinogens using surrogate endpoints, Regul. Toxicol. Pharmacol. 44 (2006) 83–96.

[3] B. Halliwell, Oxidative stress in cell culture: an under-appreciated problem? FEBS Lett. 540 (2003) 3–6.

[4] L.M. Wee, L.H. Long, M. Whiteman, B. Halliwell, Factors affecting the ascorbate- and phenolic-dependent generation of hydrogen peroxide in Dulbecco's modified Eagles medium, Free Radic. Res. 37 (2003) 1123–1130.

[5] L.H. Long, D. Kirkland, B. Halliwell, Different cytotoxicities of epigallocatechin gallate or ascorbate in various cell culture media due to variable rates of oxidation in the culture medium, Mutat. Res., submitted for publication.

[6] A. Santoro, M.B. Lioi, J. Monfregola, S. Salzano, R. Barbierri, M.V. Ursini, L-Carnitine protects mammalian cells from chromosome aberrations but not from inhibition of cell proliferation induced by hydrogen peroxide, Mutat. Res. 587 (2005) 16–25.

[7] D.J. Tweats, D.G. Gatehouse, Further debate of testing strategies, Mutagenesis 3 (1988) 95–102.

[8] S. Coecke, M. Balls, G. Bowe, J. Davis, G. Gstraunthaler, T. Hartung, R. Hay, O.-W. Merten, A. Price, L. Schechtman, G. Stacey, W. Stokes, ECVAM good cell culture practice task force report 2, ATLA 33 (2005) 261–287.

[9] OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring. No. 14 on the Application of the Principles of GLP to *in vitro* studies, 2004, http://appli1.oecd.org/olis/2004doc.nsf/43bb6130e5e86e5fc12569fa005d0044c/58d5c8b13297a995c1256f5c006008a7/$FILE/JT00174939.DOC.

[10] T. Sofuni, A. Matsuoka, M. Sawada, M. Ishidate Jr., E. Zeiger, M.D. Shelby, A comparison of chromosome aberration induction by 25 compounds tested by two Chinese hamster cell (CHL and CHO) systems in culture, Mutat. Res. 241 (1990) 175–213.

[11] S.K. Greenwood, R.B. Hill, J.T. Sun, M.J. Armstrong, T.E. Johnson, J.P. Gara, S.M. Galloway, Population doubling: a simple and more accurate estimation of cell growth suppression in the *in vitro* assay for chromosomal aberrations that reduces irrelevant positive results, Environ. Mol. Mutagen. 43 (2004) 36–44.

[12] S. Meintières, A. Biola, M. Pallardy, D. Marzin, Using CTLL-2 and CTLL-2 *bcl2* cells to avoid interference by apoptosis in the *in vitro* micronucleus test, Environ. Mol. Mutagen. 41 (2003) 14–27.

[13] F.P. Guengerich, G.A. Dannan, S.T. Wright, M.V. Martin, L.S. Kaminsky, Purification and characterization of liver microsomal cytochromes P-450: electrophoretic, spectral, catalytic and immunochemical properties and inducibility of eight isozymes isolated from rats treated with phenobarbital or beta-naphthoflavone, Biochemistry 21 (1982) 6019–6030.

[14] M.O. Bradley, B. Bhuyan, M.C. Francis, R. Langenbach, A. Peterson, E. Huberman, Mutagenesis by selected agents in V79 Chinese hamster cells: a review and analysis of the literature—a report of the Gene-Tox Program, Mutat. Res. 87 (1981) 81–142.

[15] H.R. Glatt, W. Meinl, Sulfotransferases and acetyltransferases in mutagenicity testing: technical apsects, Meth. Enzymol. 400 (2005) 230–249.

[16] H.R. Glatt, Activation and inactivation of carcinogens by human sulfotransferases, in: G.M. Pacific, M.W.H. Coughtrie (Eds.), Human Sulphotransferases, Taylor & Francis, London, 2005, pp. 281–306.

[17] S. Wilkening, F. Stahl, A. Bader, Comparison of primary human hepatocytes and hepatoma cell line HepG2 with regard to their biotransformation properties, Drug. Metab. Dispos. 31 (2003) 1035–1042.

[18] S. Knasmüller, C. Cavin, A. Chakraborty, F. Darroudi, B.J. Majer, W.W. Huber, V.A. Ehrlich, Structurally related mycotoxins ochratoxin A, ochratoxin B, and citrinin differ in their genotoxic activities and in their mode of action in human-derived liver (HepG2) cells: implications for risk assessment, Nutr. Cancer 50 (2004) 190–197.

[19] Y.C. Staal, M.H. van Herwijnen, F.J. van Schooten, J.H. van Delft, Modulation of gene expression and DNA adduct formation in HepG2 cells by polycyclic aromatic hydrocarbons with different carcinogenic properties, Carcinogenesis 27 (2006) 646–655.

[20] S. Knasmüller, V. Mersch-Sundermann, S. Kevekordes, F. Darroudi, W.W. Huber, C. Hoelzl, J. Bichler, B.J. Majer, Use of human-derived liver cell lines for the detection of environmental and dietary genotoxicants; current state of knowledge, Toxicology 198 (2004) 315–328.

[21] V. Mersch-Sundermann, S. Knasmüller, X.J. Wu, F. Darroudi, F. Kassie, Use of a human-derived liver cell line for the detection of cytoprotective, antigenotoxic and cogenotoxic agents, Toxicology 198 (2004) 329–340.

[22] J.H. van Delft, E. van Agen, S.G. van Breda, M.H. Herwijnen, Y.C. Staal, J.C. Kleinjans, Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling, Carcinogenesis 25 (2004) 1265–1276.

[23] C.L. Crespi, W.G. Thilly, Assay for gene mutation in a human lymphoblastoid line, AHH-1, competent for xenobiotic metabolism, Mutat. Res. 128 (1984) 221–230.

[24] C.L. Crespi, F.J. Gonzalez, D.T. Steimel, T.R. Turner, H.V. Gelboin, B.W. Penman, R. Langenbach, A metabolically competent human cell line expressing 5 cDNAs encoding procarcinogen-activating enzymes: application to mutagenicity testing, Chem. Res. Toxicol. 4 (1991) 566–572.

[25] P.A. White, G.D. Douglas, J. Gingerich, C. Parfett, P. Shwed, V. Seligy, L. Soper, L. Berndt, J. Bayley, S. Wagner, K. Pound, D. Blakey, Development and characterization of a stable epithelial cell line from Muta™Mouse lung, Environ. Mol. Mutagen. 42 (2003) 166–184.