4. Non-reproducible or marginal results of low concern

Occasionally, an increase in the measured genotoxicity parameter is seen that is considered significant by widely accepted criteria, such as a statistically significant increase over a negative control, or a value greater than a given fold change, *e.g.*, 2- or 3-fold, but the increase is weak or marginal. Subsequent repeats of the assay may fail to reproduce the increase. Other information, such as from other assays with a similar endpoint, or metabolism studies indicating no potentially reactive metabolite, may suggest that the singular marginal response is not a significant concern. Case Study 1 (Appendix A) presents such an example, where an initial statistically significant increase in the number of human lymphocytes with chromosome aberrations with values slightly outside the historical control range was not reproduced, and therefore was judged not to be a significant concern. The Working Group agreed that in such cases, the marginal, non-reproducible increase would be considered of no further concern for genotoxicity, and no testing beyond the standard battery of assays for that type of substance would be required. Case Study 2 (Appendix B) presents another case where there was a significant increase in the number of cells with chromosome aberrations just outside the historical control range. The Working Group concluded that it would be of low concern for human safety if repeat testing verified a weak borderline effect or showed a negative result.

There are also instances when an assay produces marginal or weak increases in response to chemical exposure, and these results are reproducible. Alternatively the chemical may produce a combination of weak and negative responses. There are several considerations that can help resolve whether these weak or equivocal results require follow-up testing. GLP study protocols generally specify in some detail conditions under which a study will be classified as positive, negative, or equivocal. While most protocols give some latitude to the professional judgment of the study director, evaluation criteria are often fairly rigid. For example, for cytogenetics studies, the criteria from one representative contract laboratory are:

> "The test article will be considered to induce a positive response when the percentage of cells with aberrations is increased in a dose-responsive manner with one or more concentrations being statistically significant and clearly outside the historical solvent control data ($p \leq 0.05$). However, values that are statistically significant but do not exceed the range of historical negative or solvent controls may be judged as not biologically significant."

Table 1
Example of a statistically but not biologically significant chromosomal aberration result due to low control values

Test article concentration	Mean (%) mitotic index	Mean (%) abnormal cells[a]
0	10	0.5
10 μg/mL	9	1
20 μg/mL	6	2
30 μg/mL	5	5.5*
Positive control	6	12

Negative historical control range 0–5%.
[a] Mean of two cell cultures: cells with structural chromosome aberrations excluding gaps.
* $p \leq 0.05$.

These criteria can lead to situations where, if one has a low value for the concomitant vehicle control, and the values for the low and mid doses of the test article appear to be dose-responsive but are well within the historical range for the vehicle control, the high dose becomes statistically different from the control but just outside the range of historical control value. An example of such data is given in Table 1.

The Working Group agreed that in case of non-reproducible or marginal results, it is advisable to develop a weight of evidence approach, considering the following points:

- The dose-response relationship needs to be examined – less concern is raised if the marginal response is not part of an increasing dose-response (*i.e.*, it is not dose-related).
- If the marginal increase is seen at high cytotoxicity (*e.g.*, approaching 50% or greater cytotoxicity in a cytogenetics assay, or >80% in the mouse lymphoma assay) but no increase is seen at lower, more moderate, toxicity, then there is less concern.
- Comparison of the magnitude of the marginal increase to historical negative control observations can help assess the probability that the result occurred by chance. Marginal responses within the historical negative control range (particularly at high concentrations and/or high toxicity) are of less concern. The Working Group agreed that more weight should be given to the historical data, and that better definition of how to use historical control data was needed. Preliminary comments on the use of historical control data are included in Appendix E. However, there is a need for further discussions in the future. Alternative approaches, such as selecting the 95 or 99% upper confidence interval

to define the range, or use of data distribution including median values, instead of the absolute upper limit, may be more appropriate. Also, it should be discussed whether observations in individual cultures rather than group means are sometimes more appropriate.

- Corroborating data can be sought from other experiments. Failure to confirm the marginal increase between replicates or experiments of the same assay or between different assays examining similar endpoints indicates a lower concern.
- Structure activity data can be useful. If no structural alerts are found for the chemical and if there is some evidence that the test compound is not a DNA damaging agent (*e.g.*, Ames test negative), then the marginal increases becomes of less concern.

Weight of evidence approaches have previously been described by Brusick et al. [43,44].

Considering the points listed above, the development of such weight of evidence arguments can lead to a conclusion that, for a marginal or equivocal response, there is a low or no level of concern, and no further testing is necessary, except possibly a repeat experiment using similar experimental conditions to check the reproducibility.

5. Follow-up strategy for a clear positive assay result

In some cases a clear and reproducible positive *in vitro* result is seen, yet the other assays in the initial battery, including any required *in vivo* test, are negative. The *in vitro* result is not automatically overruled by the negative *in vivo* result, and some follow-up testing or investigation is generally necessary to determine the relevance of the *in vitro* positive result. For example, the ICH scheme [6,7] suggests follow-up testing with a second *in vivo* test in addition to the *in vivo* cytogenetics test in the initial regulatory battery. It might be assumed that the concern about the positive *in vitro* result lessens as the number and types of negative *in vivo* assay results increase. However, this assumption may not be valid since the *in vivo* assays may have different sensitivities and /or evaluate different genotoxic endpoints. It is important that relevant endpoints are examined in the most relevant tissues *in vivo*.

To understand the basis for a positive *in vitro* result in the absence of a corresponding *in vivo* result, follow-up testing may require only a few additional studies (or tests), or more extensive research. Regardless of the question(s), the testing should be based on the full knowledge of the chemical, its physico-chemical and toxicological properties, and anticipated human exposure scenarios. An understanding of the type(s) of genotoxic insult(s) induced and the nature of the response(s), with any indications of possible mechanism, is crucial. Aspects, such as formation of DNA adducts or strand breaks, involvement of reactive oxygen or nitrogen species, nucleotide pool imbalance, inhibition of DNA synthesis or topoisomerases, and disruption of mitotic spindle need to be considered in order to identify rational testing approaches to pursue. Because each situation is likely to be different, follow-up testing is not amenable to a "one size fits all" approach, and flexibility is important to determine the most appropriate follow-up strategy to pursue.

A mode of action approach is used to determine whether a chemical that has intrinsic genotoxic properties might lead to an adverse effect, such as cancer [24]. The mode of action approach takes into account all available genotoxicity information and, in combination with other available information (*e.g.*, structure activity data, pharmacokinetics, ADME (absorption, distribution, metabolism and excretion) data, other biological responses, *etc.*) helps characterize whether a chemical is likely or not to pose a risk for exposed humans. The mode of action approach allows a full examination of whether the singular positive assay result should warrant a concern for human risk or not.

Case Studies 3 (Appendix C) and 4 (Appendix D) exemplify the weight of evidence and mode of action approaches in case of clear positive results. Case Study 4 is an example of a potential aneugen which, as discussed in several publications in recent years [33,34,37,38], might be considered to act *via* an indirect mechanism with a non-linear dose-response relationship suggesting a threshold.

Further discussion and recommendations on follow-up testing in the case of clearly positive results are given below.

The workshop discussion and case studies presented in the appendices led to the development of a framework for follow-up testing (displayed in Fig. 1). As discussed above, criteria for determining when follow-up testing beyond the initial battery is needed were agreed upon. When follow-up testing is needed, it should be based on all available information on chemical structure and mode of action, and also information learned from the nature of the response observed in the initial tests and anticipated human exposure patterns. The nature of these results (endpoint, magnitude, association with toxicity, *etc.*), and available information about the biochemical and pharmacological nature of the agent, are generally sufficient to conclude that the results observed are consistent with certain mechanisms and inconsistent

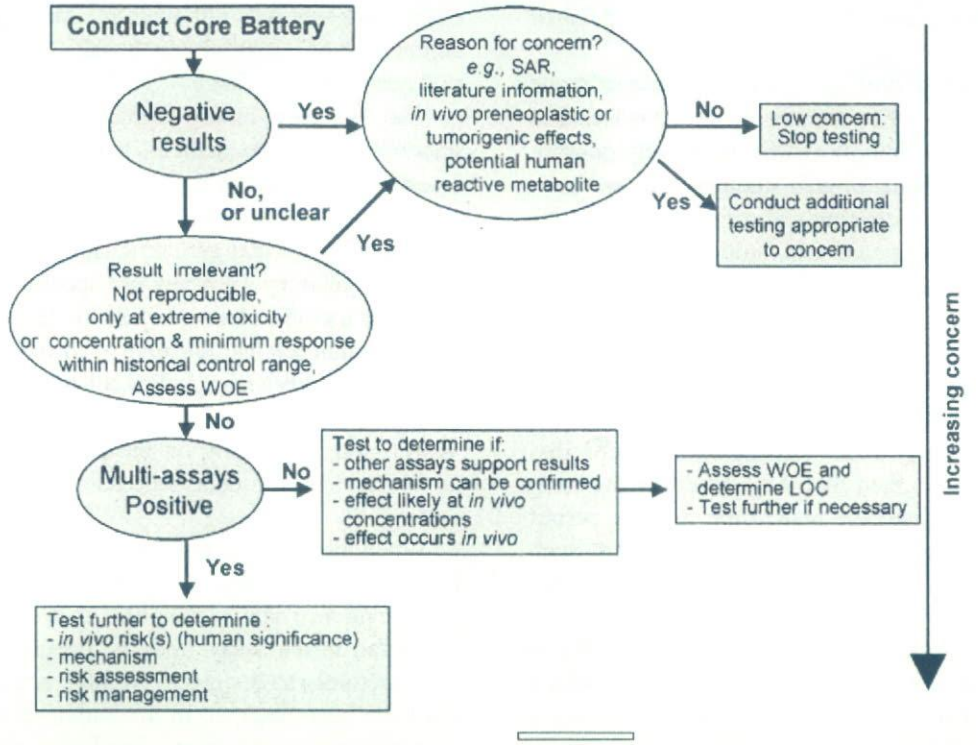

Fig. 1. Framework for interpretation of standard battery and follow-up test results. SAR = structure activity relationship, WOE = weight of evidence, LOC = level of concern.

with others. For example, it may be possible to determine whether such factors as DNA adducts or DNA strand breaks, involvement of reactive oxygen or nitrogen species, nucleotide pool imbalance and/or nucleotide mis-incorporation, interference with cell cycle kinetics, inhibition of DNA synthesis or topoisomerases, or mitotic spindle disruption are likely to be involved in the mode of action.

The different tests in the battery are selected because they measure different genotoxic effects and different genotoxic mechanisms of action, thereby providing important mechanistic information. For example, generation of chromosomal aberrations involves strand breakage and rejoining, reversion of the Ames' *Salmonella* tester strains containing specific base substitutions requires base mutations that may arise from DNA alkylation or mis-repair of bulky adducts, and reversion of the Ames' strains containing frameshift mutations requires induction of a second frameshift that is characteristic of intercalating agents and other classes, but not small molecular weight alkylating agents. These types of information should be used to guide the interpretation of the results and the selection of follow-up tests. *In vivo* endpoints should be chosen to reflect the types and the mechanisms of damage found in the initial screening battery.

It is recognized that some genotoxic agents act through disruption of cellular biochemistry that occurs only above certain treatment levels (for example, see [40]). They therefore exhibit a dose-response that is expected to be non-linear, and exhibit a threshold below which there will not be a concern. Examples of such agents include those that induce nucleotide pool imbalances (without incorporation of a nucleotide that can cause mis-pairing or strand termination at concentrations relevant to use), mitotic spindle disruption, inhibition of DNA synthesis, topoisomerase inhibition, *etc.* These agents are often referred to as "indirect gentoxins" or "non-DNA-reactive gentoxins", in contrast to those thought to act by primary reactivity with DNA. It is generally accepted that those indirect mechanisms or modes of action lead to non-linear response curves, sometimes reported as threshold effect. The Working Group recognized that when an agent can be demonstrated to act though such an "indirect" mechanism, and a dose can be established below which the effect is not observed or deemed not a concern, then appropriate margins of safety for exposure below that dose can be established. This means that human exposures sufficiently below exposures associated with the no observed effect dose should not present a significant genotoxic risk. These issues have been discussed in a special

issue of Mutation Research edited by Parry and Sarrif [45].

It will be easier to build weight of evidence or mode of action arguments when the positive result is found in a test with low specificity (*i.e.*, tests frequently positive but not confirmed in the *in vivo* studies), such as the *in vitro* chromosomal aberration test or mouse lymphoma assay. Agents that are positive in multiple tests with different endpoints, especially if they are positive *in vivo*, impart the highest level of concern and would require extensive investigations to develop a sufficient weight of evidence to establish conditions under which human exposure might be permissible.

Although these general considerations were discussed, and it was agreed that they should be the basis of selection of appropriate tests, time did not permit development of more detailed recommendations, such as identification of specific tests to follow-up on specific outcomes in standard batteries. It was agreed that consensus recommendations about specific test options would be useful, and it was suggested that the Working Group should meet again in the near future to develop such recommendations.

Considerations when choosing an *in vivo* assay as a follow-up test to an initial finding include choice of the appropriate tissue(s) in which evaluations should be performed. This depends on anticipated route(s) of exposure, tissue distribution (often known from other toxicology or pharmacokinetic studies), and metabolic degradation and/or activation in various tissues as well as target organ in long-term toxicity studies. It was noted that such considerations might dictate the use of non-standard studies (*i.e.*, not in the regulatory initial battery or standard follow-up studies but sufficiently validated) and that in those cases the use of non-standard studies would be preferable to a standard *in vivo* assay in which the endpoint or target tissue is not relevant. It was further noted that when metabolic modification of an effect is observed to occur, consideration should be given to human metabolism in relation to the laboratory models being considered. Likewise, the relationship between exposure-response information in laboratory models to blood and tissue levels from human exposures is important, and it is desirable to determine this relationship whenever possible.

6. Conclusions

A summary of the discussion and principal conclusions reached by the Working Group is as follows:

The Working Group agreed to limit the focus of the meeting to somatic cell risk in humans, with emphasis on carcinogenic risk, although the group recognized the importance of potential health risks from both somatic and germ cell mutations.

The Working Group agreed not to discuss the improvement of the standard batteries, but to focus on interpretation and appropriate follow-up testing for the tests currently in use in the standard battery. The Working Group noted that genetic toxicity test batteries vary among regulatory agencies and specific situations, but generally include a bacterial test for gene mutations and a mammalian cell test sensitive to chromosomal damage (most commonly a mammalian cell test for chromosomal aberrations and/or the mouse lymphoma $tk^{+/-}$ mutation assay), and, depending on the products and their use, an *in vivo* test in rodent bone marrow for chromosome damage.

When the initial test battery is clearly negative, there is generally no need for further testing unless there is (a) evidence that metabolites that differ from those generated in the assays may be present in the human, (b) structural alerts suggest possible activity that would not have been detected in the battery employed, or (c) evidence from the literature, previous experience, or subsequent test results suggests possible activity that would not have been identified in the battery. In these cases, follow-up testing may be indicated, and appropriate tests should be selected on the basis of the information that raised the concern.

When a non-reproducible or marginal *in vitro* positive result is obtained, and results from other assays with a similar endpoint are negative, the weight of evidence should be considered to determine if further testing is necessary or whether, based on the available data, the evidence suggests a low level of potential risk that does not require further testing. Factors that may suggest lower concern include: (a) weak effects without a strong dose relationship and values within or close to a range that could occur by chance variability (negative control historical data), (b) effects that occur only at very high levels of cytotoxicity, but not at moderate levels, in the chromosomal aberration or mouse lymphoma $tk^{+/-}$ assays (*e.g.*, approaching 50% or greater cytotoxicity in the chromosome aberration test, or >80% in the mouse lymphoma assay), (c) results that are not consistently repeatable, and (d) the absence of structural alerts or any other cause of concern. In most cases, the result is not of concern and no testing beyond the standard battery for that type of substance will be required.

When a clear positive result is obtained in *in vitro* test battery, further testing is generally indicated in order to provide a sufficient body of evidence to determine the mode of action, relevance to the human expo-

sure situation, and potential human health risk. Such testing should be based on the knowledge available about the nature and/or mode of action of the original response, *e.g.*, whether the initial result was consistent with DNA adducts versus strand breaks, involvement of reactive oxygen or nitrogen species, pool imbalance, spindle disruption, *etc.* If the evidence suggests an "indirect" mode of action (not involving direct or proximate reactivity with the DNA), such as nucleotide pool imbalance, spindle disruption, inhibition of DNA synthesis or topoisomerases, *etc.*, then tests that provide additional evidence that supports this hypothesis and that rule out direct DNA reactivity should be selected. It is often useful to determine if a positive result, found only under a specific condition in one test, is confirmed in further testing in other assays evaluating the same endpoint (*e.g.*, chromosome aberration test versus mouse lymphoma assay).

When choosing an *in vivo* assay for follow-up testing, an endpoint appropriate to the nature of the original response observed should be selected, and in addition, due consideration should be given to the route of human exposure, the expected tissues and times of highest exposure (*i.e.*, pharmacokinetic considerations), and the potential for metabolic activation and deactivation in various tissues. Although regulatory guidelines often mention *in vivo* assays that may possibly be employed for follow-up testing, non-standard studies supported by peer-reviewed published literature may, when justified, be more appropriate and informative than standard assays. The Working Group acknowledged that there are cases where mechanistic studies can be considered sufficient to support the indirect mode of action, and therefore additional *in vivo* data may not be necessary.

When addressing the relevance of findings to human health risks, the total weight of evidence should be considered. In addition to the factors noted above, when modification of an effect by a mammalian metabolizing system is observed, consideration should be given to the metabolism of the agent in the human relative to the laboratory model. Whenever possible, exposure-response information for comparison to blood and tissue levels from human exposures is desirable.

Remaining issues that were identified during the workshop, but not addressed due to time limitations, include the development of recommendations for the selection of specific assays, particularly *in vivo* assays, and the appropriate use of historical control data. It is anticipated that recommendations on these issues will be developed at a future meeting of this Working Group. Similarly, the evaluation of structural alerts was identified as a useful component of the weight of evidence approach, but not discussed in detail. This topic will have to be re-visited in the future.

Disclaimer

This manuscript represents the views of the authors and the contents do not necessarily reflect the views or policies of the authors' agencies or institutions. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

Appendix A. Case Study 1: Compound with a weak non-reproducible increase in chromosome aberration test

Compound A is an early orally applied drug development candidate for a non-life threatening therapeutic indication and low systemic exposure. There was no evidence of increased revertant numbers in the Ames test up to the limit dose level of 5000 μg/plate, using the plate incorporation method in the presence and in the absence of metabolic activation on *Salmonella typhimurium* TA98, TA100, TA1535, TA1537 and *E. coli* WP2 uvrA.

Compound A was tested for clastogenic activity in human lymphocyte cultures. Chromosome aberrations were evaluated by metaphase analysis after 3 h treatments with and without metabolic activation at concentrations ranging from 47.2 to 84.0 μg/mL and 10.6 to 75.0 μg/mL, respectively. In addition, chromosome aberrations were evaluated after 24 h treatment without metabolic activation at concentrations ranging from 8.30 to 33.3 μg/mL. In all tests, the highest test concentration evaluated produced a 48 to 59% reduction of the mitotic index. There were no significant increases in the number of cells with chromosome aberrations at any concentration evaluated in either the 3 or 24 h tests without metabolic activation. However, Compound A produced a dose-related, statistically significant increase in the number of cells with chromosome aberrations in the initial test with metabolic activation at the two highest concentrations (58 and 68.2 μg/mL); only the highest concentration produced a value outside the range of the historical data, *i.e.*, 5.5% abnormal cells as compared to the historical control range of 0–4% (Table 2, Test 1). A confirmatory 3 h test with metabolic activation using blood of a second donor produced a statistically significant increase in chromosome damage only at the highest concentration of 84.0 μg/mL, the observed value being within the acceptable range of the historical negative control data, *i.e.*, 3.5 versus 0–4% (Table 2, Test 2). This response was not clearly reproducible between

Table 2
Lack of reproducibility of a weak increase in structural chromosomal aberrations in cultured human lymphocytes with Compound A in three independent tests with metabolic activation

Compound A (μg/mL)	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression
Test 1		
1% DMSO	0.5	0
49.3	2.5	29
58.0	4.0*	52
68.2	5.5*	49
CP[b]	29.0*	45
Test 2		
1% DMSO	0.5	0
47.2	1.0	0
63.0	0.5	14
84.0	3.5*	48
CP[b]	24.3*	31
Test 3		
1% DMSO	2.5	0
52.4	1.5	21
65.5	1.0	36
81.9	2.5	56
102.0	2.0	58
128.0	1.5	65
CP[b]	58.0*	59

Negative historical control range 0–4%.
[a] Mean of two cell cultures: cells with structural chromosome aberrations excluding gaps.
[b] Cyclophosphamide at 5 μg/mL.
* Statistically significant (one-tailed Fisher's exact test $p < 0.05$).

the replicate cultures and the negative control response was low (0.5% abnormal cells as compared to the historical control range of 0–4%). A follow-up 3 h exploratory cytogenetics study with metabolic activation using blood of a third donor was performed on a larger number of concentrations in order to better characterize the cytotoxicity profile of Compound A in the lymphocyte test system. The results of this follow-up test did not reproduce the previous results using the same S9 batch but different blood donors and failed to show an increase in chromosomal aberrations at concentrations ranging from 52.4 to 128 μg/mL, which produced a 21 to 65% reduction in the mitotic index (Table 2, Test 3) and 13 to 70% reduction in cellular ATP levels (not shown). Based on the overall test results, Compound A was concluded as being equivocal for the induction of structural chromosome aberrations.

Compound A was tested for the induction of micronuclei in male and female rat bone marrow cells *in vivo*. Male and female rats were administered the vehicle only as the negative control, or Compound A at dose levels of 15, 30, and 60 mg/kg in males, and 7.5, 15, and 30 mg/kg in females by oral gavage once a day for 2 consecutive days. The numbers of micronucleated (MN) PCE in any of the test groups did not significantly increase. In conclusion, Compound A did not induce chromosome damage in the bone marrow cells of male or female rats when tested up to estimated maximum tolerated doses of 60 and 30 mg/kg, respectively. The mean maximum serum concentrations occurred 3.5 h after dosing, achieving values of 645 and 620 ng/mL in males and females, respectively, and the mean $AUC_{0-24\,h}$ exposures were 8560 and 9030 ng h/mL, respectively. The serum concentration in this study was expected to exceed the pharmacologically active concentration by approximately 60-fold.

The Working Group considered that the original increase in the number of human lymphocytes with chromosome aberrations was weak, and was not clearly reproducible in follow-up testing. Based on the overall profile of this compound in the standard genetic toxicology test battery, the Working Group concluded that the genotoxicity findings of Compound A were of low concern for human safety and did not require further follow-up testing.

Appendix B. Case Study 2: Compound with marginal increase in the number of cells with chromosome aberrations at a cytotoxic concentration

Compound B is an early drug candidate in a chronic use indication with inhaled drug application and low systemic exposure. It tested negative in the Ames assay up to levels of compound insolubility and the gene mutation test in CHO cells at *hprt* locus up to cytotoxic concentrations. Compound B did not induce chromosomal aberrations in cultured human lymphocytes when tested in a 24 h exposure without metabolic activation, or a 3 h exposure with metabolic activation (Table 3). However, when it was tested in a 3 h exposure without metabolic activation, a single concentration that induced 58% mitotic suppression produced a statistically significant increase in chromosomal aberrations just outside of the historical control range (5% of cells with chromosome aberrations versus 0–4% acceptable range for negative control data). Compound B was also tested in a mouse bone marrow micronucleus assay up to a maximally tolerated dose and did not lead to an increase in the incidence of MN PCE in the bone marrow. Toxicity was seen in the bone marrow indicating compound reached this tissue.

Based on the initial results from the 3 h exposure, the Working Group came to the consensus that the response was of low concern to human safety since it occurred

Table 3
Weak increase in structural chromosomal aberrations in cultured human lymphocytes at cytotoxic concentrations of Compound B

Compound B (μg/mL)	3 h direct		24 h direct		3 h + S9	
	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression
0.1% DMSO	0	0	2	0	1	0
25	–	–	1	15	–	–
28	–	–	–	–	0	20
32	–	–	3	30	–	–
35	–	–	–	–	1.5	32
41	–	–	2	55	–	–
44	–	–	–	–	1	54
101	0.5	8	–	–	–	–
120	0	28	–	–	–	–
144	5*	58	–	–	–	–

Negative historical control range 0–4% in 3 h direct test.
[a] Mean of two cell cultures: cells with structural chromosome aberrations excluding gaps.
* Statistically significant (one-tailed Fisher's exact test $p < 0.05$).

only at a high concentration that produced more than 50% cytotoxicity in human lymphocytes. Importantly, a concentration that produced 28% mitotic suppression did not generate a significant induction of aberrations. In addition, the increase was just outside the historical control range, and the compound did not produce chromosomal aberrations in the corresponding 24 h exposure. The Working Group suggested that a repeat of the positive arm of the test, to evaluate reproducibility, might contribute further to the assessment of Compound B. Taking into account the steep cytotoxicity dose-response curve, a new test should, if possible, include a concentration producing between 28 and 58% reduction in mitotic index.

If a similar pattern of results, or negative results, are obtained in the repeat test, the Working Group would consider that no further testing would be needed, and would conclude the response to be of low concern for human safety.

Appendix C. Case Study 3: Development of weight of evidence approach with a clearly positive result in cytogenetics assays

Compound C is a topically applied early drug development candidate intended for short term treatment of dermal scarring. It produced high local concentrations at the target site in the skin but was rapidly cleared systemically and showed only very low systemic exposure. There was no evidence of increased revertant numbers in the Ames test up to the limit dose level of 5000 μg/plate, using the plate incorporation method in the presence and in the absence of metabolic activation on *Salmonella typhimurium* TA98, TA100, TA1535, TA1537 and *E. coli* WP2 uvrA.

Compound C was tested for clastogenic activity in human lymphocyte cultures (Table 4). Chromosome damage was evaluated by metaphase analysis after 3 h with and without metabolic activation at concentrations ranging from 154 to 240 μg/mL and 33.4 to 52.2 μg/mL, respectively. In addition, chromosome damage was evaluated after 24 h without metabolic activation at concentrations ranging from 6.30 to 9.84 μg/mL. In all the tests, the highest concentrations produced a 52 to 57% reduction of the mitotic index. Compound C did not induce a significant increase in the number of abnormal cells at any concentration evaluated when treated in the presence of metabolic activation. In the absence of metabolic activation, Compound C produced concentration-related statistically significant increases in aberrant cells over a range of concentrations, which spanned both cytotoxic and non-cytotoxic test conditions in both 3 and 24 h exposures, producing 6.5 and 7.5% abnormal cells at 3 and 24 h, respectively.

Compound C was tested for the induction of micronuclei in Chinese hamster ovary (CHO-WBL) cells (Table 5). A slight increase (1.7- to 2.3-fold above mean background) was observed in the 24 h test without metabolic activation, whereas more substantial increases of 2.3- to 4.9-fold were observed in the test with activation. An additional 3 h test without activation was performed in response to the slight increase observed in the 24 h test and a significant increase (3.1-fold above mean background) was observed in cultures treated with concentrations of 68.5 and 107 μg/mL. Kinetochore analysis was performed to assess whether the

Table 4
Induction of structural chromosomal aberrations in cultured human lymphocytes by Compound C

Compound C (μg/mL)	3 h direct		24 h direct		3 h + S9	
	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression	Mean (%) abnormal cells[a]	Mean (%) mitotic suppression
0.1% DMSO	1.5	0	0.5	0	1	0
6.3	–	–	2	8	–	–
7.9	–	–	5.5*	20	–	–
9.8	–	–	7.5*	52	–	–
12.3	–	–	–	–	–	–
33.4	5.5*	23	–	–	–	–
41.8	6.0*	41	–	–	–	–
52.2	6.5*	56	–	–	–	–
154	–	–	–	–	1	16
192	–	–	–	–	0.5	34
240	–	–	–	–	3	57

Negative historical control range 0–4% in 3 h direct test.
[a] Mean of two cell cultures: cells with structural chromosome aberrations excluding gaps.
* Statistically significant (one-tailed Fisher's exact test $p < 0.05$).

observed increases were the result of structural (clastogenic) or numerical (aneugenic) chromosomal events (Table 6). Over 140 micronuclei induced by a 3 h treatment with metabolic activation at soluble concentrations ranging from 107 to 168 μg/mL were evaluated using anti-kinetochore antibody staining technique. A 3- to 5-fold increase in micronucleated cells above DMSO-treated controls was again observed and these cells were predominately kinetochore negative indicating a clastogenic response.

Induction of micronuclei *in vivo* in rat bone marrow PCE was evaluated after continuous intravenous (IV) infusion at a rate of 2 mL/kg/h for approximately 24 h up to a maximum feasible dose of 655 mg/kg. Approximately 24 and 48 h following completion of the infusion, bone marrow was extracted and assessed for the induction of MN PCE. There was no bone marrow cytotoxicity and no statistically significant increases in MN PCE at any dose level or harvest time (data not shown). Mean plasma concentrations of Compound C measured after approximately 24 h of infusion were 2.93 and 3.20 μg/mL in males and females, respectively. Attempts to assess target (bone marrow and liver) tissue exposure to Compound C resulted in concentrations below assay detection limits, possibly attributed to tissue metabolism and rapid clearance. Plasma exposures

Table 5
Induction of micronuclei by Compound C in Chinese hamster ovary cells

24 h test without metabolic activation			3 h test without metabolic activation			3 h test with metabolic activation		
Compound C (μg/mL)	Cytotoxicity[a]	% MN cells[b]	Compound C (μg/mL)	Cytotoxicity[a]	% MN cells[b]	Compound C (μg/mL)	Cytotoxicity[a]	% MN cells[b]
0	0	1	0	0	1.2	0	0	0.8
1.8	17.5	1.7	68.5	33.6	3.7*	134	0	3*
3.5	25.2	1.7	107	45.2	3.7*	168	39.7	1.9*
7.0	45.9	2.3*	168	67.3	1.2	210	24.5	3.9*
14.0	81.1	INS	210	76.3	INS	262	T	T
POS[c]	15.4	9.6*	POS[d]	43.3	21.7*	POS[e]	0	5.0*

INS: insufficient number of cells for evaluation. T: toxic, less than 20% cells when compared to negative control
[a] Cytotoxicity = $100 - 100 \times \{(CBPI_T - 1/CBPI_C - 1)\}$, with CBPI = (no. of mononucleated cells + 2 × no. of binucleated cells + 3 × no. of multinucleated cells)/total number of cells.
[b] Percent micronucleated cells; evaluating a minimum of 1000 binucleated cells.
[c] Positive control mitomycin C at 0.05 μg/mL.
[d] Positive control mitomycin C at 0.4 μg/mL.
[e] Positive control cyclophosphamide at 10 μg/mL.
* Statistically significant (one-tailed Fisher's exact test $p < 0.05$).

Table 6
Classification of micronuclei induced by Compound C in Chinese hamster ovary cells using kinetochore staining

Compound C (μg/mL)	Cytotoxicity[a]	% MN[b]	K− MN[c]	K+ MN[d]	Proportion K+ MN[e]
0	0	1.6	1.3	0.3	19
107	16.8	3.6*	3.4*	0.2	6
134	31.4	6.6*	6.0*	0.6	9
168	42.1	6.2*	5.3*	0.9	14

[a] Cytotoxicity = $100 - 100 \times \{(CBPI_T - 1/CBPI_C - 1)\}$, with CBPI = (no. of mononucleated cells + 2 × no. of binucleated cells + 3 × no. of multinucleated cells)/total number of cells.
[b] Frequency of micronuclei per 100 cells.
[c] Frequency of kinetochore negative micronuclei per 100 cells indicating a chromosomal breakage.
[d] Frequency of kinetochore positive micronuclei per 100 cells indicating chromosomal loss.
[e] Percent kinetochore positive micronuclei among total micronuclei.
* Statistically significant (one-tailed Fisher's exact test $p < 0.05$).

in the rat during the *in vivo* assay were slightly (2–4-fold) higher than skin concentrations of Compound C from human biopsy specimens of 2 of 6 subjects following 12 h topical application to the intact skin of healthy volunteers.

Compound C was also tested for gene mutational activity in the *hprt* gene mutation test in CHO cells. The test in the absence of metabolic activation was conducted over a concentration range from 5 to 80 μg/mL, and the test conducted in the presence of metabolic activation spanned concentrations ranging from 75 to 250 μg/mL. Compound C did not induce a mutagenic response up to concentrations that produced significant cytotoxicity (data not shown).

There was no evidence that Compound C induced unscheduled DNA synthesis (UDS) in male rats at dose levels of 54.6 and 437 mg/kg at the 2 or 16 h time points, respectively (data not shown). Compound C was therefore evaluated as inactive in the *in vivo/in vitro* assay for unscheduled DNA synthesis in rat primary hepatocyte cultures at two time points when administered by continuous IV infusion up to the maximum feasible dose based on compound solubility. Mean plasma concentrations of Compound C measured in the *in vivo/in vitro* UDS assay after 2 and 16 h of continuous drug infusion were 0.99 and 4.29 μg/mL, respectively. Thus, slightly higher plasma levels were achieved than in the bone marrow micronucleus test, and therefore a slightly greater margin, when compared with human skin concentrations, was achieved.

Radiolabeled ^{14}C-Compound C was incubated with calf thymus DNA (1mg/mL) in triplicate at 75, 150, and 300 μg/mL for 4 h at 37 ± 2 °C (with and without metabolic activation). After the incubation, DNA was purified by treatment with RNase A, T1, and proteinase K followed by organic extraction and precipitation. Fifty microgram DNA from each sample was counted in a liquid scintillation counter. Results showed that ^{14}C-radioactivity counts obtained from the test article-treated DNA were not significantly higher than the background level counts obtained from untreated DNA. The positive control ^{14}C-benzo[a] pyrene-treated DNA with metabolic activation showed significantly higher ^{14}C-radioactivity, most likely resulting from binding of a BaP metabolite. It is concluded that, results obtained under these experimental conditions did not show evidence for tight association or binding of ^{14}C-Compound C.

Compound C was tested for the induction of micronuclei in rat skin keratinocytes (Table 7). Male and female rats were administered Compound C (2, 4, or 8 mg/animal/day) by topical application once a day for 4 consecutive days to intact skin clipped free of fur. The topical route of administration and dose levels used were selected to reflect the intended route of clinical administration and drug exposures in humans. The day following the last dose, keratinocytes were isolated from the excised skin, dropped onto glass slides, stained with acridine orange and evaluated for the presence of micronuclei. There were no significant increases in

Table 7
Induction of micronuclei in dermal keratinocytes following topical application with Compound C

Compound C (mg/animal/day)[a]	% Micronucleated keratinocytes	
	Males	Females
0[b]	0.03 ± 0.04	0.06 ± 0.05
2	0.04 ± 0.05	0.02 ± 0.03
4	0.05 ± 0	0.02 ± 0.03
8	0.05 ± 0.08	0.02 ± 0.03
CP[c]	4.19 ± 2.95	4.83 ± 2.13

[a] mg/20 cm^2 for 3 consecutive days.
[b] Ethanol, glycerol and water (70:5:25; v/v/v).
[c] 12 mg/animal/day cyclophosphamide for 4 consecutive days.

the numbers of MN keratinocytes with either sex. Concentrations of Compound C in the target tissue after 3 consecutive days of dosing at 8 mg/animal/day ranged from 72 to 175 μg/g of skin. This exceeded intended human skin concentrations by more than 20-fold.

The Working Group concluded that the initial response in the cytogenetics test and micronucleus assay when examined in isolation warranted further follow-up testing to assess the mechanism and relevance to humans exposed to the drug by dermal application. Follow-up testing yielded negative results in the *in vivo* micronucleus assay in rats, the rat liver *in vivo/in vitro* UDS test, and the *hprt* gene mutation test in CHO cells. This data set was considered by some members of the Working Group to be insufficient to address the concerns for Compound C. The reasons were as follows:

- The rat bone marrow micronucleus study using IV dosing produced systemic exposures that were only slightly above the human skin concentrations.
- The *in vivo/in vitro* UDS and *hprt* gene mutation test in CHO cells were considered to be of limited value, since they were both likely to be insensitive to compounds inducing DNA strand breakage, as clastogens would be expected to do.

The other additional studies were considered useful. The lack of DNA binding with calf thymus DNA was considered important additional information but the argument could have been strengthened by conducting the binding experiment in human lymphocyte cultures. Finally, the Working Group considered the dermal micronucleus assay as being the crucial study to address the concern for human safety since it addresses the relatively high concentration at the site of first contact and achieved safety multiples over intended human exposure.

Based on the overall weight of evidence, and in particular the DNA binding and dermal micronucleus results, the Working Group considered Compound C to be of low concern for human safety and that no further genetic toxicology testing is needed in addition to the presented package.

Appendix D. Case Study 4: Mode of action approach: a substance that does not directly react with DNA, such as a spindle poison

Compound D is intended for long-term treatment of severe autoimmune conditions, such as multiple sclerosis. In clinical trials, plasma C_{max} was 1.65 μg/mL and plasma AUC was 37.1 μg h/mL. The drug is extensively protein bound *in vivo* giving only 2.5% free drug in rats and 1.5% in humans. Approximately 60% is unbound in cell cultures.

There was no evidence of increased revertant numbers up to the limit dose level of 5000 μg/plate, when the Ames test was conducted on *Salmonella* strains TA98, TA100, TA1535, and TA1537 plus *E. coli* WP2 uvrA, using both plate incorporation and preincubation methods in the presence of S9, and plate incorporation in the absence of S9.

In the mouse lymphoma assay (MLA), cells were treated for 3 h in the absence and presence of S9 at concentrations up to 10 mM (3569 μg/mL). In the presence of S9 the treatments were very toxic, but there were no increases in mutant frequency at concentrations producing reductions in relative total growth (RTG) to <10%. In the absence of S9, at the limit concentration (10 mM, 3569 μg/mL), only 60% reduction in RTG was seen. The 24 h treatments (two separate experiments) in the absence of S9 had some concentrations where the mutant frequency increased above control by more than the global evaluation factor [39] of 126×10^{-6} for the microwell method. However, this was only at the 250 μg/mL concentrations in both experiments where relative survival (RS) was reduced to 9.88 and 8.97%, respectively (see Fig. 2a and b). Relative total growth (RTG), which is currently the recommended measure of toxicity in the MLA, did not reduce as much as RS, and was, respectively, 0.27 and 0.20 at this 250 μg/mL concentration in the two 24 h experiments in the absence of S9. The result could therefore be considered as borderline in that these responses would be excluded from consideration if judged by RS but not by RTG. At the time the study was done, both RS and RTG were considered acceptable for judgment of toxicity. No clear information on the type of damage caused at these borderline positive concentrations could be gleaned from colony sizing as frequencies of both large and small colony mutants increased. In light of the data obtained subsequently it might be argued that this unusual response in the MLA may reflect an aneugenic mode of action.

For the *in vivo* micronucleus (MN) test, rats were dosed orally on 2 consecutive days and bone marrow sampled 24 h after the second dose. Statistically significant increases in MN frequency to 3-fold concurrent control levels and slightly above the historical control range were found at the 180 and 360 mg/kg/day (Fig. 3). Although the response was relatively weak, the study was concluded as positive.

To investigate this *in vivo* finding, MN tests were conducted *in vitro* both in V79 cells and human lymphocytes.

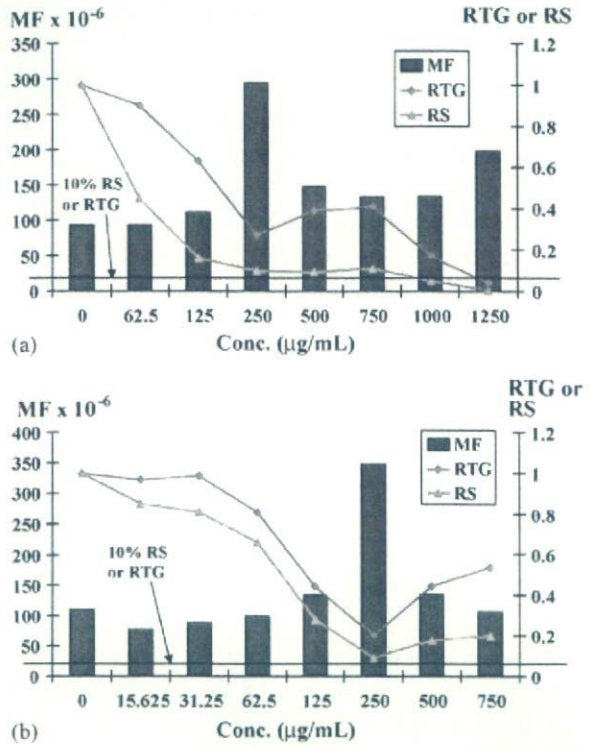

Fig. 2. Mouse lymphoma results with Compound D following 24 h treatments in the absence of S9 ((a) first independent experiment and (b) second independent experiment). MF = mutant frequency. RTG = relative total growth. RS = relative survival.

In both cases, substantial and statistically significant increases in MN frequency were seen after the 24 h treatments (but not after short treatments) in the absence of S9. The positive effects were seen at low levels of toxicity (Figs. 4 and 5). The MN in the human lymphocyte preparations were probed with a pan-centromeric

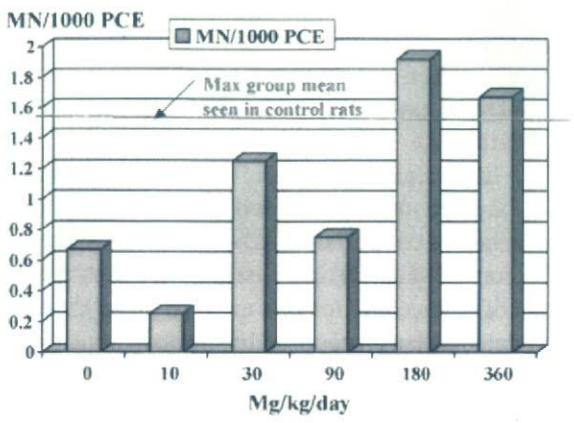

Fig. 3. Results of rat bone marrow micronucleus (MN) test with Compound D. Rats were dosed orally 2× daily and bone marrow sampled 24 h after the second dose. Data are accumulated from two separate experiments. Two thousand polychromatic erythrocytes (PCE) were scored per animal.

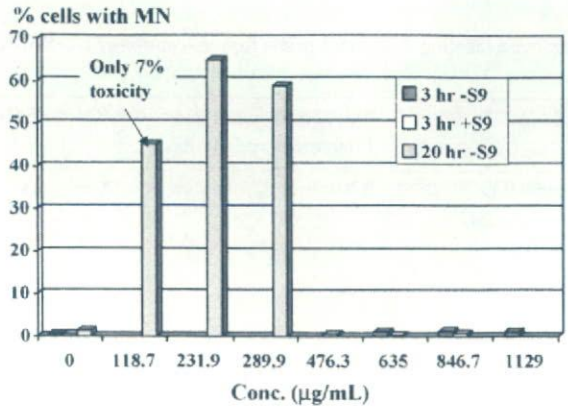

Fig. 4. Micronucleus (MN) data in V79 cells treated with Compound D.

DNA probe, using a fluorescent *in situ* hybridization method. 77% of the MN in the drug-treated cultures were centromere-positive, which is comparable to the 70% value observed with carbendazim tested as the positive control aneugen, and much higher than the 7% centromere-positive MN seen with the clastogenic positive control, cyclophosphamide (Table 8). There were too few MN from solvent controls to be probed for presence or absence of centromeres, but the historical control range for centromere-positive MN in human lymphocytes in the testing facility was 9–14.5%.

The pattern of results observed with this drug, (*i.e.*, borderline 24 h MLA positive responses in the absence of S9, the strong induction of micronuclei *in vitro* and high frequency of centromere-positive MN induced in human lymphocytes), was consistent with chromosome loss (aneuploidy). The result of the *in vivo* MN assay was very weak. At doses where no associated increase in MN was seen, the plasma exposures (C_{max} or AUC) were 75–125× the proposed human exposure levels.

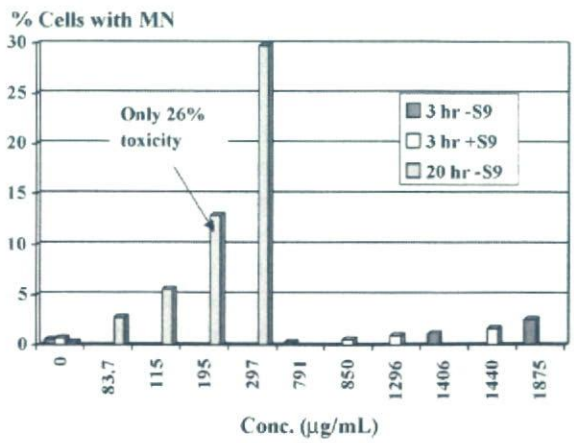

Fig. 5. Micronucleus (MN) data in human lymphocytes treated with Compound D.

Table 8
Centromere labeling with FISH probe for pan-centromeric DNA, for Compound D in micronucleus test in human lymphocytes

Treatment	No. of micronucleated binucleated cells probed	% C+	% C−
Compound D 194.5 μg/mL	100	77	23
Carbendazim 2.5 μg/mL	100	70	30
Cyclophosphamide 3.125 μg/mL	27	7	93

%C+: percent of cells with micronuclei containing a centromere. %C−: percent of cells with micronuclei containing no centromere. Pan-centromeric probing was not performed on concurrent solvent control slides. However, the historical frequency of C+ MN in human lymphocytes in the testing laboratory is 9–14.5%.

The Working Group agreed that, from the existing data, it was not clear whether the MN *in vivo* were solely due to chromosome loss. Even though 77% of the MN drug-treated cultures *in vitro* were centromere-positive, a secondary (*e.g.*, clastogenic) mechanism based on a direct DNA-mediated effect could be involved. It was suggested that it was necessary to demonstrate the absence of a direct DNA mechanism, and this could be done either *in vitro* or *in vivo*. Approaches that could be considered were:

- Demonstrate absence of DNA binding
- Demonstrate absence of induction of structural chromosomal aberrations
- Demonstrate absence of induction of DNA strand breaks.

The investigation of chromosomal aberration induction *in vivo* could be performed either in an acute study (scoring bone marrow) or in a repeat dose (*e.g.*, 28-day) study (scoring blood lymphocytes).

The Working Group discussed the evidence that is needed to convincingly describe a non-linear or threshold mechanism. It became clear that in addition to obtaining evidence in favour of a non-linear mechanism it might be necessary also to obtain evidence against a linear mechanism, *i.e.*, ruling out direct DNA reactivity. In this case study, evidence was obtained that was consistent with a chromosome loss (aneugenic) mode of action. However, this was not considered sufficient on its own. The mode of action should be more fully supported to be plausible. Further information on potential to affect the mitotic apparatus, or negative results in a rigorous test for chromosome breakage or for potential to damage DNA, would be very supportive. In this case the aneugenic mechanism of action would be accepted, no further genotoxicity testing would be required, and safety margins could be determined.

Appendix E. Use of historical control data in assessment of genotoxic response

Historical control data have been used as a convenient tool for explanation of the lack of relevance of a weak positive response, usually revealed by statistical significance when the response is compared with the concurrent negative control. For this purpose, the minimum and maximum values of the accumulated control data in the laboratory are usually used. The historical control data, however, should be more accurately and appropriately based on the distribution of the control data accumulated in the laboratory. To achieve this, certain experimental conditions, as summarized by Margolin and Risko [41] need to be achieved as follows:

- The experimental protocol must have remained fixed throughout the period covered by the historical data and the current experiment
- The method of scoring the response must be unchanged during the period
- The experimental unit must be comparable throughout the period
- The data must have been gathered by the same investigators within the same laboratory
- There must exist no known systematic differences between the various control groups, current and historical, that would produce systematic differences in response.

All control data should be included, even out-of-range data, unless there is a convincing rationale for exclusion. As long as the criteria above for obtaining historical control data have been met, these data may be combined over some different experimental conditions (*e.g.*, solvents, sampling times).

When historical control data fulfils the requirements above, they can provide a good guide for evaluation of experimental data. They can be compared with concurrent control data by using statistical methods. An example of a procedure for data analysis using historical control data has been proposed by Hayashi et al. [42].

References

[1] COM (Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment), COM guidance on a strategy for testing of chemicals for mutagenicity, United Kingdom, December 2000. (http://www.advisorybodies.doh.gov.uk/com/).

[2] K.L. Dearfield, M.C. Cimino, N.E. McCarroll, I. Mauer, L.R. Valcovic, Genotoxicity risk assessment: a proposed classification strategy, Mutat. Res. 521 (2002) 121–135.
[3] FDA, Toxicological Principles for the Safety of Food Ingredients. IV.C.1. Short-Term Tests for Genetic Toxicity, Redbook, 2000.
[4] FDA, Guidance for Industry Recommended Approaches to Integration of Genetic Toxicology Study Results. Draft Guidance. Genotoxicity. Center for Drug Evaluation and Research, November 2004.
[5] Health Protection Board (Canada), The Assessment of Mutagenicity. Health Protection Branch Mutagenicity Guidelines, Environ. Mol. Mutagen. 21 (1992) 15–37.
[6] ICH (International Cooperation on Harmonization), ICH Harmonized Tripartite Guideline. S2A. Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals, July, 1995. (http://www.ich.org).
[7] ICH (International Cooperation on Harmonization), ICH Harmonized Tripartite Guideline. S2B. Genotoxicity: A Standard Battery for Genotoxicity Testing of Pharmaceuticals, July, 1997. (http://www.ich.org).
[8] Council directive concerning the placing of plant protection products on the market (91/414/CEE), 15 July 2001.
[9] D.J. Kirkland, Genetic toxicology testing requirements: Official and unofficial views from Europe, Environ. Mol. Mutagen. 21 (1993) 8–14.
[10] D. Kirkland, M. Aardema, L. Henderson, L. Muller, Evaluation of the ability of a battery of three genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity, Mutat. Res. 584 (2005) 1–256.
[11] P.G.N. Kramers, A.G.A.C. Knaap, C.A. van der Heijden, M.R.D.F. Taalman, G.R. Mohn, Role of genotoxicity assays in the regulation of chemicals in The Netherlands: considerations and experiences, Mutagenesis 6 (1991) 487–493.
[12] OECD (Organisation for Economic Co-operation and Development), OECD Guidelines for Testing of Chemicals. Introduction to the OECD Guidelines on genetic toxicology testing and guidance on the selection and application of assays, OECD, Paris, France, 1986.
[13] T. Sofuni, Japanese guidelines for mutagenicity testing, Environ. Mol. Mutagen. 21 (1993) 2–7.
[14] E. Zeiger, Identification of rodent carcinogens and noncarcinogens using genetic toxicity tests: premises, promises, and performance, Regul. Toxicol. Pharmacol. 28 (1999) 85–95.
[15] TGD, Technical Guidance Document, second ed., European Chemicals Bureau, Joint Research Centre, Italy, December 2003 (http://ecb.jrc.it/new-chemicals/documents), 2003.
[16] 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. Schmidt, 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.
[17] Approaches to determining the mutagenic properties of chemicals: risk to future generations, Subcommittee On Environmental Mutagenesis, Committee to Coordinate Toxicology and Related Programs, U.S. Department of Health Education and Welfare (DHEW), Washington, D.C., 1977, pp. 1–58. Also published as: W.G. Flamm, L.R. Valcovic, F.J. deSerres, W. D'Aguanno, L. Fishbein, S. Green, H.V. Malling, V. Mayer, M. Prival, G. Wolff, E. Zeiger. Approaches to determining the mutagenic properties of chemicals: risk to future generations. J. Environ. Pathol. Toxicol. 1 (1977) 301–352.
[18] B.A. Bridges, ICPEMC Perspectives on the scientific basis of risk assessment in genetic toxicology, in: P. Oftedal, A. Brogger (Eds.), Risk and Reason: Risk Assessment in Relation to Environmental Mutagens and Carcinogens, Alan R. Liss, New York, 1986, pp. 175–182.
[19] D.J. Brusick, Methods for Genotoxic Risk Assessment, Lewis Publishers, CRC Press, Boca Raton, FL, 1994.
[20] J. Lewtas, D.M. DeMarini, J. Favor, D.W. Layton, J.T. MacGregor, J. Ashby, P.H.M. Lohman, R.H. Haynes, M.L. Mendelsohn, Risk characterization strategies for genotoxic environmental agents, in: D.J. Brusick (Ed.), Methods for Genotoxic Risk Assessment, Lewis Publishers, CRC Press, Boca Raton, FL, 1994, pp. 125–169.
[21] National Research Council (NRC), Risk assessment in the federal government, National Research Council, National Academy Press, Washington, D.C., 1983.
[22] J.T. Macgregor, D. Casciano, L. Müller, Strategies and testing methods for identifying mutagenic risks, Mutat. Res. 455 (2000) 3–21.
[23] U.S. Environmental Protection Agency (EPA), Guidelines for carcinogen risk assessment, EPA/630/P-03/001F, March 2005 (http://www.epa.gov/cancerguidelines).
[24] K.L. Dearfield, M.M. Moore, Use of genetic toxicology information for risk assessment, Environ. Mol. Mutagen. 46 (2005) 236–245.
[25] U.S. Environmental Protection Agency (EPA), Supplemental guidance for assessing cancer susceptibility from early-life exposure to carcinogens, EPA/630/R-03/003F, March 2005 (http://www.epa.gov/cancerguidelines).
[26] D. Jacobson-Kram, A. Jacobs, Use of genotoxicity data to support clinical trial or positive genetox findings on a candidate pharmaceutical or impurity ... Now what? Int. J. Toxicol. 24 (2005) 129–134.
[27] FDA (Food and Drug Administration) CDER (Center for Drug Evaluation and Research), Guidance for industry and review staff: recommended approaches to integration of genetic toxicology study results, January 2006.
[28] U.S. Environmental Protection Agency (EPA), Environmental Protection Agency. Guidelines for mutagenicity risk assessment, Washington, DC, Fed Register 51 (1986) 34006–34012.
[29] L. Rhomberg, V. Dellarco, C. Siegel-Scott, K. Dearfield, D. Jacobson-Kram, A quantitative estimation of the genetic risk associated with the induction of heritable translocations at low-dose exposure: ethylene oxide as an example, Environ. Mol. Mutagen. 16 (1990) 104–125.
[30] K.L. Dearfield, G.R. Douglas, U.H. Ehling, M.M. Moore, G.A. Sega, D.J. Brusick, Acrylamide: a review of its genotoxicity and an assessment of heritable genetic risk, Mutat. Res. 330 (1995) 71–99.
[31] M.C. Cimino, Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes, Environ. Mol. Mutagen. 47 (2006) 362–390.
[32] R.D. Snyder, J.W. Green, A review of the genotoxicity of marketed pharmaceuticals, Mutat. Res. 488 (2001) 151–169.
[33] D.J. Kirkland, L. Müller, Interpretation of the biological relevance of genotoxicity test results: the importance of thresholds, Mutat. Res. 464 (2000) 137–147.
[34] L. Müller, P. Kasper, Human biological relevance and the use of threshold-arguments in regulatory genotoxicity assessment: experience with pharmaceuticals, Mutat. Res. 464 (2000) 19–34.

[35] US FDA, Center for Drug Evaluation and Research (CDER), Guidance for Industry: Safety Testing of Drug Metabolites. Draft, June 2005.

[36] E.R. Nestmann, D.W. Bryant, C.J. Carr, Toxicological significance of DNA adducts: summary of discussions with an expert panel, Regul. Toxicol. Pharmacol. 24 (1996) 9–18.

[37] S. Madle, W. Von der Hude, L. Broschinski, G.R. Jänig, Threshold effects in genetic toxicology: perspective of chemicals regulation in Germany, Mutat. Res. 464 (2000) 117–121.

[38] M. Kirsch-Volders, A. Vanhauwaert, U. Eichenlaub-Ritter, I. Decordier, Indirect mechanism of genotoxicity, Toxicol. Lett. 140/141 (2003) 63–74.

[39] 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: mouse lymphoma thymidine kinase gene mutation assay: follow-up meeting of the international workshop on genotoxicity testing – Aberdeen, Scotland, 2003 – assay acceptance criteria, positive controls, and data evaluation, Environ. Mol. Mutagen. 47 (2006) 1–5.

[40] N.E. McCarroll, A. Protzel, Y. Ioannou, H.F. Stack, M.A. Jackson, M.D. Waters, K.L. Dearfield, A survey of EPA/OPP and open literature on selected pesticide chemicals. III. Mutagenicity and carcinogenicity of benomyl and carbendazim, Mutat. Res. 512 (2002) 1–35.

[41] B.H. Margolin, K.J. Risko, The statistical analysis of *in vivo* genotoxicity data: case studies of the rat hepatocyte UDS and mouse bone marrow micronucleus assay, in: J. Ashby, et al. (Eds.), Evaluation of Short-Term Tests for Carcinogens. Report of the International Program on Chemical Safety's Collaborative Study on *In vivo* Assays, vol. 1, Cambridge University Press, Cambridge, pp. 129–142.

[42] M. Hayashi, I. Yoshimura, T. Sofuni, M. Ishidate Jr., A procedure for data analysis of the rodent micronucleus test involving a historical control, Environ. Mol. Mutagen. 13 (1989) 347–356.

[43] D. Brusick, J. Ashby, E. de Serres, P.H.M. Lohman, T. Matsushima, B.E. Matter, M.L. Mendelsohn, M.D. Waters, Weight-of-evidence scheme for evaluation and interpretation of short-term results, in: C. Ramel, B. Lambert, J. Magnusson (Eds.), Genetic Toxicology of Environmental Chemicals, Part B: Genetic Effects and Applied Mutagenesis, Alan R. Liss, New York, 1986, pp. 121–129.

[44] D.J. Brusick, J. Ashby, E.J. de Serres, P.H.M. Lohman, T. Matsushima, B.E. Matter, M.L. Mendelsohn, D.H. Moore, S. Nesnow, M.D. Waters, A method for combining and comparing short-term genotoxicity test data, Mutat. Res. 266 (1992) 1–60.

[45] J.M. Parry, A. Sarrif (Eds.), Dose-response and threshold-mediated mechanisms in mutagenesis, Mutat. Res. Spec. Issue 465 (1999) 1–160.

Mutagenesis vol. 21 no. 1 pp. 11–13, 2006
Advance Access Publication 27 September 2005
doi:10.1093/mutage/gei053

Performance of flow cytometric analysis for the micronucleus assay—a reconstruction model using serial dilutions of malaria-infected cells with normal mouse peripheral blood

Dorothea Torous*, Norihide Asano[1], Carol Tometsko, Siva Sugunan, Stephen Dertinger, Takeshi Morita[2] and Makoto Hayashi[3]

Litron Laboratories, 200 Canal View Boulevard, Rochester, NY 14623, USA, [1]Toxicological Research Center, Nitto Denko Corporation, 1-1-2 Shimohozumi, Ibaraki, Osaka 567-8680, Japan, [2]Division of Safety Information on Drug, Food and Chemicals and [3]Division of Genetics and Mutagenesis, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

To confirm the performance and statistical power of a flow cytometric method for scoring micronucleated erythrocytes, reconstruction experiments were performed. For these investigations, peripheral blood erythrocytes from untreated mice, with a micronucleated erythrocyte frequency of ~0.1% were combined with known quantities of *Plasmodium berghei* (malaria) infected mouse erythrocytes. These cells had an infected erythrocyte frequency of ~0.7%, and mimic the DNA content of micronuclei (MN). For an initial experiment, samples with a range of MN/malaria (Mal) content were constructed and analysed in triplicate by flow cytometry until 2000, 20 000 and 200 000 total erythrocytes were acquired. In a second experiment, each specimen was analysed in triplicate until 2000, 20 000, 200 000 and 1 000 000 erythrocytes were acquired. As expected, the sensitivity of the assay to detect small changes in rare erythrocyte sub-population frequencies was directly related to the number of cells analysed. For example, when 2000 cells were scored, increases in MN/Mal frequencies of 3.9- or 2.7-fold were detected as statistically significant. When 200 000 cells were analysed, a 1.2-fold increase was detected. These data have implications for the experimental design and interpretation of micronucleus assays that are based on automated scoring procedures, since previously unattainable numbers of cells can now be readily scored.

Introduction

From a statistical point of view, in order to achieve a higher power of detection, sample size should be increased. For many experimental situations, it is not always feasible to increase the number of subjects studied. When the event under consideration is rare as to cause appreciable scoring error, then an alternative would be to enhance the precision of each measurement. For example, in the rodent erythrocyte micronucleus assay, the evaluation of 2000 immature erythrocytes per animal and 5 animals per dose group represents commonly cited minimum values. Owing to the rarity of micronucleated cells, even this minimal assay design results in tedious and time-consuming efforts. The use of flow cytometry (1–3) realizes the ability to evaluate high numbers of erythrocytes, something that is impossible to achieve by manual microscopy. By reducing scoring error in this manner, flow cytometry has the potential to increase statistical power.

In the present study, we evaluated the relationship between statistical power to detect a rare erythrocyte sub-population, i.e. micronucleated or malaria-infected erythrocytes (MN/Mal), and the total number of erythrocytes analysed. These experiments were accomplished using a reconstruction model whereby known quantities of malaria-infected erythrocytes were added to blood from an untreated mouse. Malaria is a known model for micronucleated erythrocytes, as they endow the target cells of interest with a micronucleus-like DNA content (4,5). The samples were analysed by flow cytometry to measure the MN/Mal frequency through the interrogation of 2000 (2k), 20 000 (20k), 200 000 (200k) and 1 000 000 (1m) erythrocytes. The results presented here show the capability of flow cytometric technology to reduce scoring error, and also the extent to which this affects the ability to detect small changes to baseline micronucleus frequencies.

Materials and methods

Staining of blood specimens

Methanol-fixed blood from untreated and malaria-infected mice used in this study were two 'biological standards' which accompany the Mouse MicroFlow®PLUS kits (Litron Laboratories, NY). MicroFlow PLUS kits were the source of these specimens.

Before analysis, malaria-infected specimens and untreated mouse specimens were washed out of fixative with ~12 ml Hank's Balanced Salt Solution. Procedures for the 3-colour labelling technique which appear in the MicroFlow®PLUS instruction manual (version 031230) were scaled up ~7-fold in order to provide at least 10 ml each of control and malaria blood in a cell density range that is recommended for this assay (between ~2000 and 6000 events/s). Anti-CD71-FITC, anti-CD61-PE and all other flow cytometry reagents were also supplied in the kits. After the labelling procedures were accomplished, the cell density of the malaria-infected sample was adjusted so that it was equal to that of the control blood sample. Initial cell densities were measured with a Coulter Counter, model ZM. After adjustment with additional propidium iodide staining solution, equal cell densities were confirmed by Coulter Counter measurements. Normalization of cell densities was an important experimental design consideration, as this allowed us to calculate the expected MN/Mal frequencies in the diluted samples once the frequencies of the original control (0.10 and 0.09% for Experiments 1 and 2, respectively) and malaria-infected (0.67 and 0.70% for Experiments 1 and 2, respectively) samples were determined with high precision (i.e. control and malaria-infected %MN/Mal frequencies are the mean value of triplicate analyses with 1m erythrocytes per analysis).

Dilution of malaria blood specimen

Malaria-infected blood (Sample H) was diluted with control blood (Sample A) in the following ratios (v/v): 1:1 (Sample G), 1:3 (Sample F), 1:7 (Sample E), 1:15 (Sample D), 1:31 (Sample C) and 1:63 (Sample B). These blood specimens were stored at 4°C until flow cytometric analysis, which occurred on the same day. Each sample was analysed three times to evaluate reproducibility.

Flow cytometric analysis

All samples were analysed according to the MicroFlow® PLUS 3-colour technique. One deviation to the kit-supplied data acquisition and analysis template was that the frequency of erythrocytes with malaria or micronuclei was determined without restriction to CD71-expression level. That is, the

*To whom correspondence should be addressed. Email: dtorous@litronlabs.com

Mal and MN frequencies measured and reported here are based on total peripheral blood erythrocytes. A second deviation from standard practices is that the default stop mode of 20 000 reticulocytes was not utilized. Rather, each specimen was analysed until the following number of erythrocytes were acquired: 2k, 20k and 200k erythrocytes in the first experiment and 2k, 20k, 200k and 1m erythrocytes in the second experiment.

Statistical analysis

The average of triplicate MN/Mal measurements associated with the control blood sample were compared with those associated with each of the other seven specimens by the Fisher's exact method. A *P*-value of 0.05 divided by 7 (number of sample groups) was considered evidence of a statistically significant difference. Expected versus observed MN/Mal frequencies were graphed for each measurement performed in the second experiment. Microsoft Excel (Microsoft Corp., Seattle, Washington) was used to determine a best-fit line. The associated equations and r^2 values were determined.

Results

Data from Experiments 1 and 2 are summarized in Table I and include the expected and observed MN/Mal frequencies. The MN/Mal frequencies shown are the average of triplicate analyses. As shown in Table I, for measurements based on 2k erythrocytes, samples with expected MN/Mal frequencies of 0.39 and 0.24% were found to be significantly different from control samples, in Experiments 1 and 2, respectively. These values correspond to fold increases of 3.9 and 2.7 for the first and second experiment, respectively. As more erythrocytes were analysed per sample, the detection limit was improved. For instance, measurements based on the evaluation of 200k erythrocytes per analysis show statistical significance for expected MN/Mal samples of 0.12 and 0.11%. These values correspond to an increase of ~1.2-fold. In fact for the second experiment, when a stop mode of 1m erythrocytes was investigated, statistical significance was observed between the control blood sample (0.09% MN/Mal) and the specimen with the lowest frequency of malaria (0.10% MN/Mal; $P = 0.00005$).

As an aid for visualizing the performance characteristics associated with the various number of cells analysed, scattergrams showing %MN/Mal measurement are presented (Fig. 1). Best-fit lines and equations are included with these graphs, and illustrate the degree to which the experimentally derived data agree with the linear relationship that is known to exist among MN/Mal frequencies for these specimens.

Discussion

To evaluate the performance and statistical power of a flow cytometric approach to score micronucleated erythrocytes, we performed a reconstruction model experiment by the serial dilution of malaria-infected mouse blood with normal mouse

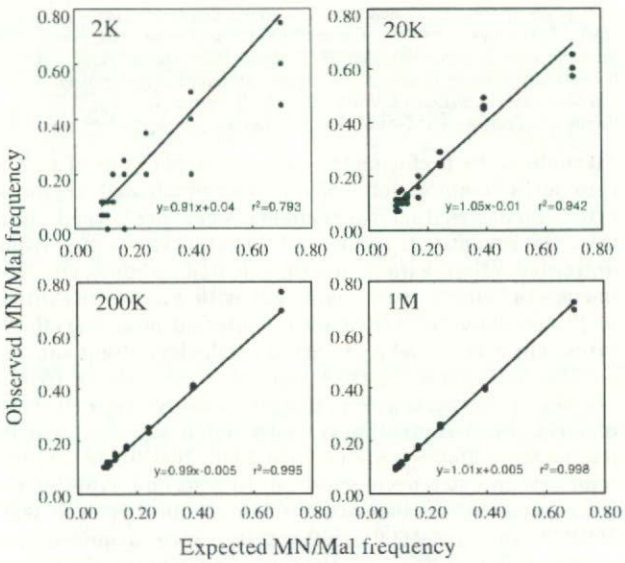

Fig. 1. Scattergram of expected versus observed MN/Mal frequencies. Each of three replicate measurements is plotted for these specimens. Best-fit linear lines are graphed, with associated equations. r^2 values document the degree of reproducibility observed.

Table I. MN/Mal frequencies (%) and *P*-values for comparisons with sample A

Sample	Expected (%)	Number of cells analysed/sample							
		2k		20k		200k		1m	
		(%)	*P*-value	(%)	*P*-value	(%)	*P*-value	(%)	*P*-value
Experiment 1									
A	0.10	0.07		0.09		0.10			
B	0.11	0.08	*0.50000*	0.11	*0.20764*	0.11	*0.13403*		
C	0.12	0.05	*0.77349*	0.10	*0.46272*	0.12	*0.00007*		
D	0.14	0.08	*0.50000*	0.13	*0.03463*	0.14	*0.00000*		
E	0.18	0.22	*0.02452*	0.21	*0.00000*	0.18	*0.00000*		
F	0.25	0.18	*0.05924*	0.30	*0.00000*	0.25	*0.00000*		
G	0.39	0.57	*0.00000*	0.36	*0.00000*	0.41	*0.00000*		
H	0.67	0.57	*0.00000*	0.73	*0.00000*	0.67	*0.00000*		
Experiment 2									
A	0.09	0.08		0.09		0.10		0.09	
B	0.10	0.08	*0.62305*	0.11	*0.20374*	0.11	*0.04822*	0.10	*0.00005*
C	0.11	0.05	*0.85547*	0.12	*0.06479*	0.12	*0.00301*	0.11	*0.00000*
D	0.13	0.13	*0.29053*	0.12	*0.08887*	0.14	*0.00000*	0.13	*0.00000*
E	0.17	0.15	*0.21198*	0.16	*0.00038*	0.17	*0.00000*	0.17	*0.00000*
F	0.24	0.30	*0.00531*	0.26	*0.00000*	0.24	*0.00000*	0.25	*0.00000*
G	0.40	0.37	*0.00076*	0.47	*0.00000*	0.41	*0.00000*	0.40	*0.00000*
H	0.70	0.60	*0.00000*	0.61	*0.00000*	0.71	*0.00000*	0.70	*0.00000*

Shading indicates those samples that are significantly different from respective control samples.

blood. As expected, the present results illustrate that the power of rare event detection is directly related to the number of cells analysed per specimen. By analysing 3m (triplicate of 1m) cells per group, 0.10% is significantly different ($P = 0.00005$) when compared with 0.09%. Even so, it must be appreciated that the biological significance of minute changes must be considered in addition to statistical significance.

Previously, we have shown that individual differences were negligible in the mouse micronucleus assay when 1000 cells per animal and 5 or 6 mice per dose group were analysed (6–8) and the statistical unit for the evaluation can be assigned to a cell but not to an animal. According to the present results and also results by Asano *et al.* (9), the variability of the data was high when 2k cells were analysed. Under these circumstances, the difference among animals is not apparent, as they are likely to be smaller than the scoring error. While, in the case of the present malaria dilution experiments, when 200k or 1m cells per sample were analysed, the scoring error decreased and converged to a value. This, however, is not true in the case of the actual micronucleus assays using model chemicals (9). When 200k or 1m immature erythrocytes were analysed, differences between individual animals became apparent and there was data variability within each dose group. Therefore, even if the experimental size in the animal experiments is increased, we cannot expect the same increment of detecting power. This finding suggests that optimizing the statistical procedure also includes evaluating individual differences.

Based on the present results, we confirm the accuracy and high performance of the micronucleus assay system using flow cytometry and we propose that the number of reticulocytes analysed for the micronucleus assay using flow cytometry be a minimum of 20k. We suggest that the analysis of 20k reticulocytes is approximately equivalent to the manual microscopic analysis according to test guideline OECD 474 (9,10). We anticipate that the experimental size of the MN assay will be recommended and set by expert committees based on the evaluated data. In addition to statistical sensitivity, biological variability between animals and as a consequence of treatment should also be considered. There appears to be diminishing value to analyses based on 200k or even 1m per animal. These may be useful in certain special circumstances, for instance when looking for evidence of threshold or practical threshold effects (9).

References

1. Grawé,J., Zetterberg,G. and Arnéus,H. (1992) Flow-cytometric enumeration of micronucleated polychromatic erythrocytes in mouse peripheral blood. *Cytometry*, **13**, 750–758.
2. Dertinger,S.D., Torous,D.K. and Tometsko,K.R. (1996) Simple and reliable enumeration of micronucleated reticulocytes with a single-laser flow cytometer. *Mutat. Res.*, **371**, 283–292.
3. Torous,D.K., Hall,N.E., Illi-Love,A.H. *et al.* (2005) Interlaboratory validation of a CD71-based flow cytometric method (Microflow) for the scoring of micronucleated reticulocytes in mouse peripheral blood. *Environ. Mol. Mutagen.*, **45**, 44–55.
4. Tometsko,A.M., Torous,D.K. and Dertinger,S.D. (1993) Analysis of micronucleated cells by flow cytometry. 1. Achieving high resolution with a malaria model. *Mutat. Res.*, **292**, 129–135.
5. Dertinger,S.D., Torous,D.K., Hall,N.E., Tometsko,C.R. and Gasiewicz,T.A. (2000) Malaria-infected erythrocytes serve as biological standards to ensure reliable and consistent scoring of micronucleated erythrocytes by flow cytometry. *Mutat. Res.*, **464**, 195–200.
6. Hayashi,M., Yoshimura,I. Sofuni,T. and Ishidate,M. (1989) A procedure for data analysis of the rodent micronucleus test involving a historical control. *Environ. Mol. Mutagen.*, **13**, 347–356.
7. Hayashi,M., Hashimoto,S., Sakamoto,Y., Hamada,C., Sofuni,T. and Yoshimura,I. (1994) Statistical analysis of data in mutagenicity assays: rodent micronucleus assay. *Environ. Health Perspect.*, **102** (Suppl. 1), 49–52.
8. Adler,I.-D., Bootman,J., Favor,J., Hook,G., Schriever-Schwemmer,G., Welzl,G., Whorton,E., Yoshimura,I. and Hayashi,M. (1998) Recommendations for statistical designs of in vivo mutagenicity test with regard to subsequent statistical analysis. *Mutat. Res.*, **417**, 19–30.
9. Asano,N., Torous,D.K., Tometsko,C.R., Dertinger,S.D., Morita,T. and Hayashi,M. (2005) Practical threshold for micronucleated reticulocyte induction observed for low doses of mitomycin C, Ara-C, and colchicine. *Mutagenesis*, **21**, 15–20.
10. OECD (1997) *Guideline for the Testing of Chemicals: Mammalian Erythrocyte Micronucleus Test. Guideline 474.* Organisation for Economic Cooperation and Development, Paris, France.

Received on May 23, 2005; revised on July 21, 2005; accepted on July 22, 2005

Bull.Natl.Inst.Health Sci., **125**, 35-50 (2007) **Review**

有機スズ化合物の生殖発生毒性

江馬 眞

Reproductive and Developmental Toxicity of Organotin Compounds

Makoto Ema

Organotin compounds are chemicals widely used in agriculture and industry. Widespread use of organotins has caused increasing amounts to be released into the environment. Organotins show many aspects of toxicity, such as immunotoxicity, neurotoxicity, and reproductive/developmental toxicity. However, the reproductive and developmental toxicity of organotins is not well understood. The findings of the studies on reproductive and developmental effects of organotin compounds in mammals were summarized in this review.

Keywords: Organotin, reproductive toxicity, developmental toxicity, implantation failure, teratogenicity

1. はじめに

有機スズ化合物は農業や工業の分野で広く使われている[1, 2]．四価のスズ化合物は主に他の有機スズ化合物生産の中間体として使用されている．三価の有機スズ化合物は殺生物作用を有しており，防黴剤，ダニ駆虫剤，ネズミ駆散剤，軟体動物駆除剤等として，また，船底防汚剤として広く用いられている．特に，トリフェニルスズ（TPT）とトリブチルスズ（TBT）は藻類駆除剤，軟体動物駆除剤として，防汚剤製品中によく使われてきた．二価の有機スズ化合物は商業上で最も重要な誘導体であり，主にプラスティック工業の分野でポリマーの劣化を防止するためにポリ塩化ビニル（PVC）プラスティックの熱，光安定剤として使われている．一価の有機スズ化合物はPVCの安定剤として使用されている．有機スズ化合物の生産量をTable 1に示した。

近年の有機スズ化合物の広範な使用により有機スズ化合物による環境汚染の懸念が高まっている．農薬としての使用以外の有機スズ化合物の環境汚染の経路は，PVCプラスティックの安定剤として使われた有機スズ化合物の水中への溶出であり[3]，また，船底防汚剤としての使用が水環境汚染の原因となっている[4]．海棲生物[5-7]や海産物[8-12]からTBTやTPTが検出されており，カキ[13]，泥ガニ[14]，ムールガイ[15]，チヌークサーモン[16]，イルカ，マグロ及びサメ[17]における食物連鎖によるTBTの生物濃

To whom correspondence should be addressed:
Makoto Ema; Kamiyoga 1-18-1, Setagaya, Tokyo 158-8501, Japan; Phone: +81-3700-9878; Fax: +81-3-9700-1408; E-mail: ema@nihs.go.jp

Table 1. スズ化合物の生産量

物 質 名	CAS	生産量（トン）
Dibutyltin dichloride	683-18-1	10,000 - 15,000
Dibutyltin dilaurate	77-58-7	1000 - 5000
Dibutyltin maleate	78-04-6	500 - 1000
Dibutyltin oxide	818-08-6	1000 - 5000
Dibutyltin bis (2-ethylhexylmercap-acetate)	10584-98-2	7,500 - 12,500
Dibutyltin bis (isooctyl mercap-acetate)	25168-24-5	Not available
Dimethyltin dichloride	753-73-1	1,000 - 5,000
Dimethyltin bis (2-ethylhexyl mercap-acetate)	57583-35-4	5,000 - 10,000
Dimethyltin bis (isooctyl mercap-acetate	26636-01-1	Not available
Dioctyltin dichloride	3542-36-7	5,000 - 10,000
Dioctyltin bis (2-ethylhexyl mercap-acetate)	15571-58-1	7,500 - 12,500
Dioctyltin bis (isooctyl mercap-acetate)	26401-86-5	Not available
Monobutyltin trichloride	1118-46-3	10,000-15,000
Monobutyltin tris (2-ethylhexyl mercap-acetate)	26864-37-9	2,500-7,500
Monobutyltin tris (isooctyl mercap-acetate)	25852-70-4	Not available
Monomethyltin trichloride	993-16-8	1,000 - 5,000
Methyltin Reverse Ester Tallate	201687-57-2	7,500 - 10,000
Monomethyltin tris (2-ethylhexylmercap-acetate)	57583-34-3	5,000 - 10,000
Monomethyltin tris (isooctylmercap-acetate)	54849-38-6	Not available
Mono-octyltin trichloride	3091-25-6	1,000 - 5,000
Mono-octyltin tris (2-ethylhexylmercap-acetate)	27107-89-7	2,500 - 7,500
Mono-octyltin tris (isooctylmercap-acetate)	26401-86-5	Not available
Tributyltin chloride	1461-22-9	2500 - 3000
Tetrabutyltin	1461-25-2	10,000 - 12,500
Tetraoctyltin	3590-84-9	2,500 - 7,500
Tin Tetrachloride	7646-78-8	20,000 - 25,000

出典：ORTEP Association. 2004. Global production data

縮，またコイ[10]及びカブトガニ[18]における食物連鎖によるTPTの生物濃縮が報告されている．ヒトは海産物を通じて有機スズを摂取しており，滋賀県人のTBTの1日摂取量は4.7-6.9 μg（1991年），2.2-6.7 μg（1992年），TPTの1日摂取量は4.7-6.9 μg（1991年），2.2-6.7 μg（1992年）であり[19]，また，1998年のトータルダイエット方式調査による日本人の1日摂取量は，TPT: 0.09 μg，ジフェニルスズ（DPT）: 0 μg，TBT: 1.7 μg，ジブチルスズ（DBT）: 0.45 μgと報告されている[19]．これらの値はFAO/WHO合同残留農薬専門家会議によるTPTの許容1日摂取量（25 μg）[20]及びtributyltin oxide（TBTO）の経口曝露指針値（18 μg）[21]よりも低く，海産物中の有機スズ濃度はヒトの健康に悪影響を及ぼすほど高くはない[11, 19]と考えられるが，Belfoidら（2000）[9]は，ヒトの健康影響の可能性について結論を下すためには海産物中のTBT含量についての更なる研究が必要であると述べている．

近年，環境汚染物質による内分泌系の障害の結果，野生動物の生殖に対する悪影響が惹起される可能性が指摘されている[22]．TBT及びTPTは内分泌撹乱作用が疑われる物質とされており[23]，低濃度で巻貝のインポセックス（imposex: 雌にペニスと輸精管が形成される現象），さらに繁殖障害を引き起こす[24]ことから，哺乳類における生殖発生毒性が懸念されている．

有機スズ化合物の一般毒性については古くから比較的よく知られている[2, 21, 25-28]が，生殖発生毒性の理解は十分ではない．本稿では，Ema M and Hirose A（2006）[29] Reproductive and developmental toxicity of organotin compounds. In Metals, Fertility, and Reproductive Toxicity, CRC Pressの実験動物における生殖発生毒性の項を基に最近の新たな知見を加えて，有機スズ化合物の生殖発生毒性について概説した．

2. フェニルスズ化合物の生殖毒性

2-1 トリフェニルスズ（TPT）の生殖毒性

TPTは昆虫の不妊化剤として知られている[30]．Table 2にTPTの生殖毒性試験の結果を示した．雄に対する影響として，100または200 ppmのtriphenyltin hydroxide（TPTH）を含む飼料を64日間与えた雄Sharmanラットを無処置雌ラットと繰り返し5回交配させたところ，体重増加及び摂餌量の著しい低下とともに，受精率，出産生児数及び交配あたりの生児数の低下が認められたが，摂餌量の回復とともに受精率が回復したことが報告されている[31]．Holtzmanラットに20 mg/kgのtriphenyltin acetate（TPTA）またはtriphenyltin chloride（TPTCl）を19日間混餌投与したとき，体重及び精巣重量への影響が顕著であった．精巣では，精細管の精上皮細胞層の減少，ステージの進行した精上皮細胞の減少及び精細管腔の狭小化等の精巣の退行性変化がみられ，TPTAを投与したときにより強い毒性が観察されている[32]．同様に，20 mg/kgのTPTAまたはTPTClのHoltzmanラットへの20日間混餌投与により精子形成が障害されたが，70日間正常飼料を与えると精子形成の完全な回復がみられた[33]．ICR/Ha SwissマウスにTPTA（2.4,12 mg/kg）またはTPTH（1.3,8.5 mg/kg）を単回腹腔内投与，もしくはTPTA（6mg/kg）またはTPTH（11 mg/kg）を5日間連続強制経口投与した後に，無処置雌と交配させ，妊娠13日に剖検した結果，優性致死作用は認められなかった[34]．

TPTの雌動物における生殖毒性についても報告がある．20 mg/kgのTPTAまたはTPTClのHoltzmanラットへの4日間混餌投与により，成熟卵胞の減少，初期卵胞の閉鎖の増加，黄体の著しい減少が観察されている[35]．このような現象は排卵の減少，延いては受胎率の低下の原因となる．

ラットの妊娠初期にTPTClを投与したときの妊娠の成立及び維持に対する影響が検討されている[36]．Wistarラットの妊娠0-3日に3.1, 4.7, 6.3 mg/kgまたは妊娠4-6日に6.3, 12.5, 25.0 mg/kg のTPTClを強制経口投与したところ，用量依存的な着床阻害が引き起こされ，妊娠0-3日の4.7 mg/kg以上，妊娠4-6日の12.5 mg/kg以上で妊娠率の低下が観察された．着床前胚死亡率の増加は妊娠0-3日の4.7 mg/kg以上でみられたが，TPTCl投与群の妊娠が成立した雌における着床数，生存胎児数，着床前及び着床後の胚死亡率は対照群と同様であった．これらの結果は妊娠初期に投与したTPTClは着床阻害作用を示し，着床前に投与した方が強い影響を及ぼすことを示している．

子宮内膜の正常な機能は胚生存のために重要であり，子宮の脱落膜化は正常な着床及び胎盤形成，その後の正常妊娠の維持に必須である．偽妊娠動物における内膜創傷による子宮内膜の変化は，胚の着床によって惹起される妊娠子宮における脱落膜反応と同様であり[37, 38]，着床に関連した母体の生理学的変化のモデルとなりうる[37]．この方法を用いて脱落膜反応を誘起することにより化学物質の生殖発生毒性を母体と胚/胎児とに分けて検討することが可能となる[37-41]．TPTClの着床阻害作用の原因を明らかにするために，子宮機能に対する作用が偽妊娠Wistarラットを用いて検討されている[42]．ラットの偽妊娠0-3日にTPTCl（3.1, 4.7, 6.3 mg/kg）を強制経口投与し，偽妊娠4日の11:00から13:00の間に麻酔下で偽妊娠ラットの子宮内膜を創傷することにより脱落膜反応を誘起し，偽妊娠9日の子宮重量を子宮脱落膜化の指標として測定した[43]．その結果，子宮重量の低下（子宮脱落膜化の抑制），偽妊娠4及び9日の血清中プロゲステロン低下が4.7 mg/kg以上の投与でみられた．この投与量は妊娠0-3日に投与したときには着床前胚致死を引き起こす

Table 2 フェニルスズ化合物による生殖毒性

物質名	動物種	投与量	投与日	投与経路	生殖発生毒性	著者
TPTH	Sharman ラット	100-200 ppm	64-238 日	経口（混餌）	↓生存児のある母体数 ↓同腹児数/交配	Gains & Kimbrough (1968)
TPTA TPTCl	Holtzman ラット	20 mg/kg	19 日	経口（混餌）	↓精巣サイズ ↑精巣の形態学的変化	Pate & Hays (1968)
TPTA TPTCl	Holtzman ラット	20 mg/kg	20 日	経口（混餌）	↑精子形成過程の障害	Snow & Hays (1983)
TPTA	ICR/Ha Swiss マウス	2.4-12 mg/kg 6 mg/kg	1 日 5 日	腹腔内 強制経口	優性致死作用なし 優性致死作用なし	Epstein et al. (1972)
TPTH		1.3-8.5 mg/kg 11 mg/kg	1 日 5 日	腹腔内 強制経口	優性致死作用なし 優性致死作用なし	
TPTA TPTCl	Holtzman ラット	20 mg/kg	4-24 日	経口（混餌）	↓成熟卵胞数 ↑初期卵胞の閉鎖数 ↓黄体数	Newton & Hays (1968)
TPTCl	Wistar ラット	4.7-6.3 mg/kg 12.5-25 mg/kg	妊娠 0-3 日 妊娠 4-6 日	強制経口 強制経口	↓妊娠率，↓胎児体重 ↓妊娠率	Ema et al. (1997a)
TPTCl	Wistar ラット	4.7-6.3 mg/kg	偽妊娠 0-3 日	強制経口	↓子宮内膜脱落膜化 ↓血清プロゲステロン	Ema et al. (1999a)
DPTCl	Wistar ラット	16.5-24.8 mg/kg	妊娠 0-3 日	強制経口	↓妊娠率，着床前胚死亡 ↓胎児体重	Ema et al. (1999b)
		33.3 mg/kg	妊娠 4-7 日	強制経口	同上，↑着床後胚死亡	
DPTCl	Wistar ラット	4.1-24.8 mg/kg	偽妊娠 0-3 日	強制経口	↓子宮内膜脱落膜化 ↓血清プロゲステロン	Ema & Miyawaki (2002)

TPTH: Triphenyltin hydroxide, TPTA: Triphenyltin acetate, TPTCl: Triphenyltin chloride, DPTCl: Diphenyltin dichloride.

量であった[36]．これらの結果は，TPTClはプロゲステロン低下を伴う子宮内膜の脱落膜化抑制を惹起し，これらがTPTClによる着床阻害に関与していることを示唆している．TPTClの子宮の脱落膜化抑制及び着床阻害作用に対する卵巣ホルモンの作用を検討したところ，プロゲステロンとエストロンの投与はTPTClを投与した卵巣摘出ラットの脱落膜化を維持し，4.7 mg/kg以上のTPTClとプロゲステロンを併用投与したラットの妊娠率及び着床数はTPTClを単独投与したラットよりも高かった[44]．これらの結果から，TPTClによる子宮内膜の脱落膜化抑制は，少なくとも部分的には，卵巣ホルモンを介しており，またプロゲステロンはTPTClによる着床阻害を防御することが示された．

2-2 ジフェニルスズ（DPT）の生殖毒性

ラットに経口摂取されたTPTは，ジフェニルスズ（DPT），モノフェニルスズ（MPT）さらに無機スズに代謝される[45-47]．DPT化合物の生殖毒性試験の結果をTable 3に示した．Diphenyltin dichloride（DPTCl）の妊娠成立及び妊娠維持に対する影響についてラットを用いて検討した[48]．DPTClをWistarラットの妊娠0-3日に4.1，8.3，16.5，24.8 mg/kg，妊娠4-7日に8.3，16.5，24.8，33.0 mg/kgを強制経口投与したところ，妊娠率の低下が妊娠0-3日の24.8 mg/kg，妊娠4-7日の33.0 mg/kgの投与でみられた．妊娠0-3日の16.5mg（48 μmol）/kg以上の投与で着床前の胚致死が増加したが，妊娠の成立した雌の着床前胚死亡率は対照群と同様であった．着床後胚死亡率は妊娠4-7日の33.0 mg/kg 投与で上昇した．これらの結果から，妊娠初期に投与したDPTClは着床阻害を引き起こし，着床前の投与は着床中及び着床直後の投与よりも作用が強く発現することが明らかになった．妊娠0-3日の投与ではDPTClの親化合物であるTPTClも4.7 mg（12 μmol）/kg以上で着床前胚致死作用を示す[49]．モル投与量の比較により，TPTClの作用がDPTClよりも強いことが明らかなので，DPTClまたはその代謝物がTPTClの着床阻害作用の原因物質である可能性は低いと考えられる．しかしながら，TPT化合物はDPTClを投与したラットの肝で生成される[47]ので，投与されたDPT化合物の一部がTPTとして有害作用を発現している可能性があり，DPTの毒性研究の際にはこのことを考慮する必要がある．TPTとDPTによる生殖毒性の差異を明らかにし，その原因物質を明らかにするためには更なる研究を要する．

DPTClの子宮機能に対する影響について偽妊娠ラットを用いて検討されている．Wistarラットの偽妊娠0-3日に4.1，8.3，16.5，24.8 mg/kgのDPTClを強制経口投与した結果，16.5 mg/kg以上の投与で子宮内膜脱落膜化の抑制，偽妊娠4日及び9日の血清プロゲステロンの低下が観察された[50]．これらの投与量はラットの妊娠0-3日に投与したときには着床前胚致死作用を示す量であった[48]．これらの知見は，DPTClはプロゲステロン低下を伴う子宮内