Mutation Research 721 (2011) 27–73

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

A core *in vitro* genotoxicity battery comprising the Ames test plus the *in vitro* micronucleus test is sufficient to detect rodent carcinogens and *in vivo* genotoxins

David Kirkland[a,*], Lesley Reeve[b], David Gatehouse[c], Philippe Vanparys[d]

[a] *Kirkland Consulting, PO Box 79, Tadcaster LS24 0AS, United Kingdom*
[b] *Covance Laboratories Limited, Otley Road, Harrogate HG3 1PY, United Kingdom*
[c] *Old Barn, Cherry Orchard Lane, Wyddial, Near Buntingford, Herts SG9 0EN, United Kingdom*
[d] *ALTOXICON BVBA, Boskant 101, B2350 Vosselaar, Belgium*

ARTICLE INFO

Article history:
Received 22 September 2010
Received in revised form
12 November 2010
Accepted 15 December 2010
Available online 14 January 2011

Keywords:
Genotoxicity
Ames test
In vitro micronucleus test
Battery of tests

ABSTRACT

In vitro genotoxicity testing needs to include tests in both bacterial and mammalian cells, and be able to detect gene mutations, chromosomal damage and aneuploidy. This may be achieved by a combination of the Ames test (detects gene mutations) and the *in vitro* micronucleus test (MNvit), since the latter detects both chromosomal aberrations and aneuploidy. In this paper we therefore present an analysis of an existing database of rodent carcinogens and a new database of *in vivo* genotoxins in terms of the *in vitro* genotoxicity tests needed to detect their *in vivo* activity. Published *in vitro* data from at least one test system (most were from the Ames test) were available for 557 carcinogens and 405 *in vivo* genotoxins. Because there are fewer publications on the MNvit than for other mammalian cell tests, and because the concordance between the MNvit and the *in vitro* chromosomal aberration (CAvit) test is so high for clastogenic activity, positive results in the CAvit test were taken as indicative of a positive result in the MNvit where there were no, or only inadequate data for the latter. Also, because *Hprt* and *Tk* loci both detect gene-mutation activity, a positive *Hprt* test was taken as indicative of a mouse-lymphoma *Tk* assay (MLA)-positive, where there were no data for the latter. Almost all of the 962 rodent carcinogens and *in vivo* genotoxins were detected by an *in vitro* battery comprising Ames + MNvit. An additional 11 carcinogens and six *in vivo* genotoxins would apparently be detected by the MLA, but many of these had not been tested in the MNvit or CAvit tests. Only four chemicals emerge as potentially being more readily detected in MLA than in Ames + MNvit – benzyl acetate, toluene, morphine and thiabendazole – and none of these are convincing cases to argue for the inclusion of the MLA in addition to Ames + MNvit. Thus, there is no convincing evidence that any genotoxic rodent carcinogens or *in vivo* genotoxins would remain undetected in an *in vitro* test battery consisting of Ames + MNvit.

1. Introduction

In 2007, a working group of the German-speaking section of the European Environmental Mutagen Society (GUM) recommended [1] that Stage-1 *in vitro* genotoxicity testing should consist of an Ames test plus an *in vitro* micronucleus test (MNvit), since these tests cover all of the essential mutagenic endpoints (gene mutations, structural chromosome damage, and aneuploidy), and moreover cover testing in both prokaryotic and eukaryotic systems. The MNvit has been retrospectively validated by ECVAM [2] and an OECD guideline (TG487) has now been adopted [3]. The main difference between the recommended two-test battery and other testing approaches is the lack of a mammalian cell gene-mutation test, most often the mouse-lymphoma assay (MLA). However, Pfuhler et al. [1] reasoned that a gene-mutation test with mammalian cells can be omitted because the bacterial gene-mutation test detects all relevant modes of action specifically leading to gene mutations. Moreover, most of the substances that are positive in mammalian gene-mutation tests also induce clastogenic effects, and these compounds would be detected with a high level of efficiency by the MNvit. A further advantage in reducing the requirement for mammalian cell testing to a single assay is the potential to reduce the frequency of "irrelevant" or "misleading" positive results [4–6], and thereby reduce unnecessary follow-up testing in animals.

The recommendations of the GUM [1] were based on sound theoretical reasoning. However, these recommendations would gain further support if confirmed by analysis of data. In this paper, therefore, we have analysed the database of carcinogens published by Kirkland et al. [4], and also compiled and analysed a new database of *in vivo* genotoxins (described herein), to determine whether an *in vitro* battery comprising Ames + MNvit is sufficiently sensitive

* Corresponding author. Tel.: +44 1937 830849.
E-mail address: dkirkland@genetoxconsulting.co.uk (D. Kirkland).

1383-5718/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2010.12.015

to detect those chemicals with published *in vivo* carcinogenic or genotoxic activity, or whether an additional mammalian cell test (the MLA) is also needed.

2. Methods

2.1. Rodent carcinogens

The *in vitro* Ames, micronucleus (MNvit) and chromosomal aberration (CAvit) test-results for carcinogens were taken from Kirkland et al. [4]. However, the mouse-lymphoma assay (MLA) results from this paper have been revised. Because the requirements for an acceptable MLA and the criteria for a positive response have been updated in recent years by various expert working groups (mainly from the International Workshops on Genotoxicity Testing, IWGT) [7–11], and because many of the published MLA results were from the National Toxicology Program (NTP) which took place prior to IWGT recommendations, the NTP MLA results have recently been re-evaluated [12]. This has led to a large number of outcomes (or "calls") changing, and a new category (uninterpretable, U) being included.

In Kirkland et al. [4], the results of MLA tests were taken, where possible, from the expert panel review of Mitchell et al. [19]. A large number of NTP MLA studies were included in that review, but because it was common practice by that time not to accept positive MLA results where test concentrations exceeded 10 mM, or where cytotoxicity exceeded 90%, a number of the NTP calls were revised by Mitchell et al. [19]. In addition, these authors identified several MLA results, including some from NTP studies, as inconclusive, not-testable, limited by solubility, limited by osmolality, limited by acidic pH shift, or limited because the chemical reacts with plastic, rapidly hydrolysed at neutral pH, or was only stable at acidic pH. These categories did not fit with the four categories used by Kirkland et al. [4]. Thus, Kirkland et al. [4] reviewed the original reports or data in order to classify these studies as positive, negative, equivocal or technically compromised (i.e. test results that were questionable due to failure to meet essential standard criteria for an adequate study). In terms of positive responses, Kirkland et al. [4] were aware that mutagenic responses seen only at RTG <10% were unreliable, and so some positive "calls" by NTP and in Mitchell et al. [19] were not accepted by Kirkland et al. [4]. However, the recommendation that, for a biologically meaningful response, the induced mutant frequency should exceed the Global Evaluation Factor (GEF) had not been published at the time of preparation of the Kirkland et al. [4] paper, and so some results accepted as positive by Kirkland et al. [4] would not be considered positive if the GEF were applied. As a result, several of the MLA "calls" in Kirkland et al. [4] were different from the original NTP "calls" and are different from "calls" made today using the latest recommendations.

For the purposes of the current analysis, therefore, the re-evaluated MLA calls have been used, and the same criteria have been applied by the current authors to non-NTP MLA studies.

It should be noted that re-evaluation of other tests, such as the *in vitro* chromosomal aberration (CAvit) test would not have the same impact on the original calls from the NTP and Kirkland et al. [4]. The protocol for the CAvit has changed since the time of the NTP studies in that later sampling times and longer treatment times are now used. This reduces the risk of "false negatives" compared to previously. However, the requirement for at least 50% cytotoxicity and the criteria for a positive response have not changed. Therefore, re-evaluation of the CAvit studies would not be likely to remove any of the positive calls, but does render some of the negative calls inadequate (see later).

2.2. In vivo genotoxins

In building the *in vivo* genotoxins database (Appendix A) we looked for positive results for the micronucleus (MN), chromosomal aberration (CA), unscheduled DNA synthesis (UDS), transgenic mutation and Comet endpoints. Whilst data on UDS were usually in liver and CA were usually in bone marrow, we included MN, transgenic mutations and comets in any tissue. Some papers demonstrated clear induction of DNA strand-breaks measured by means of the alkaline elution method. In such cases we accepted this as a surrogate for a Comet assay-positive, but these cases are identified in Appendix A. Occasionally the same data set was published in more than one paper by the same group. In these cases we have endeavoured to cite the first publication of the data, and not bias the overall response by presenting the same conclusions more than once. However, in some review papers the original source was not given, and may have related to the individual papers we found. We acknowledge that we will not have found every published positive *in vivo* result, because some specialised or local publications will not have been reviewed or abstracted by the search materials we used. However, we believe that our approach has provided an extensive database of *in vivo* genotoxins.

For an initial search we used a number of detailed review papers, e.g.:

- Mavournin et al. [13] for the *in vivo* MN test.
- Lambert et al. [14] for the transgenic mutation assay.
- Madle et al. [15] for the *in vivo* UDS test.
- Snyder [16] for marketed pharmaceuticals.

Where possible we checked the original papers cited in these review papers, but they were not always given, and in those cases we accepted the conclusions of the review author(s). We then searched for more recent papers and for data on the other endpoints through the ToxLine [http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE] and CCRIS [http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS] databases of the US National Library of Medicine, and through NTP [http://ntp-apps.niehs.nih.gov/ntp_tox/index.cfm], PubMed [http://www.ncbi.nlm.nih.gov/pubmed] and IARC [http://monographs.iarc.fr/] websites. We included all references we could find where positive, equivocal or inconclusive results were reported. It is possible that the findings in some of these papers may be disputed. However, our objective was to test the hypothesis (whether Ames + MNvit could detect *in vivo* genotoxins) by using as extensive a database as possible, including compounds where the evidence of *in vivo* genotoxicity is thus far inconclusive or equivocal, and we did not want to exclude any potential positive *in vivo* findings at this stage of analysis.

We then searched, using the same sources, for published *in vitro* results from Ames, MLA, MNvit and CAvit tests. Wherever possible we checked the data in the original papers to ensure positive results met current criteria, and that a negative result was from a robust study design closely compliant with current IWGT and OECD recommendations. Some chemicals were tested in the MLA as part of the NTP program, and have been re-evaluated by Schisler et al. [12], and other mammalian cell results have been evaluated by the current authors using the same criteria. For some chemicals there were no published MLA *Tk* mutation data, but there were data from HPRT mutation studies. Because the *Hprt* and *Tk* loci detect gene mutagens, we accepted a positive HPRT test (and XPRT or GPT mutations in AS52 and other cells) as being indicative of an MLA-positive. However, because the *Hprt* locus is not as effective at detecting genotoxins that induce large deletions, we did not accept negative HPRT results into the database.

2.3. Response categories

Where possible, for all *in vitro* and *in vivo* data we used the same four response categories as described previously in Kirkland et al. [4], namely:

+: a definitive positive response, either in a single publication or across the majority of publications with the chemical in question. Any negative results could be outweighed by overwhelming dominance of positive publications, or by viewing the data in detail and deciding that the negative test was not adequate.
−: a clearly negative response in all publications found.
E: equivocal, indicating that the results with a given chemical were not consistent (both positive and negative results obtained) either within an experiment, in repeated experiments in the same laboratory, or between laboratories (and therefore between publications). Weak responses, where there was some evidence of a chemically induced effect but no clear dose–response or where biologically significant levels were not reached, were also categorised E. If a published study was considered inconclusive, for convenience we called it equivocal.
TC: technically compromised, indicating a test result that was questionable due to failure to meet essential standard criteria for an adequate study. Some examples would be if a test compound was not tested in a sufficient number of Ames strains, was negative but only tested in the absence of S9, was only tested over short periods in the absence of S9 in mammalian cells, if insufficient sampling times were used, if the test compound did not reach adequate levels of toxicity, or was not tested according to accepted criteria for upper concentrations/doses for non-toxic or insoluble compounds. In such cases where negative results could not be completely judged as conclusive, we called these TC.

All of the data for the new *in vivo* genotoxins database are presented in Appendix A, and any comments of note (e.g. particular cells, tissues or species) are given there. Because the NTP MLA studies have been re-evaluated [12], and a category of "uninterpretable" was used by those authors, we have included this additional response category for MLA data in the database. As for the rodent carcinogens, we accepted the re-evaluated "calls" of NTP MLA studies [12] and applied the same criteria to non-NTP studies.

Because "current" criteria have been applied to old studies, many of the "calls" in Appendix A are different from the author's original calls, or even those by expert review panels. However, we believe it is more appropriate to judge the usefulness of the various *in vitro* tests by evaluating the data according to current criteria.

Where different publications on the same chemical in the same test system gave different results we have attempted to make an overall "call" based on the strength of the evidence. For example, a clear positive result achieving biologically significant levels, according to current criteria, would outweigh a negative result from a study that did not meet current criteria, and the overall call would be positive. If there were several publications on the same chemical, some of which were valid but others were TC, we decided our overall response category for that chemical and endpoint based on the valid studies. However, if two papers equally met current criteria but one was positive and the other negative, the strength of evidence call would be equivocal (E). Where there were different outcomes in different papers for the same chemical and endpoint, we have recorded the original "call" against the cited reference in Appendix A.