

6.0 RECOMMENDATIONS

In the previous chapters of this review, we have examined all aspects of TGR assays and their use in the prediction of genotoxicity and carcinogenicity. In this chapter, we provide recommendations for the conduct of a TGR assay in a regulatory context and also discuss how the assays would be used in a test battery. In addition, it is important to consider what experiments would enhance our understanding of TGR assays and thus increase confidence in the conclusions derived from TGR experiments, both by themselves and in the context of a test battery.

6.1 Recommendations for the conduct of TGR assays

The following recommendations for the conduct of TGR assays are based on the analysis of the data shown in Chapter 4 and the consensus of two meetings that were convened in Washington, DC (1999) and Plymouth, UK (2002) to provide internationally harmonised guidelines for the conduct of transgenic mutation assays for the purpose of regulatory assessment of safety. More detailed rationales for the recommendations are included in published manuscripts derived from these discussions (Heddle *et al.*, 2000; Thybaud *et al.*, 2003) and references therein.

6.1.1 Accepted characteristics of a TGR mutation assay

6.1.1.1 Criteria for inclusion of an assay

In those assays described in Section 2.2, the target gene is bacterial, and the means of recovery is by incorporation of the target gene in a λ phage or plasmid shuttle vector. The procedure involves the extraction of genomic DNA from the tissue of interest in the rodent, *in vitro* processing (packaging of λ vectors, or restriction/ligation of plasmids) of the genomic DNA to recover the shuttle vector and detection of mutations in bacteria under suitable conditions. Assays to be considered are those that are based on a neutral transgene that is readily recoverable from most tissues. The assay must be generally available for use and be used by several laboratories as demonstrated by published work so that it is clear that the assay can be transferred from the laboratory of origin successfully. It is accepted that the *lacI*, *lacZ* (lambda and plasmid), *cII*, *gpt* delta (*gpt*) and *gpt* delta (Spi) assays fall into this category as performed under standard conditions.

6.1.1.2 Limitations of test characteristics of the transgenes

The goal of transgenic systems is to emulate the mutagenic response of endogenous regions of the genome by measuring the response in a surrogate bacterial transgene. A key assumption in this approach is that the bacterial target gene reflects most of the important parameters affecting mutation burden at native loci. It is acknowledged that every assay, whether transgenic or not, has its own spectrum of detectable mutations, and, accordingly, some differences are expected. Bacterial transgenes possess attributes that differ from those of most mammalian genes. These include higher GC content; higher density of the dinucleotide CpG and associated 5-methylcytosine; a multicopy, head-to-tail concatemer structure that leads to hypermethylation; and, because they are neutral genes, a lack of transcription and the associated transcription-coupled repair.

6.1.1.3 Types of mutations detected by transgenic systems

The mutations scored in the *lacI* and *lacZ* transgenic systems consist primarily of base pair substitution mutations with a few frameshift mutations and small insertions/deletions. The relative proportion of these mutation types among spontaneous mutations is similar to that seen at the native *Hprt* gene. Large deletions are not readily detected/recognised with

these assays, but are detectable with the Spi and plasmid assays. Mutations of interest are *in vivo* mutations that arise in the mouse; *in vitro* and *ex vivo* mutations, which arise during phage/plasmid replication or repair, are uncommon.

6.1.1.4 Sensitivity of transgenic mutation assays

The sensitivity of a mutation assay is defined to a large extent by the magnitude of induced mutational response compared with the magnitude of background levels. Existing evidence suggests that the level of induced point mutations in transgenic targets and endogenous genes occurs at similar frequencies following acute treatment, but that the spontaneous mutation frequency in transgenes appears to be somewhat higher than has been previously observed in endogenous targets for a limited number of tissues. This higher background mutation frequency may make it more difficult to achieve a significant induced response with certain weak mutagens.

6.1.2 Treatment protocols

6.1.2.1 Justification

The justification for the number of treatments used, the time of sampling and the tissues sampled must be included in the description of the protocol used. All available data on the toxicokinetics of the test substance, such as absorption, distribution, metabolism and excretion, should be used in the study design, together with any information about mitogenic activity.

6.1.2.2 Selection of species

A variety of transgenic mouse models are currently available, and these systems have been more widely used than transgenic rats. If a rat is clearly a more appropriate model than a mouse (*e.g.* when investigating the mechanism of carcinogenesis for a tumour seen only in rats), the use of available transgenic rat models should be considered.

6.1.2.3 Selection of sex

Male animals should normally be used, consistent with guideline recommendations for other *in vivo* genotoxicity tests. However, if there are significant differences between the sexes in terms of toxicity or metabolism, then both males and females will be required. There may be cases where females alone would be justified – for example, when testing human sex-specific drugs, or in the case of sex-specific metabolism. These recommendations are applicable to the rat as well as to the mouse.

6.1.2.4 Number and size of treatment groups

Assays should use groups of 5–10 animals. A full set of data must be generated from a negative control group and a minimum of two dose levels. The top dose should be the MTD. The other doses should be one-third of the MTD and two-thirds of the MTD. If all of the three dose groups are complete, analysis of only the top and second dose would be sufficient, although the samples from the lowest dose should be retained for possible further analysis. If, however, enough animals die in the top dose group such that statistical power is reduced below an acceptable level, this dose group would not be analysed, and the two lower doses would be sufficient to define a negative result. In this latter case, delay in analysis of the samples from the lowest dose could compromise sample blocking (see Section 6.1.3.4); if blocking is to be maintained, it would be appropriate to analyse all samples at the same time.

6.1.2.5 Duration of treatment

Based on observations that mutations accumulate with each treatment, a repeated-dose regimen is strongly encouraged, with daily treatments for a period of 28 days generally considered adequate both for producing a sufficient accumulation of mutations by weak mutagens and for providing a sampling time adequate for detecting mutations in slowly proliferating organs. Alternative treatment regimens, such as weekly dose administration, may be appropriate for some evaluations, and these alternative dosing schedules should be justified in the protocol. Treatments should not be shorter than the time required for the complete induction of all of the relevant metabolising enzymes, and shorter treatments may necessitate the use of multiple sampling times that are suitable for organs with different proliferation rates. Treatment times longer than 8 weeks should be employed with caution, since long treatment times are known to produce an apparent increase in mutant frequency through clonal expansion.

6.1.2.6 Positive control

For laboratories that have demonstrated competence with these assays, concurrent positive control animals are not normally necessary; however, it is recommended that positive control DNA be included with each plating to confirm the success of the method. It is recommended that laboratories new to these test systems include concurrent positive controls during validation.

6.1.3 Post-treatment sampling

6.1.3.1 Sampling time

The time between the last treatment and the time of sampling, the sampling time, is a critical variable. The time required to reach the maximum mutant frequency is tissue specific and seems to be related to the turnover time of the cell population, with bone marrow and intestine being rapid responders and the liver being much slower. Following 28 consecutive daily treatments (as recommended in Section 6.1.2.5), sampling at 3 days following the final treatment should be suitable for both rapidly and slowly proliferating tissues, although the maximum mutation frequency may not manifest itself in slowly proliferating tissues under these conditions. If slowly proliferating tissues are of particular importance, then a longer sampling time (*e.g.* 28 days) may be more appropriate. In the case of germ cells where the kinetics are well defined, the sampling time should be selected according to the stage of interest.

6.1.3.2 Rationale for tissue selection

In TGR assays, it is possible to use virtually any route of administration and to sample any tissue. Therefore, the selection of tissues to be sampled should be based upon the reason for conducting the study and any existing mutagenicity, carcinogenicity or toxicity data for the compound under investigation. Important factors for consideration should include the route of administration, the likely tissue distribution, the possible mechanism of action or the likely human exposure to the compound. In the absence of any background information, at least one rapidly dividing (*e.g.* bone marrow) and one slowly dividing tissue (*e.g.* liver) should be evaluated. If a compound is negative in bone marrow and liver, a third tissue should be evaluated. The choice of tissue would be based on the route of administration: for example, small intestine if administration is oral, lung if the administration is through inhalation or skin if topical application has been used. This third tissue would allow the evaluation of compounds that are direct-acting *in vitro* mutagens, rapidly metabolised, highly reactive or poorly absorbed, or those for which the target tissue is determined by route of administration (Dean *et al.*, 1999). The rationale for tissue selection should be made clear.

6.1.3.3 Storage of tissues

Tissues should be stored at or below $-70\text{ }^{\circ}\text{C}$ and may be stored under these conditions for several years. Isolated DNA, stored refrigerated in appropriate buffer, should be used for mutation analysis within 1 year, but may generate useful data if stored longer than this.

6.1.3.4 Methods of measurement

Standard laboratory or published methods for the detection of mutants are available for the recommended transgenic models (Vijg and Douglas, 1996; Nohmi, Suzuki and Masumura, 2000). Modifications should be justified and properly documented. There is no biological justification to set a minimum acceptable number of plaque-forming units or colony-forming units from an individual packaging: all data can be used and aggregated. Tissues should be processed and analysed using a block design, where samples from the negative control group, the positive control group and each treatment group are processed together.

6.1.3.5 Requirements for reporting

Reporting of a regulatory study should be as defined for all Good Laboratory Practice studies. The report should include the total number of plaque-forming units or colony-forming units and the mutant frequency for each tissue and for each animal. Data for individual packaging should be retained, but need not be reported.

6.1.3.6 Statistics

The application of statistics to *in vivo* transgenic mutation assays is consistent with previously reported statistical approaches for *in vivo* genotoxicity studies (Bielas and Heddle, 2000), with specific modifications. Pairwise analysis is appropriate for one dose, and a test for a dose–response is appropriate if two or more doses were used. Non-parametric statistical tests such as the generalised Cochran-Armitage trend test allow analysis of variable data such as those typically obtained with these assays. Statistical tests used should consider the animal as the experimental unit. A positive result is one in which the data for one or more tissues show a statistically significant dose–response relationship or a statistically significant increase in any dose group as compared with concurrent negative controls using an appropriate statistical model. A negative result is one that is not statistically significant and in which the mean mutant frequencies at all doses for at least three tissues (see Section 6.1.3.2) are within two standard deviations of the mean mutant frequency in the control.

6.1.3.7 Sequencing of mutants

When testing drugs or chemicals for regulatory applications, the sequencing of mutants is not normally required, particularly where a clear positive or negative result is obtained. Sequencing data may be useful when high interindividual variation is observed. In these cases, sequencing can be used to rule out the possibility of jackpots or clonal events by identifying the proportion of unique mutants from a particular tissue. Sequencing up to 10 mutants per tissue should be sufficient for simply identifying clonal mutants; sequencing as many as 25 mutants may be necessary for correcting mutant frequency mathematically for clonality. When sequencing is to be included as part of the study protocol, special care should be taken in the design of the sequencing component, in particular with respect to the number of mutants sequenced per sample.

6.2 Recommendations for further research regarding test protocol

6.2.1 *Time of administration*

What is the influence of the duration of treatment on the observed mutation frequency for weak mutagens?

The data presented in Chapters 4 and 5 were heavily weighted by strong mutagens; in these cases, it appeared that a treatment time as short as 14 days (sometimes even shorter) was sufficient to detect a significant mutagenic response. However, the consensus recommendations specify a treatment duration of 28 days, based on extrapolations of data from the TRAIT, limited direct application with weak mutagens and theoretical considerations, and because this treatment time is commonly used in toxicological testing. It has not been determined conclusively if data (especially negative results) from experiments using an administration time of less than 28 days should be discounted, if a 28-day treatment period is sufficiently long to permit the detection of weak mutagen-induced mutations in all tissues or if any weak mutagens could in fact be detected using treatment times shorter than 28 days. This is particularly important because the majority of mutagenic chemicals in the environment are likely to be weak mutagens. Additional confidence in the recommendation described in Section 6.1.2.5 would be provided by systematic studies using weak mutagens in which the time of administration is varied. To date, there have been two studies that have confirmed that the recommended consensus protocol is effective for detecting weak mutagens (acrylamide: Thybaud *et al.*, 2003; urethane: Singer, 2006).

6.2.2 *Frequency of treatment*

What is the influence of the frequency of treatment? That is, is a weekly dosing regimen (four weekly doses) equivalent to a daily dosing regimen?

Some laboratories have favoured weekly, rather than daily, administrations. The difference between weekly and daily administrations in terms of their effect on mutation frequency and on the ultimate conclusions of TGR experiments has not yet been thoroughly investigated. Confidence regarding this question would be increased by experiments in which the same total dose was administered over 28 days but using different frequency of administration. These experiments should be done using weak mutagens, since there is ample evidence that single treatments of strong mutagens will yield positive results.

6.2.3 *Sampling time*

Would a 3-day sampling time be sufficient to detect a significant increase in mutation frequency in both slowly and rapidly dividing tissues after administration of weak mutagens?

There have been few experiments examining the effects of sampling time on mutation frequency, and most of those experiments that have been conducted used strong mutagens such as ENU with single doses or small numbers of repeated doses. There is a need to carry out experiments with other mutagens (especially weak mutagens). In addition, there are limited data in the literature describing the optimal sampling time following 28 consecutive daily treatments. It will be particularly important to evaluate mutagenesis in different tissues using this protocol. It should be noted that the IWGT proposed the “28 (administration) + 3 (sampling)” protocol as a single sampling time that is intended to accommodate both rapidly and slowly proliferating tissues. An alternative protocol that may be appropriate in some circumstances would be to add an additional group of animals per dose, allowing for both short and long sampling periods. At the current time, there are not sufficient comparative data to rule out either protocol.

6.3 Use of TGR assays as a component of genotoxicity test batteries

An important consideration when selecting tests for inclusion in a genotoxicity test battery is the degree to which predictivity for the endpoint in question (mutagenicity or carcinogenicity) is improved by combining the tests, rather than using the tests alone. In the context of genotoxicity test battery interpretation, *in vivo* assays are typically given more weight than *in vitro* assays, such that *in vivo* assays are often used to confirm or discount the results of *Salmonella* and *in vitro* chromosomal aberration assays.

6.3.1 Test battery approaches: Mutagenicity per se vs. prediction of carcinogenicity

In the conduct of mutagenicity test batteries, the primary objective should be to correctly identify agents that are mutagens and those that are non-mutagens (*i.e.* genotoxins and non-genotoxins). While agents so identified are potentially germ cell mutagens, this approach will also provide the best opportunity to identify potential genotoxic carcinogens. This approach differs, in principle, from approaches that attempt to select mutagenicity test batteries based primarily on their statistical predictivity for carcinogenicity, rather than their optimal ability to detect mutagenicity (genotoxicity) *per se*. Because the former approach is not based on extrapolation across endpoints (*i.e.* mutagenicity tests detecting mutagenicity), sensitivity and positive predictive value are the primary considerations; specificity and negative predictive value are of relatively lower importance. In contrast, the latter approach, which requires extrapolation across different endpoints (*i.e.* mutagenicity predicting carcinogenicity), is more dependent on issues relating to specificity and the prevalence of carcinogens and, accordingly, is inherently more difficult to validate.

6.3.2 Conclusions based on analysis of existing TGR data

The following conclusions can be drawn from Chapter 5:

- 1) As shown in Table 5-10, TGR was usually positive for those mutagens that were positive in *Salmonella* and *in vitro* chromosomal aberration assays (0.84, 54/64). In contrast (Table 5-11), the *in vivo* micronucleus assay had a lower predictivity for mutagens that are positive in both *in vitro* tests (0.78, 38/49). If *in vivo* confirmation of positive results from both *Salmonella* and *in vitro* chromosomal aberration is warranted, the TGR assay may be a better choice than the *in vivo* micronucleus assay, but any difference is marginal.
- 2) For chemicals having positive *Salmonella* and negative *in vitro* chromosomal aberration results (presumptive gene mutagens), selecting either the TGR assay (Table 5-10) or the *in vivo* micronucleus assay (Table 5-11) as the *in vivo* confirmation assay did not markedly affect the proportion of correct carcinogenicity predictions; however, the numbers of chemicals in this category are very small (four and six, respectively), providing little in the way of precision.
- 3) For chemicals having positive *in vitro* chromosomal aberration and negative *Salmonella* results (presumptive clastogens), selecting *in vivo* micronucleus (Table 5-11) as the *in vivo* confirmation assay led to a markedly higher proportion of correct *in vitro* mutagenicity predictions than did selecting the TGR assay (Table 5-10) (micronucleus: 0.57, 4/7; TGR: 0.23, 3/13), although the numbers of chemicals tested are very small.
- 4) For those chemicals with negative results in both *Salmonella* and *in vitro* chromosomal aberration, the micronucleus assay was only marginally better at predicting the combined results of the *in vitro* battery (micronucleus: 0.90, 19/21; TGR: 0.82, 18/22).

6.3.3 Possible strategies for test battery interpretation

There are potentially two uses for TGR assays in a genotoxicity testing strategy. A primary use of the assay could be for confirming or refuting *Salmonella* gene mutation results using an *in vivo* test system. Alternatively, TGR assays could be used as the first *in vivo* assay in a test battery. In cases where the results of *in vitro* testing indicate a greater potential for gene mutations than chromosomal aberrations (*i.e.* mutagenic to *Salmonella*, non-mutagenic to *in vitro* chromosomal aberration), a TGR assay could be substituted for the *in vivo* micronucleus assay to identify whether the chemical causes gene mutations *in vivo*. These strategies are included simply to illustrate the range of possible ways in which TGR assays could be used in a regulatory context; they are not meant as prescribed methods or recommendations to national regulatory authorities.

6.3.3.1 A new test battery interpretation framework – the Selective Replacement Model

A proposed new strategy for the use of a short-term test battery in identifying chemicals with mutagenic potential is presented in Figure 6-1. The test battery consists of various combinations of four assays – *Salmonella*, *in vitro* chromosomal aberration, *in vivo* micronucleus and TGR. It is assumed that the *Salmonella* and *in vitro* chromosomal aberration assays would be a standard component of any test battery.

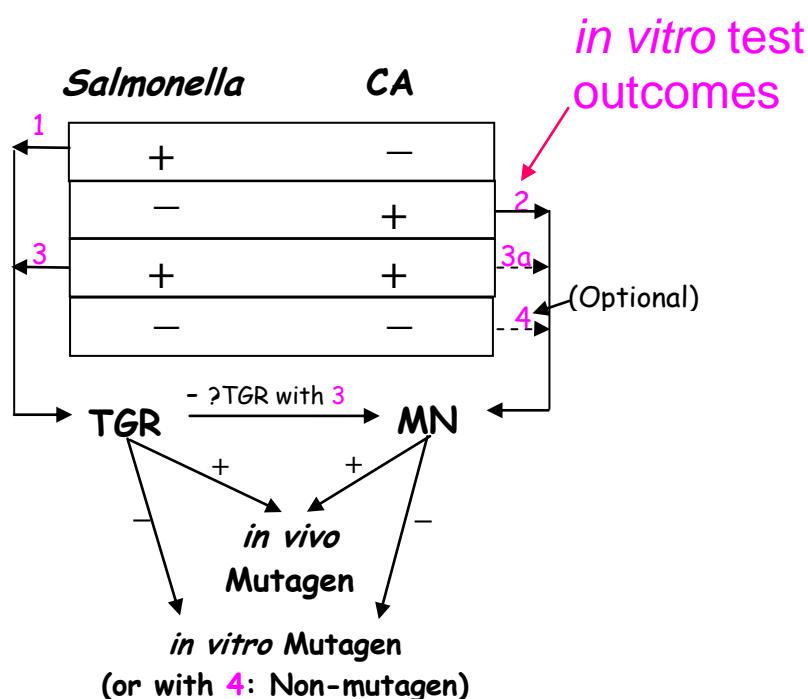


Figure 6-1. Possible mutagenicity test strategy in which transgenic rodent gene mutation assays serve as a primary *in vivo* test (CA, chromosomal aberration assay; MN, micronucleus assay)

For chemicals with positive results in both *in vitro* assays, the first *in vivo* test conducted would be the TGR assay (Figure 6-1, *in vitro* test Outcome 3, rather than Outcome 3a), because it was found to have a slightly lower false-negative rate for such chemicals compared with the *in vivo* micronucleus assay (Section 6.3.2). This suggests that the probability of needing both *in vivo* assays to identify a true *in vivo* mutagen would be lower when the TGR assay is the first *in vivo* assay than when the micronucleus assay is the first *in vivo* assay. If the TGR assay is negative with Outcome 3, an *in vivo* micronucleus assay may be

required as an option to ensure that the chemical is a *bona fide* negative mutagen *in vivo*. If both *in vivo* assays are negative, the chemical is concluded to be an *in vitro* mutagen only.

Also, in this strategy, chemicals that have a positive result in only one of the *in vitro* assays could possibly proceed to further testing using the *in vivo* assay examining the same genetic endpoint (*i.e.* same mode of action). For example, chemicals positive only in *Salmonella* (Outcome 1) would be tested using the TGR assay, whereas chemicals with a positive result only in the *in vitro* chromosomal aberration assay (Outcome 2) would be tested using the *in vivo* micronucleus assay. A negative result in the *in vivo* assay selected would lead to a conclusion that the chemical is an *in vitro* mutagen only.

This approach embraces the well-established concept that some chemicals exhibit a preference for inducing either gene mutations or chromosomal aberrations, which should be considered in the selection of tests in test batteries. As discussed in Chapter 5, it follows logically that test batteries selected to be the most indicative of the range of genotoxic endpoints should be the test combinations that can best detect genotoxic carcinogens.

Chemicals with negative results in both *Salmonella* and *in vitro* chromosomal aberration assays have a low probability of being mutagenic. Because the *in vivo* micronucleus assay is less costly and requires less time than the TGR assay, this assay could be selected as the *in vivo* confirmatory assay in this situation. A negative *in vivo* micronucleus result would lead to the conclusion that the chemical is not mutagenic.

While this scenario may be more financially costly than the strategy described below (Figure 6-2), it has the potential to use fewer animals to arrive at the final conclusion regarding *in vivo* mutagenicity.

6.3.3.2 Use of TGR assays as an adjunct in existing test batteries – the Addition Model

TGR assays may also find uses in resolving conflicts between *in vitro* and *in vivo* tests that are currently components of the standard genotoxicity test battery – *Salmonella*, *in vitro* chromosomal aberration and *in vivo* micronucleus. In situations where the standard test battery has been conducted and there are conflicting results – particularly in situations where *Salmonella* has a positive result but *in vivo* micronucleus is negative – the TGR assay may be conducted as an additional test to resolve the conflict.

According to this model, chemicals that have at least one positive result in the two *in vitro* tests (Outcome 1, 2 or 3 in Figure 6-2) would proceed, as has become standard practice in most test strategies, to *in vivo* testing using *in vivo* micronucleus. If the *in vivo* micronucleus test was positive, the chemical would be concluded to be an *in vivo* mutagen, but if the *in vivo* micronucleus test was negative, there would be a conflict between the results of this assay and the *Salmonella* assay. Since there is the potential for a chemical to preferentially induce gene mutations, a negative *in vivo* micronucleus assay should not be used to refute a positive *Salmonella* result. The use of the TGR assay as an additional confirmatory *in vivo* assay would allow determination of whether the chemical also induces gene mutations *in vivo*.

The utility of this confirmatory approach is demonstrated in the following example. There were 12 cases identified in the TRAIID where chemicals had positive results in both *Salmonella* and *in vitro* chromosomal aberration, but negative *in vivo* micronucleus results (Table 6-1). Since it is inappropriate to use the *in vivo* micronucleus assay to discount *Salmonella* results because these tests assess mechanistically distinct endpoints, an accurate decision regarding the genotoxicity of the chemical cannot be made. The addition of the TGR assay to the test battery for the confirmation of *Salmonella* results correctly predicted the results of the *in vitro* tests (and carcinogenicity) in 9/12 cases. Notably, the *in vivo* micronucleus assay provided no predictive value in these cases, either because of a false-negative genotoxicity result or because these compounds were in fact not clastogenic *in vivo*. It is

recognised that the TGR assays have the advantage of not being limited to the bone marrow or blood, as is the case with the micronucleus assay; accordingly, some of the positive TGR results may reflect tissue specificity rather than endpoint specificity. Nevertheless, the value of the TGR assays in resolving *in vivo* micronucleus test results that do not agree with the outcome of *in vitro* testing is quite obvious.

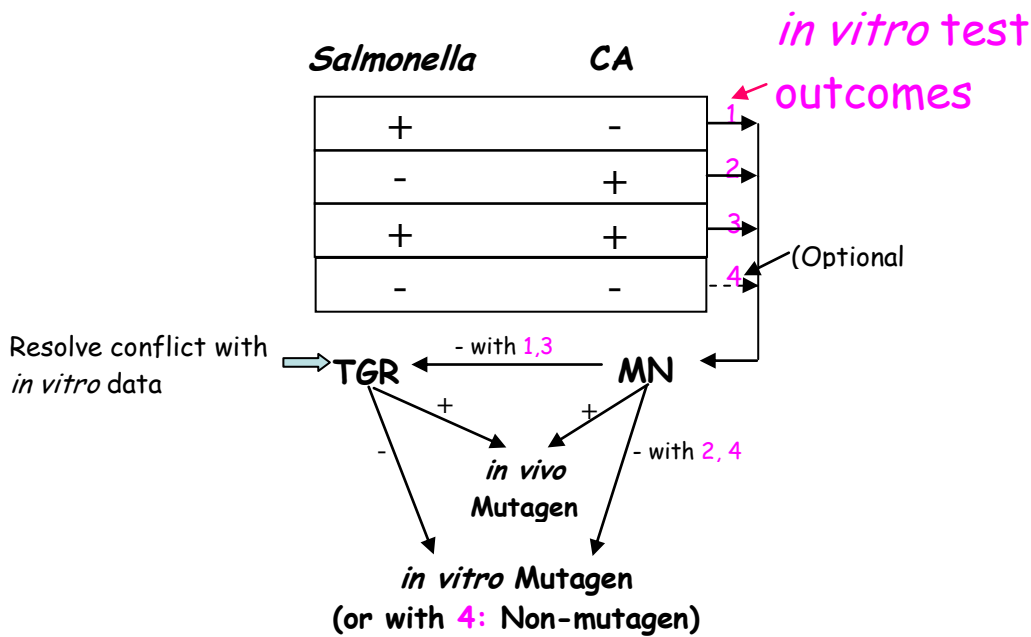


Figure 6-2. Possible mutagenicity test strategy in which transgenic rodent gene mutation assays act as a secondary *in vivo* test (CA, chromosomal aberration assay; MN, micronucleus assay)

Table 6-1. Chemicals for which the standard *Salmonella* + *in vitro* chromosomal aberration, *in vivo* micronucleus test battery did not provide a clear conclusion

Chemical	<i>Salmonella</i>	<i>In vitro</i> CA	<i>In vivo</i> MN	Conclusion	TGR	Carcinogenicity
1,2-Dibromoethane	+	+	-	?	+	+
2-Amino-3,8-dimethyl-imidazo(4,5-f)quinoxaline (MeIQx)	+	-	-	?	+	+
2-Amino-3-methyl-imidazo(4,5-f)quinoline (IQ)	+	+	-	?	+	+
2-Nitro- <i>p</i> -phenylenediamine	+	+	-	?	+	+
Acrylonitrile	+	+	-	?	-	+
Diesel exhaust	+	+	-	?	+	+
Diethylnitrosamine (DEN)	+	+	-	?	+	+
Dipropylnitrosamine (DPN)	+	+	-	?	+	+
Genistein	+	+	-	?	-	+
Metronidazole	+	+	-	?	-	+
<i>o</i> -Anisidine	+	+	-	?	+	+
<i>p</i> -Cresidine	+	+	-	?	+	+

CA, chromosomal aberration assay; MN, micronucleus assay

6.4 Recommendations for further experimentation to enhance confidence in the test battery

The above proposals are based on an analysis of the existing data. However, confidence in these proposals would be enhanced significantly by additional experimental data. Several unresolved questions warranting further experimentation are outlined in the sections below.

6.4.1 Testing non-carcinogens

Would the selection of additional non-carcinogenic chemicals for testing in the TGR assay provide data that would alter the predictive values presented in Chapter 5?

Carcinogens were unavoidably over-represented in the data presented in Chapter 5. Consequently, positive predictivity was probably overestimated and negative predictivity underestimated. However, the extent of the expected difference is not known. The addition of new TGR data for non-carcinogens to bring the proportion of non-carcinogens up to that found in the NTP database would allow a better estimation of predictive values for carcinogenicity to be made. In fact, as confirmation of this prediction, the addition of a number of non-carcinogens to the number of non-carcinogens presented in the previously published version of this review (Lambert *et al.*, 2005) did result in a significant increase in the specificity for TGR assays (as well as the other genotoxicity assays). Furthermore, based on OECD test validation criteria, the test performance should be based on the most biologically related endpoint (OECD, 2005). Accordingly, as mentioned in Section 6.3.1, specificity is a less important issue with respect to the predictivity of a new mutagenicity test for an existing assay that detects exactly the same endpoint (*e.g. in vivo* gene mutation predicting *in vivo* gene mutation).

6.4.2 Other short-term test data

Would testing to fill data gaps for the chemicals with missing data from the Salmonella, in vitro chromosomal aberration or in vivo micronucleus assays (and having known TGR assay results) alter the conclusions regarding test battery performance?

The number of chemicals having a full set of short-term genotoxicity test results was very small, and the majority of these chemicals were mutagenic in one or more test systems. If sufficient data were available to contribute to a more balanced database, it would be informative to re-examine the agreement between TGR and other short-term assays and the performance of the TGR assay in a test battery.

6.4.3 Test results of additional chemicals using a harmonised protocol (particularly with weak mutagens)

As mentioned in Chapter 5, to date, two studies have confirmed the robustness of the IWGT recommended protocol to detect weak mutagens (Thybaud *et al.*, 2003; Singer, 2006). Further studies such as these will allow for a reprise of the predictive exercise carried out in this review and confirmation of the role of the assay in a test battery

6.5 Development of an OECD Test Guideline on Transgenic Rodent Gene Mutation Assays

This extensive review fulfils the requirements for the preparation of an OECD Detailed Review Paper according to the OECD *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals* (OECD, 2006), which states that:

A DRP [Detailed Review Paper] can be developed when a specific area of hazard identification needs to be reviewed, prior to the development of a Test Guideline. A DRP is not needed if there is already an agreement on the test method to be developed for the intended purpose. When the area of concern

needs further review before a particular test method raises interest for development of a Test Guideline, the following aspects should be covered in the DRP:

- ✓ a description of the scientific progress and new techniques available in the area under review (**described in Chapters 2, 4**);
- ✓ an inventory of existing test methods in that area, together with an appreciation of, *inter alia*, the scientific validity, sensitivity, specificity and reproducibility of these methods (**Chapters 3, 5, 6**);
- ✓ an inventory of (inter)national data requirements with respect to the environmental safety and human health area under review, including those data used as part of existing hazard assessment procedures (**Chapters 3, 6**);
- ✓ identification of gaps with respect to significant endpoints not yet sufficiently covered by OECD Test Guidelines (**Chapters 3, 6**);
- ✓ identification of methods that are currently covered by OECD Test Guidelines but are to be replaced or updated in order to comply with current scientific views (**not applicable**);
- ✓ proposals with respect to the development of new Test Guidelines and/or the updating of existing ones (*see below*);
- ✓ indication of the relationship between the proposed and existing tests and of their limitations of use (**Chapters 5, 6**).

In order to bring the TGR assays into mainstream regulatory practice, it is important to establish an OECD Test Guideline on Transgenic Rodent Gene Mutation Assays. Based on the extensive information and analyses in this review, there is sufficient evidence to support the recommendation that OECD undertake the development of a Test Guideline on Transgenic Rodent Gene Mutation Assays:

- These assays fill a gap in current regulatory practices and in existing OECD Test Guidelines – namely, a test for gene mutations *in vivo*.
- They provide data comparable in quality and predictivity for carcinogenicity with those of other standard mutagenicity tests.
- Where *in vivo* tests are used or required, they provide economical strategies that can result in the use of fewer animals.
- Transgenic animal models provide the basis for the establishment of *in vitro*–equivalent assays.

Accordingly, it is recommended that OECD establish an Expert Working Group to develop such a Test Guideline and serve as an international forum for undertaking any additional research that would lead to the development of a fuller understanding of the variables surrounding the conduct of TGR mutation assays.