Follow-Up Actions from Positive Results of In Vitro Genetic Toxicity Testing

Kerry L. Dearfield,1 Véronique Thybaud,2 Michael C. Cimino,3 Laura Custer,4 Andreas Czich,5 James S. Harvey,6 Susan Hester,7 James H. Kim,✉8 David Kirkland,9 Dan D. Levy,10 Elisabeth Lorge,11 Martha M. Moore,12 Gladys Ouédraogo-Arras,13 Maik Schuler,14 Willi Suter,15 Kevin Sweder,4 Kirk Tarlo,16 Jan van Benthem,17 Freddy van Goethem,18 and Kristine L. Witt19

1 U.S. Department of Agriculture, Food Safety and Inspection Service, Washington, District of Columbia, USA
2 sanofi-aventis, Vitry-Alfortville Research Center, Drug Safety Evaluation, Vitry sur Seine, France
3 U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, District of Columbia, USA
4 Department of Genetic Toxicology, Bristol-Myers Squibb Co., Research & Development, East Syracuse, New York, USA
5 sanofi-aventis, Deutschland GmbH, R&D Drug Safety Evaluation FFM, Hattersheim, Germany
6 GlaxoSmithKline R&D, Herts, United Kingdom
7 U.S. Environmental Protection Agency, Research Unit Cores/IO, Research Triangle Park, North Carolina, USA
8 ILSI Health and Environmental Sciences Institute (HESI), Washington, District of Columbia, USA
9 Covance Laboratories Limited, Harrogate, United Kingdom
10 U.S. Food and Drug Administration, Office of Nutrition, Labeling and Dietary Supplements, College Park, Maryland, USA
11 Servier Group, Drug Safety Assessment, Biologie Servier, Cedex, France
12 U.S. Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arkansas, USA
13 L’Oréal, International Safety Research Department, Aulnay sous bois, France
14 Pfizer Global Research & Development, Drug Safety Research and Development, Groton, Connecticut, USA
15 Novartis Pharma AG, Preclinical Safety – GeneSafe, Basel, Switzerland
16 Amgen Inc., Thousand Oaks, California, USA
17 National Institute for Public Health and the Environment (RIVM), Laboratory for Health Protection Research, Bilthoven, The Netherlands
18 Johnson & Johnson Pharmaceutical R&D, Genetic & Exploratory Toxicology, Beerse, Belgium
19 National Institute of Environmental Health Sciences, National Toxicology Program, Research Triangle Park, North Carolina, USA

Abbreviations: ADME, absorption, distribution, metabolism, excretion; HESI, Health and Environmental Sciences Institute; ILSI, International Life Sciences Institute; IVGT, Project Committee on the Relevance and Follow-up of Positive Results in In Vitro Genetic Toxicity Testing; IWGT, International Workshops on Genotoxicity Testing; MOA, mode-of-action; PB/PK, physiologically based pharmacokinetics; REACH, European Union legislation Registration, Evaluation, Authorisation and restriction of Chemicals; SAR, structure-activity relationships; WOE, weight-of-evidence.

Disclaimer: This manuscript has been reviewed by the agencies and companies of the authors and approved for publication. The views expressed in the manuscript do not necessarily reflect the policy of these agencies and companies.

David Kirkland is currently at Kirkland Consulting, PO Box 79, Tadcaster LS23 0AS, United Kingdom.
*Correspondence to: James Kim, 1156 15th St, NW, 2nd Floor, Washington, DC 20005, USA. E-mail: j.kim@hesiglobal.org

Received 19 March 2010; provisionally accepted 16 June 2010; and in final form 18 June 2010

DOI 10.1002/em.20617

Published online 20 October 2010 in Wiley Online Library (wileyonlinelibrary.com).
INTRODUCTION

Genetic toxicity testing of agents to characterize the potential adverse outcome(s) in exposed humans is a mainstay of regulatory programs worldwide [Cimino, 2006]. Genetic toxicity data are used for a variety of regulatory purposes ranging from identifying potential adverse outcomes in humans with minimal safety information to providing mechanistic information and performing a mode-of-action (MOA) assessment for data-rich chemicals known to cause cancer, heritable effects, and/or other adverse health outcomes. Genetic toxicity evaluation relies heavily on the use of in vitro tests in the initial battery for the detection of agents that alter DNA and its expression. Because in vitro tests are used in such testing scenarios, it is critical to be able to interpret in vitro findings in relation to in vivo risk of genetic damage that may be involved in potential adverse risk to humans.

Experience shows that the interpretation of results from in vitro genetic toxicity testing is not a straightforward exercise and that experts can differ in their judgment when looking at the same set of test results. The situation is most difficult when decisions must be made based on minimal additional information beyond the genetic toxicity data. It is unrealistic to expect total consensus on interpretation of in vitro genotoxicity test results in all instances, because each situation will be somewhat different and it is always important to consider all the information for decision making. However, it would be of value to provide a consistent, transparent decision process about what to do when confronted with a set of in vitro test results (i.e., a transparent decision-making thought process).

The International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) convened workshops in May 2008 and February 2009, for the Review Subgroup to discuss approaches for interpreting positive in vitro genotoxicity tests and to provide direction for follow-up testing. These workshops are part of a continuing HESI program for examining the relevance and follow-up of positive in vitro genotoxicity tests. Although it should not be assumed there was consensus on all aspects, this article provides the collective considerations of the Review Subgroup at the workshops. This effort drew on the many collaborative efforts over the past several decades that provide the foundation for these considerations. They are provided as a reference tool and are described in brief, in Appendix I.

As a result and taking into consideration the existing publications and guidance, the objective of this effort was to develop, using the existing assays, a decision flow chart for follow-up testing to be applied in case of clear positive results in vitro for whatever type of compound. Considering that recent publications from the International Workshops on Genotoxicity Testing (IWGT) effort provide adequate direction for possible additional testing when negative, marginal, or inconclusive results from the initial set of genotoxicity tests are encountered [Kasper et al., 2007; Ku et al., 2007; Thybaud et al., 2007], these aspects are not discussed here. Similarly, follow-up to in vivo positive results, a topic considered at the 2009 IWGT meeting in Basel [Thybaud et al., in press], is not within the remit of the Review Subgroup, and is not discussed here. The present decision flow chart for follow-up testing of clear in vitro positive results takes into account all available information, considers different options and factors, and allows flexibility in follow-up direction(s) to assist in an overall risk-based decision. To this end, the Review Subgroup also evaluated the strengths, weaknesses/limitations of existing genetic toxicity assays, and their potential use in follow-up testing.
DECISION PROCESS FOR FOLLOW-UP ACTIONS IN CASE OF CLEAR POSITIVE RESULTS IN VITRO

The Review Subgroup constructed a general conceptual flow of logic that can be followed when clear positive results are obtained in the initial genotoxicity tests, usually used in the standard regulatory battery (Section “Decision Process Flow Chart”).

**Decision Process Flow Chart**

The flow chart (Fig. 1) provides guidance on how data from existing assays and other available information can be used to construct an appropriate strategy, which could consist of in vitro and/or in vivo assays, to follow-up on in vitro positive results. The process begins with the interpretation of the existing data (more detail in Section “Step 1: Interpretation of Results from Initial Set of Genotoxicity Tests and Additional Available Data”), that is, genotoxicity and other toxicity data, literature data/information, and possible confounding factors. Among the available data may be information on chemical structure (physico-chemistry and presence of structure alerts for genotoxicity) and chemical behavior in animals (ADME, PB/PK parameters).

Step 2 [Section “Step 2: Weight of Evidence with Mechanism and Mode of Action (Issues) to Consider”] consists of a weight of evidence (WOE) determination. The primary objectives of this step are to determine if the positive in vitro responses may be artifactual (e.g., the treatment conditions in the in vitro system may not be physiologically relevant), to better characterize the nature of the observed in vitro genotoxicity, and if it exists, to use the available body of evidence to determine the mode of action (MOA; e.g., DNA-reactive versus non DNA-reactive genotoxic mechanisms).

Using the WOE approach will allow one to decide (Section “Step 3: Decision Point”) if there is “enough” information to conclude that the positive findings are of concern to the human exposure situation and potential human health risk, or not. If there is sufficient information to make either decision, no further testing is required. If not, this WOE step will help identify any data or knowledge gaps that need to be addressed, and provide justification for the decision to perform follow-up testing (Section “Step 4: Possible In Vitro and/or In Vivo Follow-Up Testing”).

On most occasions, when a clear positive result is obtained in an in vitro test, further testing will be necessary [Thybaud et al., 2007]. If further testing is deemed
necessary, all available data for WOE and MOA considerations are evaluated to identify the most appropriate additional testing (Step 4). Further testing could consist of additional in vitro and/or in vivo assays. In vivo assays may not necessarily be part of the follow-up testing, but can help improve the WOE or confirm the MOA, and provide key data for the assessment of potential risk for humans associated with compound exposure. In vitro testing, on the other hand, has the advantage of providing results more expeditiously, and satisfying ethical concerns (i.e., reduction in use of animals). Moreover, follow-up testing does not necessarily mean a genotoxicity test. Whatever test(s) is chosen, it must address the data/knowledge gap(s) identified in Step 2 so as to improve the WOE, or support the MOA, and thus improve the assessment of risk for humans.

After deliberation in Step 4, the additional testing is performed (Section “Step 5: Run Additional Tests”). Once the additional testing is concluded, the additional data are used in a re-interpretation of the entire data set (Section “Step 6: Re-Interpret All Data and Final Decision Point”). This should provide enough evidence to decide whether the tested chemical is genotoxic, and the genotoxicity implies a relevant concern for humans, or to conclude that the chemical would be of low concern for humans during the intended usage.

**Step 1: Interpretation of Results from Initial Set of Genotoxicity Tests and Additional Available Data**

In providing guidance for the interpretation of genetic toxicity test results and then making a decision or recommendation for any follow-up testing, it is critical that the nature of the research and/or regulatory question(s)/decision(s) be carefully considered [NRC, 2008]. The initial interpretation in Step 1 should identify whether or not those questions are being addressed. For example, if assessment of the initial results provides clear answers to the expected regulatory questions, then further genotoxicity testing may not be needed. Ultimately, the strategy for follow-up testing is likely to be quite different when the sole purpose is hazard identification or prediction of potential carcinogenicity (in the absence of cancer data) than it would be for a mode-of-action assessment and key event identification (e.g., that might not be the induction of DNA damage) for a chemical known to be a mutagen and a carcinogen [Dearfield and Moore, 2005; Moore et al., 2008]. Moreover, as recommended in Dearfield et al. [1991], when discussing the U.S. Environmental Protection Agency’s genotoxicity testing battery: “Results from the initial battery and confirmatory testing (if performed) are reviewed along with all other available relevant information before decisions on subsequent regulatory action are made.” Although not always sufficient to lead to definite conclusions, all this information would be carefully considered in this interpretive step to help develop a decision which may include follow-up testing.

Assays usually required in the initial battery were selected because they are considered to provide a sensitive assessment of the potential for chemicals to induce a wide range of genotoxic effects from small-scale genetic damage (e.g., point mutations) to larger scale genetic alterations (e.g., chromosomal damage) (Table I). Therefore, in interpreting the results from the assays in the initial battery, it is important to understand the strengths and limitations of the assays that are used. These strengths and limitations are presented in Appendix II in Tables V and VI, and a full explanation of the tables is given in Table IV.

Before it can be interpreted, the quality of the data needs to be scrutinized. Proper conduct of the assays [e.g., following good laboratory practices (GLPs), OECD assay procedures], will allow more meaningful interpretation. Since the requirements for conducting assays, and therefore subsequent interpretation of data, have changed over the years, it is important to check that the quality of the primary data is sufficient based on current guidelines and recommendations. Deficiencies or data gaps may be identified (e.g., not all the tests in the standard battery may have been performed) and clarification of results or modification of the tests that have already been performed (e.g., additional or different sampling times) may be needed.

Positive results in different in vitro assays may signify different concerns for humans and require different follow-up strategies. For example, a mutagenic response in the bacterial reverse mutation assay represents a significant hurdle to overcome in the safety assessment process since bacterial mutagens are generally regarded as DNA-reactive and the predictivity of the Ames test for rodent carcinogenicity is high. Generally, 60% concordance and 80% positive predictivity are reported [Zeiger, 1998; Begnini et al., 2010].

In addition to the in vitro genotoxicity test data, the following should be considered in Step 1:

- Data from available physiologically based pharmacokinetic (PB/PK) or pharmacodynamic studies
- Data from absorption, distribution, metabolism, excretion (ADME) studies (these data aid in identification of possible targets—cell types or tissues)
- Structure–activity relationships (SARs)
- Existing knowledge on MOA (e.g., for a cytotoxic anticancer drug)
- Data from any other toxicity tests that are available
- Data on physico-chemical properties (e.g., electrophilic capability)
- Data from *in silico* evaluations
- Any relevant information from the literature.
Positive results in long-term in vivo assays (e.g., chronic bioassays or two generation reproductive studies) provide more definitive assessments of carcinogenic or reproductive risk, respectively, and will significantly contribute to the weight of evidence. If such studies will in any case be conducted on the chemical of concern, specific follow-up testing of a positive in vitro genetic toxicity should be focused on determining the mode of action responsible for any events that could be seen in these long-term assays. It should nevertheless be noted that those data are not always available at the time of decision making (e.g., during drug development). In the instances where chronic bioassays will not in any case be conducted, or if decisions need to be taken before such data are available, specialized short or medium-term tests for genetic damage need to be considered when assessing the genotoxic risk.

**Confounding Factors to Consider During Interpretation**

The possibility that confounding factors could explain or contribute to the positive in vitro responses needs to be addressed at an early stage of the interpretation. The Review Subgroup identified several confounding factors that should be considered, and these are discussed below.

a. Non-physiological conditions

When tests are conducted under extreme non-physiological conditions, artifactual positive results can be generated that are not reliable or relevant for assessing genotoxic risk [Brusick, 1987; Galloway et al., 1987; Scott et al., 1991; Kirkland et al., 2007]. For example, in mammalian cell tests, it is important that pH and osmolality are controlled, and if deviations from normal are observed (particularly low pH in the presence of S9 or high osmolality), that these are considered for their potential to induce artifactual positive responses. Changing the culture conditions or neutralizing the pH shift may eliminate the positive response. It should be noted that most current guidances clearly state that assays should not be conducted under extreme non-physiological conditions.

b. Feeding effects

Feeding effects (e.g., release of histidine from test material) can give rise to increased revertant numbers in a Salmonella assay that are not due to genotoxic properties of the material [Aeschbacher et al., 1983; Gatehouse, 1987]. To exclude such effects, a treat-and-plate protocol can be used.

c. Interactions with test article

Interactions of the test article with the culture medium can occur and may need to be examined. Likewise it is important to check that the chemical is stable in the selected solvent, and chemical reaction did not occur between the test article and the solvent that could lead to a genotoxic reaction product not representative of the hazard potential of the test chemical per se either in rodents in vivo or in humans. Selecting a different solvent may avoid such a confounding effect. It should be noted that most guidances clearly state that absence of interaction of the test article with the solvent and test system should be examined.

Further, test article and solvent interactions may have an impact on metabolism (e.g., inhibition of enzymatic activities) and there could be a non-specific reaction between test article and S9 proteins (i.e., no enzyme activity). This might be suspected if a positive result is obtained in the presence of S9 but there is no evidence that the chemical is metabolized. In addition, some organic solvents inhibit certain P450s in the metabolic activation systems (e.g., S9) routinely used [Chauret et al., 1998; Easterbrook, 2001], and could confound the results.

d. Auto-oxidation

Some chemicals that are prone to auto-oxidation (e.g., polyphenols, catechols) have been shown to be oxidized by the culture medium to produce clastogenic levels of hydrogen peroxide [see Long et al., 2007]. The activity of cells or S9 is not involved in this clastogenic response. Thus, due to the artificial nature of the culture medium, this is not likely to occur in vivo. If such auto-oxidation by the culture medium is suspected, it can be checked by incubating the test compound with culture medium without cells or S9, and measuring production of hydrogen peroxide [see Long et al., 2007]. As not all culture media have the same potential for auto-oxidation (e.g., Hams F10 and F12 were shown by Long et al., 2007 to produce much lower levels of hydrogen peroxide than other media), changing the cell system or selecting other media may eliminate the problem. Whether or not positive results are confirmed in other cell/culture systems detecting clastogenic damage could provide insight as to whether auto-oxidation is a likely explanation for the positive results.

Follow-up testing may be needed to simply address the individual assay result before proceeding to Step 2. If suspicion exists that the positive response is the result of culture conditions or interactions with culture media or solvents, the follow-up may be conducted with other cell-types/cell-lines and/or under different conditions (e.g., other metabolic enzyme system).

**Metabolism**

Some chemicals (promutagens) need to be converted into reactive metabolites to induce mutagenic activity, for example, via oxidative metabolism or in few cases due to Phase 2 metabolism [e.g., acyl glucuronides (Ritter, 2000), benzylic alcohols (Glatt, 2000)]. Generally, in vitro genetic toxicity studies employ mammalian based systems, and a metabolic activation system (S9 homogenate...
and cofactors) is usually included during conduct of in vitro genotoxicity assays to provide a source of metabolizing enzymes, typically from Aroclor 1254 or phenobarbital/β-naphthoflavone-induced rat liver, focusing on enhanced oxidative metabolism.

A simple assay outcome that suggests formation of a mutagenic metabolite (activation) occurs when a test chemical is positive in the presence of S9, but negative in the absence of S9 under the same experimental conditions (e.g., same treatment time). Conversely, a negative result in the presence of S9, but a positive result in the absence of S9 obtained under the same experimental conditions (e.g., same duration of treatment) suggests detoxification of the parent compound by the S9 activation system. It might be useful to consider whether different study types give a consistent pattern of results (e.g., in both prokaryotic and eukaryotic cells, or for different endpoints in mammalian cells such as chromosomal aberrations and gene mutations). These considerations would reinforce either interpretation (active metabolite or detoxification) and add to the weight of evidence for Step 2.

In addition to activation and detoxification, other implications that need to be considered include the nature of the exogenous metabolic system, the nature of the metabolic profile for human risk assessment, or the generation of unique, possibly not relevant, metabolites (see following).

a. Nature of the exogenous metabolic system
   The source of the exogenous metabolic system (e.g., purchased microsomes or S9 mix made in the test laboratory) should be evaluated for the uniformity of its preparation (particularly if being used across laboratories and different experimental runs) and the levels of constituent and inducible enzyme activities. In general, co-factors are added to ensure oxidative Phase 1 metabolism.

   The percentage of S9 used is known to potentially affect the metabolic profile. Whether the S9 mix is made from induced or non-induced sources, or from different species (e.g., human), can influence the metabolic capability of the exogenous system.

   Because of high concentrations of chemical used in most in vitro studies, saturation of metabolic capacity and even enzyme substrate inhibition might occur, and the detoxification limit can be exceeded [as for paracetamol, Bergman et al., 1996]. This can influence the metabolic pattern of a compound (generation of reactive oxygen, the lack of mainly detoxifying phase 2 metabolism in standard S9 preparations), and produce a different metabolic pattern from that seen at lower and possibly more physiologically relevant concentrations [Ku et al., 2007]. If such effects are suspected, the experiment could be repeated under modified experimental conditions, and cofactors for Phase 2 metabolism (e.g., glutathione), antioxidants or free radical scavengers could be added to overcome the lack of detoxification, to mimic more physiological situations, and possibly to demonstrate detoxification processes for mechanistic considerations.

b. Nature of the metabolic profile
   It can be useful to compare the metabolic profile in humans with those in the animal models being used. Where the metabolic profile in rodents is different from the profile in humans, or where metabolites generated by other systems (e.g., plants) to which humans are exposed are different from those produced by rodent metabolizing systems, it might be appropriate to use human microsomes, human cells with metabolic capability, and/or human-derived metabolizing systems to address human relevance and provide a useful comparison with the animal systems. Nevertheless, even test systems of human origin need to be evaluated for relevance of possible adverse effects in humans.

   Once the metabolite profile and metabolizing systems have been analyzed, the interpretation still may not define the metabolism issue precisely. For example, it is not easy in many instances to demonstrate the absence of a metabolite in vivo that may be considered responsible for the genotoxicity in vitro, taking into account the sensitivity of analytical methods. However, negative results from in vivo genotoxicity and/or carcinogenicity studies may be sufficient to demonstrate the absence of risk from genotoxic metabolites in in vitro tests [Dobo et al., 2009].

c. Unique metabolite(s)
   In some cases, a metabolite may be uniquely generated by the in vitro test system. Rarely is a compound uniquely metabolized by bacteria to a genotoxic substance capable of inducing reverse mutations [Blumer et al., 1980; Hrelia et al., 1998]. However, if such a situation is suspected, demonstration that bacterial specific metabolism is responsible for the mutagenic response could be pursued to judge the relevance of the positive findings for mammalian cells [Suter et al., 2002]. A non-mutagenic response using metabolic inhibitors or bacterial strains deficient in the metabolic activity suspected to be responsible for the unique mutagenic metabolite could suggest that the original response was likely to be due to a bacterial specific metabolite. Moreover, metabolite profiles generated by standard tester bacteria, metabolic mutant bacteria, and human or rat hepatocytes could be evaluated for differences. For additional insight, the compound could be tested in an in vitro mammalian system that detects point mutations (e.g., hprt, tk).

Impurities

If the positive result (for example, in a bacterial reverse mutation assay) is suspected of being due to an impurity, further testing would be needed. Such a situa-
tion may be indicated if there is evidence or information to suggest the test compound itself or projected metabolite(s) may not be the cause of the positive response (e.g., no known structural alert, compound class is generally not considered mutagenic). If an impurity is suspected, it should be isolated, if possible, and tested separately to help determine whether the parent compound and/or the impurity are responsible for the observed mutagenic activity. For further information, see EMEA [2006] “Guideline on the Limits of Genotoxic Impurities” (CPMP/SWP/5199/02). Frequently, the impurity cannot be identified or isolated and the test material is purified and tested again. If the test result is negative, there is usually no need for further testing, even if the impurity cannot be identified.

**Step 2: Weight of Evidence with Mechanism and Mode of Action (Issues) to Consider**

Selection of follow-up testing should be based on the knowledge available about the nature and/or mechanism/mode of action of the original response. Consideration of a hypothesized mode of action (MOA) is a powerful approach to direct possible follow-up testing, and helps clarify the key questions that need to be addressed.

Many international efforts have provided guidance for conducting an MOA analysis [Sonic-Mullin et al., 2001; USEPA, 2005; Boobis et al., 2006, 2008]; these largely discuss frameworks for carcinogenic MOAs, but the frameworks can be applied easily to MOAs for mutation and to MOAs for other adverse effects. By proposing a mutagenic MOA, specific follow-up strategies can be identified that will provide evidence to support or refute the proposed mutagenic MOA. Finally, by comparing possible MOAs in animal models to humans [e.g., Meek, 2008], the relevance of the in vitro genotoxicity findings for human risk can be more clearly addressed [see e.g., USEPA, 2007]. This MOA approach provides a preferential framework that organizes the data and thought process to help focus on the most appropriate follow-up testing.

If a MOA is not evident, or a hypothesized MOA is not supported by sufficient data, a weight of evidence (WOE) approach can be followed to determine whether genotoxicity in vitro indicates a cause for concern for humans under the expected or measured exposure scenario(s) for humans. The data from Step 1 are used to determine if “enough” data exist to indicate a concern for humans. A “weighting” scheme proposed by EPA [USEPA, 2007; based on the EPA Mutagenicity Risk Assessment Guideline] would be appropriate here. If the WOE approach indicates there is not “enough” data to indicate a concern, then data/information gaps to address that deficiency need to be identified—this will become the basis for follow-up testing.

Some general issues the Review Subgroup considered for a WOE/MOA determination that may help direct follow-up testing are described below.

**Direct DNA Reactive Versus Non-Direct DNA Reactive Mode of Actions**

Genotoxic activity may result from a direct effect on the genetic material or an indirect effect. It is important to distinguish between these two general MOAs, namely whether a compound acts via a direct mechanism (i.e., direct DNA reactive), where DNA is the primary target (i.e., DNA adducts, thymidine dimers, and in some cases strand breaks), or an indirect mechanism (i.e., non-direct DNA reactive) where primary targets other than DNA are involved (DNA being secondarily damaged, e.g., via free radicals, reactive oxygen or nitrogen species, nucleotide pool imbalance, spindle disruption, inhibition of DNA synthesis or topoisomerases). In both cases, the end result is still a genotoxic insult and possible genetic alteration. However, using an MOA approach to distinguish between the two can provide better guidance on appropriate follow-up testing and also provide insight into possible dose response relationships (linear versus non-linear) that can contribute to the ultimate assessment of human risk.

Generally, evidence for direct DNA reactions can be readily obtained (e.g., adduct formation) and used in establishing a MOA. However, if the evidence suggests a MOA not involving direct or proximate reactivity with the DNA, then tests that can provide evidence to exclude direct DNA reactivity should be considered. It can be difficult to demonstrate and explain a non-DNA reactive mechanism, particularly if it cannot be identified with the well-described indirect mechanisms listed above. In such cases, follow-up with appropriate robust assays that can demonstrate the absence of interaction with DNA are advisable (see assays to evaluate DNA reactivity in Table II).

For example, a compound may induce mutations by non-DNA reactive mechanisms that are unique to bacteria or are manifested to a greater extent in bacteria. This could be via inhibition of a bacterial enzyme such as the type II topoisomerase (DNA gyrase). Because of its central role in nucleic acid metabolism, inhibition of gyrase can lead to base pair substitutions, frameshifts, and small deletions. When tested in the bacterial reverse mutation assay, some intercalators have been shown to be mutagenic only in one strain, such as TA102 or TA1537 [Ferguson et al., 1990; Albertini et al., 1995]. Although gyrase inhibitors can interact with mammalian DNA, inhibition of mammalian topoisomerase II requires much greater concentrations of intercalators than is typically achieved in a clinical setting or at environmentally relevant exposures. At high concentrations, gyrase inhibitors may lose specificity, inhibit the mammalian enzyme, and
show ‘off target’ activity not relevant to in vivo conditions, or to clinical exposure levels. Demonstration of intercalation-induced frameshift mutations could require additional in vitro and in vivo testing, such as a DNA dye displacement assay, in vitro bleomycin amplification micronucleus assay [Snyder and Strekowski, 1999; Snyder and Diehl, 2000], or an in vivo micronucleus assay.

As discussed in Kirkland et al. [2007], while numerous potential non-DNA targets, when disrupted, have been postulated as sources of genotoxic responses in mammalian systems, especially chromosome damage, obtaining clear evidence is often difficult. If damage to a non-DNA target or process is suspected, follow-up testing may provide experimental evidence to support this hypothesized MOA. In the absence of clear evidence of a non-DNA mechanism, it may be necessary to examine whether clastogenic responses are associated with DNA reactivity in the cells used. For example, if inhibition of protein synthesis is suspected, the concentrations at which protein synthesis is inhibited can be compared with those at which chromosomal aberration induction occurs to see if they correlate. It should be noted that disruption of a non-DNA target may lead to genotoxic responses in vivo, as well as in vitro, if high enough concentrations can be achieved in the target tissue. Thus, inhibition of protein synthesis by cycloheximide [Sato et al., 1990] or inhibition of topoisomerase by fluoroquinolones [Mukherjee et al., 1993; Singh et al., 2003] can lead to induction of micronuclei in vivo. Moreover, as mentioned earlier, if impact on a detoxification process (e.g., saturation) is suspected, the addition of free radical scavengers, antioxidants, glutathione, etc., may eliminate or abrogate the genotoxic response.

Aneuploidy/Polyploidy Issue

In some cases, an in vitro mammalian chromosome aberration test may show an increase in polyploid cells. This could also be accompanied by an increase in mitotic index. Induction of polyploidy per se is not considered a serious health risk, as many normal human tissues (e.g., liver) contain polyploid cells [Muehlbauer et al., 2008], but induction of aneuploidy is a concern for health. Chemically induced polyploid cells could have arisen by multiple events during replication and cell division, or may be genomically unstable leading to aneuploidy. Therefore, in such cases it is important to establish whether an aneuploidy mechanism is involved. This can be done most conveniently by conducting an in vitro micronucleus (MN) test. If MN are not induced, then the polyploidy may be due to cytotoxic effects during which cytokinesis is disturbed. Evidence of a ‘polyploidy only’ effect could be further obtained by use of a flow cytometric method to determine the DNA content of the cells, which is likely to be more sensitive than counting polyploid cells on slides under a microscope, unless a large number of cells is scored. However, if MN are induced and no structural chromosome aberrations were observed in the in vitro mammalian chromosome aberration test, the compound is most likely an aneugen. To improve the weight of evidence, MN could possibly be probed for the presence of centromeres to confirm that they do not result from chromosome breaks. If an aneuploidy mechanism is demonstrated (i.e., MN without structural chromosome aberration and/or predominant induction of centromere-positive MN), it suggests that a non-linear dose response may be applicable depending on the mode of action. This can be achieved by performing a non-disjunction assay using chromosome-specific probes in binucleated cells [Bentley et al., 2000].

Step 3: Decision Point

Step 3 is a major decision point to decide on whether further testing is required. As long as there are sufficient data for a decision (i.e., to address the questions being posed), two options are considered:

1. whether the chemical under scrutiny is considered genotoxic,
2. whether the evidence in the context of the WOE/MOA analysis presents a low (negligible) concern for humans under the usage expected or measured.

In either case, no further testing is necessary and a decision (e.g., for regulatory purposes) is rendered.

In the case of option 1, this decision would lead to the assumption that the chemical is a potential human genotoxic carcinogen and/or germ cell genotoxicant. The decision to stop testing could also be driven by the possibly disproportionate cost of follow-up activity to further evaluate the effect (e.g., in other test systems), in the context of compound usage. Such a decision, therefore, needs to be taken in the context of the projected use of the compound (e.g., clinical indication, labeling, restriction of use, etc.). In such cases, consideration should be given to whether the compound provides enough benefit to balance potential risk and the total time and cost required to fully investigate the potential mutagenic and/or carcinogenic risk posed by the compound.

Generally, if data are sufficient to establish a MOA, then a decision for its relevance to an adverse effect (i.e., concern or not) can be made. If a decision regarding whether the compound is a concern for genotoxic human risk cannot be made, then follow-up testing is needed to address the data/information gap(s) identified in the WOE/MOA determination (Step 4).
TABLE I. Table of Endpoints and Relationship Between Standard In Vitro Genotoxicity Assays

<table>
<thead>
<tr>
<th>Standard in vitro assays</th>
<th>Bacterial reverse mutation assay (Ames assay)</th>
<th>Mammalian gene mutation assays (e.g., hprt)</th>
<th>Mouse Lymphoma Assay - Large Colonies</th>
<th>Mouse Lymphoma Assay - Small Colonies</th>
<th>Micronucleus assays</th>
<th>Chromosome aberration assays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoints detected by in vitro assays</td>
<td>Gene mutations</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Structural chromosome damage</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>Numerical chromosome damage</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>Polyploidy</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+++ efficiently detected; ++ detected depending on test design; + not always detected; – not detected.

Step 4: Possible In Vitro and/or In Vivo Follow-Up Testing

As described above, several factors need to be considered when evaluating the nature of the in vitro genetic activity seen, and in selecting follow-up tests. These factors include the nature of the compound itself (e.g., its potential reactivity and instability; its ability to form electrophilic entities; the presence of structural alerts for genotoxicity), the ability of the assay to detect specific genetic endpoints, the nature of the response in the in vitro assay(s) (e.g., level of the lowest positive concentration; magnitude of the effect; the nature of the dose-response curve), and the intended use of the compound.

The Review Subgroup noted that follow-up testing does not necessarily mean further tests for genotoxic activity. If further genotoxicity testing is considered necessary, then it needs to be decided if in vitro tests can address the data/information gap(s), or whether in vivo testing will be necessary. Regardless of the follow-up testing chosen, it must address the data/knowledge gap(s) identified in Step 2 and improve the WOE and assessment of risk for humans.

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 (MLA)]. Generally, chemicals identified as inducing primarily point mutations would better be evaluated in follow-up assays that detect point mutations. Similarly, chemicals that primarily induce chromosomal mutations would better be further evaluated using an endpoint that detects chromosomal mutations. Table I provides general advice on the endpoints most readily evaluated by the various standard assays; this information should be helpful in identifying the most appropriate assays for endpoint-specific follow-up action.

An example may prove useful. In developing a follow-up strategy for positive results in in vitro mammalian gene mutation assays, it is important to first determine what general types of genetic alterations are induced. There are two commonly used in vitro mammalian gene mutation assays, the mouse lymphoma assay (MLA) using the thymidine kinase (tk) gene as the reporter gene, and various cell lines using the hypoxanthine-guanine phosphoribosyl-transferase (hprt) gene as the reporter. Because these two reporter genes detect different types of genetic damage, it is important to understand the properties of both genetic loci. The tk gene, probably because of its autosomal location, detects a full array of genetic damage including point mutations, deletions, chromosomal rearrangements, mitotic recombination and nondisjunction [Applegate et al., 1990; Chen et al., 2002; Wang et al., 2009]. The tk mutants generally fall into two categories based on colony size (small and large). Chemicals that are known to induce point mutations (and have little or no clastogenicity) will induce primarily large colony mutants. Chemicals that are known to be clastogens (and have little or no ability to induce point mutations) will induce primarily small colony mutants and, as indicated above, will also be negative in the bacterial reverse mutation test. Though the proportion of small versus large colonies is helpful information, it should be acknowledged that most chemicals induce both point mutations and chromosomal changes, and therefore induce both small and large colony tk mutants.

Because of its location on the non-homologous part of the X-chromosome (i.e., there is only a single copy), the hprt gene detects point mutations and some deletions, but is not efficient at detecting large deletions and cannot detect mitotic recombination [Moore et al., 1989]. It is also important to recognize that bacterial reverse mutation assays detect only point mutations—deletions of the target his or tryp genes would not result in reverse mutation. Thus, when a chemical is negative in a bacterial assay but positive in an in vitro mammalian assay (particular the MLA and cytogenetic endpoints), it could represent a difference between prokaryotic and eukaryotic cells, but it is also likely that the chemical induces chromosomal damage rather than point mutations. Even if rarely conducted,
except for research purposes, further investigation of the specific types of damage induced by a particular chemical, could potentially be achieved by isolating a series of mutants and evaluating them using a combined strategy of cytogenetic and molecular analysis [Wang et al., 2009].

Because of many variables that need to be considered, it is clear that a defined list of follow-up actions cannot be provided for every in vitro positive scenario. The Review Subgroup recognized, however, that some general guidance can be given. Table II provides suggestions for follow-up assays when different positive in vitro findings are encountered. The table indicates the in vitro and in vivo assays most appropriate to confirm the induction of gene mutations, or numerical or structural chromosome damage. It also indicates the appropriate mechanistic studies that can be conducted to evaluate possible mode(s) of action (e.g., presence or absence of DNA reactivity in vitro or in vivo).

After performing the WOE determination, it is necessary to decide whether and which in vivo assays need to be performed. As mentioned previously, the same endpoint(s) (gene mutations and structural or numerical chromosome aberrations) that is/are positive in in vitro tests should be investigated in an adequate in vivo test. Therefore, if the chemical appears primarily to be a point mutagen (positive in the bacterial reverse mutation test, positive for hprt mutations, and/or induces predominantly large colony tk mutants), it is appropriate to perform a test system that optimizes for the detection of point mutations such as an in vivo gene mutation assay (e.g., transgenic gene mutation assays), since cytogenetic endpoints conducted in vivo would not provide appropriate follow-up.

Conversely, if the chemical does not induce point mutations (negative in the bacterial reverse mutation test and mammalian cell hprt mutation tests, and/or induces predominantly small colony tk mutants), then the appropriate in vivo follow-up would be a test with a cytogenetic endpoint. Cytogenetic evaluations in vivo are usually limited to cells of the peripheral blood or bone marrow, although micronuclei can be measured in liver and other tissues [Hayashi et al., 2007]. If other tissues need to be studied (e.g., because distribution studies indicate tissue accumulation), then DNA damage assays such as the Comet assay can be considered for follow-up of cytogenetic damage.

For the majority of chemicals inducing both point mutations and chromosomal effects, both in vivo cytogenetic and gene mutation assays should be considered, but may be influenced by the question to be answered and the degree of follow-up desired. Table III can be used to help select the most appropriate in vivo assay(s). If a test predominantly detecting one endpoint is negative, a second in vivo test that investigates the other endpoint is necessary. However, it is possible to select one in vivo test that detects (e.g., gpt transgenic animals) both point mutations and chromosomal effect, or precursor events resulting in gene mutations and/or chromosome damage (e.g., at least some types of DNA adducts, Comet assay) (Table III). Another alternative would be to evaluate two different end-points in the same study and animals (e.g., a study combining a bone marrow micronucleus test and a liver Comet assay, or when the Pig-a assay is validated, a combination of Pig-a gene mutation and micronucleus test in bone marrow).

Absorption, distribution, metabolism, and elimination (ADME) and physiologically based pharmacokinetic (PB/PK), pharmacodynamic, and other toxicological studies (to address mechanism, MOA) may be included in the follow-up testing in addition to any other genotoxicity tests, and are extremely useful to assess possible follow-up action. Use of such data can help determine appropriate targets to study if in vitro or in vivo. For example, PB/PK data can be used to ensure exposure at target tissues, particularly bone marrow or peripheral blood, but it should be noted that the plasma concentration may be an underestimate of the exposure at the first tissue of contact. If in vivo results are available (as they may be in many circumstances) and show good exposure of the target tissue(s), but the in vitro and in vivo studies give conflicting (positive and negative) results, the metabolism of the chemical under the conditions of the in vitro assay could be investigated (see metabolism section in Step I) in order to better understand the discrepancy.

Metabolic profiles could help identify if the parent or a metabolite should be tested in follow-up assays. However, results obtained with isolated metabolites should be interpreted with caution, because the oral administration of a metabolite that is usually produced in liver may for example lead to different tissue distribution and metabolic consequences. Although it may prove useful to test one or more of the metabolites in the in vitro assays, knowledge of the amounts of compound or metabolite reaching various targets can help determine the choice of cell type or tissue to study. In many instances (e.g., pesticide and drug registrations), such ADME and PB/PK data are available, and understanding the kinetics of compound distribution and elimination will help optimize the design of the follow-up assays. In addition, if there are ADME and PB/PK data for humans, comparisons between the animal models and humans can be made to determine possible relevance of the in vitro genotoxicity results and the appropriate targets and endpoints to study in follow-up testing. For this purpose, both maximum concentration (C_max) and area under the curve (AUC) should be considered.

**Step 5: Run Additional Tests**

Once the analysis in Step 4 is complete, additional testing is performed. The Review Subgroup agreed that as a general principle for follow-up testing, optimal informa-
### TABLE II. Potential Follow-Up Assays to Clarify In Vitro Positive Findings

| Assays that can be chosen to follow-up in case of in vitro positives: for mechanistic information and/or confirmation of in vitro findings | DNA adduct assay | UDS assay | Comet assay | DNA primary damage | Point mutations | Point mutations | Deletions | Micronuclei assays | Structural chromosome damage | Numerical chromosome damage | Structural chromosome damage | Chromosome aberration assays | Assays for non-DNA reactive mechanisms |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Endpoints detected by the above assays** | Adducts | Breaks | Gene mutation assays | In vitro assays (e.g., hprt) | Transgenic models (a) | Micronuclei without centromere | Micronuclei with centromere | Chromosome aberration assays | Assays for non-DNA reactive mechanisms |
| In case of positive in vitro gene mutation assays | In vitro assays | To evaluate DNA reactivity | To confirm gene mutation endpoint in vitro | – | – | – | – | – | To evaluate the evidence supporting non-DNA mechanisms |
| | In vivo assays | To further evaluate DNA reactivity in vivo | – | To further evaluate gene mutation endpoint in vivo | – | – | – | – | – |
| In case of positive in vitro chromosome damage assays | In vitro assays | To evaluate DNA reactivity | – | – | – | To confirm the induction of chromosome damage endpoint in vitro, and to differentiate clastogen from aneugen mechanism |
| | In vivo assays | To further evaluate DNA reactivity in vivo | – | To further evaluate chromosome damage endpoint in vivo | – | – | To further evaluate chromosome damage endpoint in vivo, and to differentiate clastogen from aneugen mechanism, if positive |

(a) for more details on transgenic mutation assays and their ability to detect point mutations and deletions, see Heddle et al. (2000) and Thybaud et al. (2003).
Useful in some circumstances, as originally presented in testing performed in light of the WOE determination. For an understanding of the relevance of the follow-up actions is paramount for a judgment of genetic toxicology experts becomes a critical concern. It is particularly difficult to interpret positive in vitro results that are not supported by appropriate in vivo testing), or may require a full ‘research project’ [Dearfield and Moore, 2005]. Because each WOE/MOA determination is different depending on the questions being addressed, the specific requirements for follow-up testing in each situation do not fit well into a one-size-fits-all approach. Flexibility in approach is important for identifying follow-up testing that is appropriate for the particular circumstances of the chemical or drug being evaluated.

Step 6: Re-Interpret All Data and Final Decision Point

After all the follow-up testing has been conducted (i.e., after in vitro and/or in vivo assays and/or any other additional testing), the existing and new data will be evaluated to assess possible human risk. At this point, sufficient evidence should be available to decide on the level of concern for the chemical in question, given its use and the likelihood of human exposure. There is no absolute ‘yardstick’ to measure whether the information provides ‘enough’ (sufficient) evidence to decide that there is a low or high concern. It is particularly difficult to interpret positive in vitro results that are not supported by appropriate in vivo test results (e.g., where the in vivo test assayed a different endpoint from that found positive in the in vitro test). The extent of negative in vivo data that are needed to fully address the positive in vitro results also cannot be defined for all cases, and will depend on the issues being addressed. It is important to note that in these cases, the scientific judgment of genetic toxicology experts becomes a critical factor. Clear justification of the follow-up actions is paramount for an understanding of the relevance of the follow-up testing performed in light of the WOE determination.

Viewing the evidence in a hierarchical fashion may be useful in some circumstances, as originally presented in the EPA’s mutagenicity guideline [USEPA, 1986]. Briefly, there is increasing concern for human risk when:

- the in vivo genotoxicity tests are positive, thus confirming the in vitro tests,
- the positive results are obtained in appropriately performed tests,
- the positive results are not caused by extreme conditions of exposure to target cells/tissues, such as high pH and osmolality for tissue cultures and above maximum tolerated doses for in vivo studies,
- the WOE for a concern is strengthened if a mode of action is similar to that found in other (toxicity) tests, and is most likely a genotoxic key event.

Although negative in vivo genotoxicity tests do not automatically negate clear positive in vitro tests [Elespuru et al., 2009], the possible concern for human risk is likely to be lessened when all the available data, including appropriate negative in vivo follow-up test results, are consistent in pointing towards a low level of concern. Other considerations that may lower a level of concern include:

- the negative result in the in vivo test is not due to lack of exposure of the target cells,
- in silico methods that do not predict a genotoxic mode of action,
- DNA binding, adduct, or strand-break studies do not support observed genotoxic activity via direct interaction with genetic material, and/or
- the compound under investigation belongs to a group of chemicals which are considered to be non-genotoxic.

**SUMMARY AND PERSPECTIVES**

The Review Subgroup developed a decision process flow chart that can be used to address positive results obtained in one or more in vitro genetic toxicity assays. General considerations for analyzing results from an initial battery of genotoxicity tests have been discussed in relation to the choice

<table>
<thead>
<tr>
<th>Follow-up to in vitro positives</th>
<th>Unscheduled DNA synthesis</th>
<th>Transgenic models (a)</th>
<th>DNA adducts</th>
<th>Comet assay</th>
<th>Micronucleus assays</th>
<th>Chromosome aberration assays</th>
</tr>
</thead>
<tbody>
<tr>
<td>In vitro gene mutation assay</td>
<td>+++</td>
<td>++</td>
<td>+</td>
<td>++*</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>In vitro structural chromosome damage</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+*</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>In vitro numerical chromosome damage</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>++*</td>
</tr>
<tr>
<td>In vitro polyploidy</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

+++ efficiently detected; ++ detected depending on test design or * not enough data available; + not always detected; – not detected (a) for more details on transgenic mutation assays and their ability to detect point mutations and deletions, see Heddle et al. (2000) and Thybaud et al. (2003).
and design of follow-up tests. Specific considerations, for example, confounding factors, metabolic issues, and modes of action, have been elaborated to suggest options for follow-up actions. Moreover, the Review Subgroup evaluated the relevance, reliability, and scope of existing genotoxicity assays to help define their usefulness for any in vitro or in vivo follow-up testing. The Review Subgroup also indicated different in vitro and in vivo follow-up actions for different scenarios since no single follow-up action will satisfy all cases. The Review Subgroup also indicated considerations relevant to human risk assessment, for example:

- are the effects biologically relevant considering human usage (e.g., magnitude of the effects, concentrations compared to human exposure)?
- are the effects cell specific?
- when in vivo data are available, are the effects seen in vitro also seen in vivo?
- for either an indirect or a direct interaction with DNA mechanism, is it reasonable to expect these effects to occur in humans under normal conditions of use (e.g., therapeutic dose, environment contamination, food intake)?

In addition to the discussions provided here on the improved use and interpretation of the existing assays for better human risk assessment, the ILSI/HESI workshop convened two other working groups, a Quantitative Subgroup and an Emerging Technologies and New Strategies Workgroup. The Quantitative Subgroup discussed the possibility of moving from qualitative to a more quantitative evaluation of genotoxicity data. Also, depending on the mode of action, dose response relationships in the different assays might be compared with expected or observed human exposure(s) to determine potential genetic risk at different dose levels in humans. This could be done at different levels, when evaluating in vitro and in vivo data, and when extrapolating from in vitro to in vivo data, and from animals to humans. The quantitative evaluation paradigm, when available, can be integrated in the decision process flow chart presented here to further improve human risk assessment. This work is still in progress. The Emerging Technologies and New Strategies Workgroup examined the development and potential of new assays and/or the refinement of the existing assays (i.e., those considered in the current article) that could contribute to a more relevant genotoxicity evaluation, and more insight into in vitro positive findings. A summary of the discussion of the new/emerging technologies working group is presented in a separate publication in this issue.

APPENDIX I: ACTIVITIES ADDRESSING THE TESTING SCHEME FOR GENETIC TOXICOLOGY

The discussion provided in this paper came about from the long and fruitful efforts in the genetic toxicology community to examine the usefulness and relevance of the tests routinely used to determine genotoxicity potential. Many collaborative efforts over the past several decades provided the foundation for these considerations, including:

NATIONAL, REGIONAL AND INTERNATIONAL ENVIRONMENTAL MUTAGEN SOCIETIES

National, regional and international Environmental Mutagen Societies are composed of scientists from academic laboratories, regulatory authorities and industries. These Societies organize regular meetings (e.g., annual meetings for the national and/or regional societies; the international meetings are held every few years), provide a forum for the presentation of new data, and host discussions on new methods, approaches and strategies. They also organize and coordinate collaborative efforts. They play a key role in maintaining the links and networks between genetic toxicology experts and the training of young scientists.

OECD

The Organisation for Economic Co-operation and Development (OECD) develops and coordinates environmental health and safety activities on an international level. These activities include (1) harmonizing chemical testing and hazard assessment procedures; (2) harmonizing classification and labeling; (3) developing principles for Good Laboratory Practices (GLPs); (4) cooperating on the investigation of existing chemicals; and (5) sharing and exploring possible cooperative activities on risk management (Cimino, 2001). As part of its program to evaluate chemical safety, the OECD has developed Guidelines for the Testing of Chemicals (OECD, 2005), for use by government, industry, and independent laboratories when testing the safety of new and existing chemicals, pesticides, pharmaceuticals, and food additives.

IWGT

A series of International Workshops on Genotoxicity Testing (IWGT, formerly IWGTP) has taken place since 1993. Groups of experts from academia, government regulatory bodies, and industry, representing the American, European and Asian continents, were brought together to achieve data-driven consensus recommendations on test design and interpretation (Kirkland, 1994; also see references for IWGT reports generated by the workshops). The recommendations have been important in revisions to existing OECD 1997 guidelines and in drafting new guidelines such as for Comet and transgenic mutation assays and the in vitro micronucleus test.
**ICH**

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) workgroup produced two guidelines for genotoxicity testing: S2A “Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals” (ICH, 1995), and S2B “Genotoxicity: A standard battery for genotoxicity testing of pharmaceuticals” (ICH, 1997). These documents provide the basis for genotoxicity testing and assessment of pharmaceuticals in the EU, Japan, the U.S., and elsewhere (Müller et al., 2001). A new draft, S2(R1) “Guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use,” has been published recently combining the first two documents, with the goal of optimizing the standard genetic toxicology battery for prediction of potential human risk, and providing guidance on interpretation of results in view of improving risk characterization for carcinogenic effects that are based upon changes in genetic material (ICH, 2008). The revision provides standards for follow-up testing and for interpretation of positive results in vitro and in vivo, including assessment of non-relevant findings.

**WHO**

The U.N. World Health Organisation/International Programme on Chemical Safety (WHO/IPCS) has considered the need for and nature of revision to its genetic toxicology testing battery, as originally published in Ashby et al. (1996). WHO/IPCS has involved the international community of scientists, government regulatory bodies, industry, and the public to participate in the effort (WHO, 2007). The recommendations have been published recently (Eastmond et al., 2009).

**GENETOX PROGRAM**

This database was created by the U.S. Environmental Protection Agency (EPA) and contains genetic toxicology test data resulting from expert peer review of the open scientific literature on over 3,000 chemicals. The program was established three decades ago to select assay systems for evaluation, to review data in the scientific literature, and to provide a basis for recommending proper testing protocols and evaluation procedures. The complete list of published reports (with the sole exception of the update for the MLA (Mitchell et al., 1997) is provided in Cimino and Auletta (1994). Some of the assays examined are routinely employed in current test batteries, some are no longer commonly conducted, and some are used for special purposes. The Gene-Tox database is publicly available on-line via the U.S. National Library of Medicine’s (NLM) Toxicology Data Network (TOXNET) (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX).

**NTP**

The National Toxicology Program of the U.S. Department of Health and Human Services was established in 1978 as a cooperative effort to (1) coordinate toxicology testing programs; (2) strengthen the science base in toxicology; (3) develop and validate improved testing methods; and; (4) provide information about potentially toxic chemicals to health, regulatory, and research agencies, scientific and medical communities, and the public. The NTP publishes studies for diverse toxicological endpoints, including genetic toxicity and carcinogenicity [http://ntp.niehs.nih.gov].

**NICEATM, ICCVAM, ECCVAM AND JaCVAM**

The NICEATM (NTP Interagency Center for the Evaluation of Alternative Toxicological Methods), which was established in 1998, and ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods), which NICEATM administers, seek to promote the validation and regulatory acceptance of toxicological test methods that will enhance their member agencies’ ability to assess risks and make decisions, and that will refine, reduce, and/or replace animal use. The ultimate goal is the validation and regulatory acceptance of test methods that are more predictive of adverse human and ecological effects than currently available methods [http://iccvam.niehs.nih.gov]. The comparable activities in the European Union and Japan are EECVAM (European Centre for the Validation of Alternative Methods [http://ecvam.jrc.it]) and JaCVAM (Japanese Centre for the Validation of Alternative Methods [http://jacvam.jp/en/]; Kojima, 2007), respectively. All three activities cooperate with one another and pool their data, efforts, and resources.

**GOVERNMENT REGULATORY TESTING SCHEMES**

Many activities were (and are) conducted to support and aid in the design of government regulatory testing schemes (see Dearfield et al., 1991; Cimino, 2006; and references therein). Meetings such as the EPA-sponsored one in Williamsburg, VA in January, 1987 (Workshop on the Relationship Between Short-Term Test Information and Carcinogenicity) have spurred reconsideration of the role of short-term mutagenicity tests in toxicity testing (Auletta and Ashby, 1988; Kier, 1988). More recent activities, such as the review and update of the mutagenicity testing scheme by the United Kingdom COM (Committee on Mutagenicity of Chemicals in Food, Consumer Prod-
REACH

REACH is a new European Union regulation concerning the Registration, Evaluation, Authorisation and restriction of Chemicals. It came into force on 1st June 2007 and replaces a number of European Directives and Regulations with a single system.

REACH has several aims:

- To provide a high level of protection of human health and the environment from the use of chemicals.
- To make the people who place chemicals on the market (manufacturers and importers) responsible for understanding and managing the risks associated with their use.
- To allow the free movement of substances on the EU market.
- To enhance innovation in and the competitiveness of the EU chemicals industry.
- To promote the use of alternative methods for the assessment of the hazardous properties of substances, e.g., quantitative structure-activity relationships (QSAR).

Details of mutagenicity testing are contained within Section R7.7 of the Guidance on information requirements and chemical safety assessment: Chapter R.7a: Endpoint specific guidance (http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08).

REFERENCES FOR APPENDIX I


Follow-up to In Vitro Genotoxicity 191


APPENDIX II: REVIEW OF STRENGTHS AND WEAKNESSES OF GENOTOXICITY ASSAYS

As one of the main objectives of this effort, the Review Subgroup evaluated the existing assays (i.e., those mainly described in regulatory guidelines and/or in late validation phase) for several factors to assist in the interpretation and relevance of the assays for weight of evidence determinations and usefulness of particular assays for follow-up testing.

The Review Subgroup evaluated the strengths and weaknesses of genotoxicity assays that are described in regulatory guidelines and/or that are relatively well-established and show no major technical issues. Several factors were considered for the evaluation such as model characterization, understanding of the end-point and underlying mechanism, validation status and reliability, scientific and regulatory recognition, potential use in the standard battery and/or as follow-up test, knowing that these factors could depend on the situation. It was acknowledged that the analysis will reflect the current reality and that it may change as new test systems or new data become available in the future.

Based on these considerations, 4 different categories were established to define the usefulness of any in vitro or in vivo tests in follow-up testing (see Table IV below).

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Well-characterized.</td>
</tr>
<tr>
<td>#2</td>
<td>May be well-characterized.</td>
</tr>
<tr>
<td>#3</td>
<td>May be less characterized than category #1 and 2 tests.</td>
</tr>
<tr>
<td>#4</td>
<td>Some may be well-characterized.</td>
</tr>
</tbody>
</table>

Based on the characteristics given in Table IV, all of the main genotoxicity assays were assigned a category. They may not fulfill all criteria for a given category, but most of them. For each category, the table also describes the potential use of the tests classified in each category. The in vitro and in vivo assays are listed in Tables V and VI, respectively in the present Appendix II. The assays are classified in three groups depending on the end-points detected; gene mutations, chromosome damage, and DNA primary damage. For each end-point, the assays are ordered based on their categorization (1 to 4); only a few category 3 assays are reported in the tables.

The assays listed in Table V and VI of this Appendix, if not yet part of the initial battery of conducted tests, could be considered as potential follow-up assays in case of in vitro positive results in the initial battery. They could be used to confirm and complement the assays conducted in the battery, and/or to answer specific questions and contribute to WOE and MOA determinations, as described in Step 2 “Weight of Evidence with Mechanism and Mode of Action (Issues) to Consider.”

It should be noted that in vitro cell transformation tests, sometimes called in vitro carcinogenicity tests, have been proposed to supplement the standard test battery to identify carcinogens that act by mechanisms other than mutagenicity or clastogenicity. The most widely used of these tests employ the Syrian Hamster Embryo (SHE) and Balb/c-3T3 cell lines, and their endpoints scored are colonies or foci of transformed cells (for review, see OECD, 2007). These assays are not addressed in this document because they would not be considered as a component of a genetic toxicology testing battery. There are no current guidelines for these tests, although international efforts are currently underway to develop such guidance.

While some OECD guidelines exist for assays conducted on yeast (Kafer et al. 1982), Scott et al. (1982), Brockman et al. (1984), and Zimmermann et al. (1984)) and Drosophila (Bentley et al. 1994), and they may be useful research models, better alternatives exist for risk assessment (i.e., mammalian cell assays). Those assays were therefore not described in this appendix.

### TABLE IV. Categories of assays

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Well-characterized.</td>
</tr>
<tr>
<td>#2</td>
<td>May be well-characterized.</td>
</tr>
<tr>
<td>#3</td>
<td>May be less characterized than category #1 and 2 tests.</td>
</tr>
<tr>
<td>#4</td>
<td>Some may be well-characterized.</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Gene mutations</strong></td>
<td></td>
</tr>
<tr>
<td>Bacterial Reverse Mutation Assay in <em>Salmonella typhimurium</em> and <em>Escherichia coli</em></td>
<td>1</td>
</tr>
<tr>
<td><em>In Vitro</em> Gene Mutation Assay in Mammalian Cells (excluding mouse lymphoma assay)</td>
<td>1</td>
</tr>
<tr>
<td><em>In Vitro</em> Mouse Lymphoma L5178Y tk&lt;sup&gt;−/−&lt;/sup&gt; Assay</td>
<td>1</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Chromosome damage</strong></td>
<td></td>
</tr>
<tr>
<td>In Vitro Micronucleus Assay in Mammalian Cells</td>
<td>1</td>
</tr>
<tr>
<td>In Vitro Chromosome Aberration Assay in Mammalian Cells</td>
<td>1</td>
</tr>
<tr>
<td>In vitro Mouse Lymphoma L5178Y tk^- Assay</td>
<td></td>
</tr>
<tr>
<td><strong>DNA primary damage</strong></td>
<td></td>
</tr>
<tr>
<td>In Vitro Comet Assay in Mammalian Cells</td>
<td>2</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>In Vitro DNA Adducts in Mammalian Cells</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial DNA Damage or Repair Assay</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>In Vitro Alkaline Elution in Mammalian Cells</td>
<td>4</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>In Vitro</strong></td>
<td>Mammalian</td>
</tr>
<tr>
<td><strong>In Vitro</strong></td>
<td>Sister Chromatid Exchange Assay in Mammalian Cells</td>
</tr>
</tbody>
</table>

**TABLE VI. Relevance of Commonly Employed and Recently Developed In Vivo Genetic Toxicity Assays**

<table>
<thead>
<tr>
<th>Type of assay</th>
<th>Category</th>
<th>OECD Guideline</th>
<th>Endpoint(s)</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Opportunities</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Mutation Assays in Transgenic Models</td>
<td>2</td>
<td>No guideline</td>
<td>Gene mutations (point mutations including base pair substitutions and frameshift mutations) in mammals in vivo. Reporter genes (e.g., lacZ, lacI, gpt) in shuttle vectors (e.g., lambda phage). Some models (e.g., api, plasmid) also have the ability to detect deletions.</td>
<td>Can be applied to any tissue. Relevant end-point: gene specific. No selective pressure on mutations, therefore accumulation of damage over time. Uses a small number of animals.</td>
<td>Labor intensive and expensive. Requires multiple dosing. Requires transgenic animals. Need to optimize protocols for different tissues, or to apply the recommended design (28 treatment days, sampling after 3 and/or 28-day recovery period). Mutamouse, Big Blue and gpt delta models do not detect large deletions. Relatively high mutant frequency background shown to impact the sensitivity.</td>
<td>Mutant sequencing for mechanistic information (mutational spectrum) and confirmation of mutation (increase in mutant frequency versus clonal effect). Quantitation of dose response possible.</td>
<td>Heddle et al. (2000), Thybaud et al. (2003), Lambert et al. (2005), OECD (2009)</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
<td>OECD Guideline</td>
<td>Endpoint(s)</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Opportunities</td>
<td>References</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Expanded Simple Tandem Repeat (ESTR) Assay</td>
<td>3</td>
<td>No guideline</td>
<td>Non-coding tandem repeat DNA mutation in mammals in vivo.</td>
<td>Conducted in germ cells. Potentially relevant end-point. Detects heritable mutations at ambient exposure levels. Uses relatively few animals.</td>
<td>Seldom used. Some tandem repeat mutations also occur in, or near, coding genes. Although there are parallels with mutations in coding genes, the human health outcomes require further study.</td>
<td>Very specific follow up. Can be conducted in humans.</td>
<td>Dubrova et al. (1998), Yauk (2004), Singer et al. (2006), Gomes-Pereira and Monckton (2006)</td>
</tr>
<tr>
<td>Mouse Spot Test</td>
<td>4</td>
<td>OECD 484</td>
<td>Gene mutations (point mutations including base pair substitutions and frameshift mutations) in mammals in vivo.</td>
<td>None identified by the Subgroup.</td>
<td>Seldom used. Requires large number of animals.</td>
<td>None identified by the Subgroup.</td>
<td>Russell et al. (1981a)</td>
</tr>
<tr>
<td>Mouse Biochemical Specific Locus Test</td>
<td>4</td>
<td>No guideline</td>
<td>Gene mutations (point mutations including base pair substitutions and frameshift mutations) in mammals in vivo.</td>
<td>Relevant end-point in germ cells. Provides data for quantitation of inherited mutation frequency.</td>
<td>Seldom used. Requires large number of animals.</td>
<td>Very specific follow up. Quantitation of dose response possible.</td>
<td>Johnson et al. (1981)</td>
</tr>
<tr>
<td>Mouse Visible Specific Locus Test</td>
<td>4</td>
<td>No guideline</td>
<td>Gene mutations (point mutations including base pair substitutions and frameshift mutations) in mammals in vivo.</td>
<td>Relevant end-point in germ cells. Provides data for quantitation of inherited mutation frequency.</td>
<td>Requires large number of animals.</td>
<td>Very specific follow up. Quantitation of dose response possible.</td>
<td>Russell et al. (1981b)</td>
</tr>
<tr>
<td>Mammalian Bone Marrow Chromosomal Aberration Test</td>
<td>1</td>
<td>OECD 475</td>
<td>Structural and numerical chromosome damage (i.e., clastogenicity and polyploidy) in vivo in somatic cells.</td>
<td>Long history of use. Regulatory acceptance. Relevant end-point in somatic cells. Detects polyploidy inducers, including some aneugens. Applicable in wild-type animals.</td>
<td>Lack of sensitivity, does not detect all rodent carcinogens, e.g., might not detect organ specific compounds and unstable compounds/metabolites. Resource intensive and time consuming scoring of chromosome aberrations. Requires skilled scientists.</td>
<td>Methodology potentially applicable in other tissues. FISH/chromosome painting can provide additional mechanistic information.</td>
<td>Tice et al. (1994)</td>
</tr>
<tr>
<td>Mammalian Erythrocyte Micronucleus Test</td>
<td>1</td>
<td>OECD 474</td>
<td>Structural and numerical chromosome damage (i.e., clastogenicity and aneuploidy) in vivo in somatic cells.</td>
<td>Long history of use. Regulatory acceptance. Relevant end-point in somatic cells. Detects both clastogens and aneugens. Applicable in wild-type animals.</td>
<td>Resource intensive and time consuming scoring of micronuclei. Lack of sensitivity, does not detect all rodent carcinogens, e.g., might not detect organ specific compounds and unstable compounds/metabolites. Possible confounding effects (e.g., hypo- and hyperthermia, impact on erythropoiesis).</td>
<td>Methodology potentially applicable in other tissues. Scoring automation (e.g., flow cytometry, image analysis).</td>
<td>Hayashi et al. (1994, 2000, 2007), Tweats et al. (2007)</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
<td>OECD Guideline</td>
<td>Endpoint(s)</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Opportunities</td>
<td>References</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Mammalian Spermatogonial Chromosomal Aberration Test</td>
<td>2</td>
<td>OECD 483</td>
<td>Structural and numerical chromosome damage (i.e., clastogenicity and polyploidy) in vivo in mammalian testicular cells.</td>
<td>Long history of use. Regulatory acceptance. Relevant end-point in germ cells. Detects both clastogens and aneugens. Applicable in wild-type animals. One of few currently available assays in germ cells.</td>
<td>Resource intensive and time consuming scoring of chromosome aberrations. Seldom used.</td>
<td>None identified by the Subgroup.</td>
<td>Adler et al. (1994)</td>
</tr>
<tr>
<td>Rodent Dominant Lethal Assay</td>
<td>2</td>
<td>OECD 478</td>
<td>Generally accepted to be the result of chromosomal damage (structural and numerical anomalies) but gene mutations and toxic effects cannot be excluded.</td>
<td>Relevant end-point in germ cells.</td>
<td>Seldom used. Requires large number of animals.</td>
<td>None identified by the Subgroup.</td>
<td>Green et al. (1985)</td>
</tr>
<tr>
<td><em>In Vivo</em> Mammalian Comet Assay</td>
<td>2</td>
<td>No guideline</td>
<td>DNA single- and double-strand breaks, alkali-labile lesions, incomplete DNA repair sites and DNA crosslinking.</td>
<td>Rapid and easy to conduct. Can be applied to any tissue. Cell division not required, and analysis made in individual cells. Low number of cells required. Can be automated.</td>
<td>Indicator test detecting premutagenic lesions. Resource intensive and time consuming scoring of comets. May be unable to detect mutagens that do not produce strand breaks or alkali-labile lesions. Apoptosis/necrosis as indicators of tissue cytotoxicity, needs to be controlled (histopathology). Interlaboratory variability and lack of guidelines.</td>
<td>Can be used for mechanistic studies. Elucidate a genotoxic mechanism; analyze possible target organ-specific genotoxicity; investigate the in vivo relevance to positive in vitro genotoxicity; can be combined with in vivo MN assay.</td>
<td>Singh et al. (1988), Tice et al. (2000), Sasaki et al. (2000), Hartmann et al. (2003), Brendler-Schwaab et al 2005, Burlinson et al. (2007)</td>
</tr>
<tr>
<td>Type of assay</td>
<td>Category</td>
<td>OECD Guideline</td>
<td>Endpoint(s)</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Opportunities</td>
<td>References</td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>In vivo DNA Adducts in Mammalian Cells (different methods)</td>
<td>2</td>
<td>No guideline</td>
<td>DNA adducts</td>
<td>Sensitivity (ability to detect the different types of adduct) and specificity (ability to define the chemical structure of the adduct) depends on the method used. Can be highly sensitive ($^{32}$P-postlabelling or AMS) or chemically specific (MS). Other methods include immunochemical techniques, fluorescence, ECD (for 8-OH-dG).</td>
<td>Indicator test detecting premutagenic lesions. Generally technically difficult. High quantity of DNA required for some methods. Interpretation of low levels of DNA adducts, and definition of positive/negative results is difficult. Various methods available. Sensitivity (ability to detect the different types of adducts) and specificity (ability to define the chemical structure of the adduct) depends on the method used.</td>
<td>DNA structure identification for mechanistic purpose (mass spectrometry).</td>
<td>Phillips et al. (2000)</td>
</tr>
<tr>
<td>In vivo Sister Chromatid Exchange Assay</td>
<td>4</td>
<td>No guideline</td>
<td>Inter-chromatid exchange.</td>
<td>Long history of use. Useful for some classes of chemicals.</td>
<td>Resource intensive and time consuming scoring of SCE. Mechanism not fully understood. Might be due to mechanisms other than genotoxicity (e.g., inhibition of cell proliferation).</td>
<td>None identified by the Subgroup.</td>
<td>Tucker et al. (1993)</td>
</tr>
<tr>
<td>In vivo Alkaline elution</td>
<td>4</td>
<td>No guideline</td>
<td>DNA strand breaks in mammals in vivo (see in vitro section).</td>
<td>Can be applied to many tissues.</td>
<td>Does not examine individual cells, but rather looks at populations of cells (might result in lack of sensitivity, because of dilution effect).</td>
<td>None identified by the Subgroup.</td>
<td>USEPA (1987)</td>
</tr>
</tbody>
</table>
REFERENCES


Environmental and Molecular Mutagenesis. DOI 10.1002/em

202 Dearfield et al.


Accepted by—
S. Galloway