mice in vivo, it seems less toxic in vitro, particularly in non-liver cells. That may be because cyclic heptapeptide microcystines do not generally penetrate most cells including bacteria, and a specific transport system may be required [25]. We used 2 parameters to estimate MCLR cytotoxicty—RS and RSG. RS is relative plating efficiency just after exposure, while RSG is relative cell growth for the 3 days following exposure. RSG exhibited stronger response than RS, suggesting that MCLR has an inhibitory effect on cell growth [26]. Because the cytotoxicity was not severe, the genotoxic responses to MCLR must have been due to physiological effects. In the *TK* gene mutation assay, MCLR elevated not only the frequency of mutants, but also the fraction of SG mutants, suggesting that MCLR induced predominantly gross structural changes, such as large deletions, recombinations, and rearrangements.

Molecular analysis strongly supported this hypothesis. Most of the *TK* mutants induced by MCLR were the result of LOH, while the fraction of non-LOH mutants hardly changed (Fig. 3). LOH is an important genetic event in tumorigenesis and is frequently observed in a variety of human tumors. The two major mechanisms for generating LOH are deletion (hemizygous LOH) and inter-allelic recombination (homozygous LOH) [18,19]. Both mechanisms involve the repair of chromosomal double strand breaks (DSBs), either non-homologous end-joining and homologous recombination (HR), although their regulation and role have not been clarified [27]. Other mechanisms may be involved, too, including illegitimate recombination and mitotic non-disjunction [26]. DSB-inducing agents, such as ionizing irradiations, effectively produce LOH mutations through the repair pathways [17,18]. MCLR clastogenic activity may also involve DSBs. Honma and Little [28] demonstrated that 12-*O*-tetradecanoyl-phorbol-13-acetate (TPA), which is the most active tumor promoter known, preferably induces homozygous LOH through HR. MCLR also has tumor promoting activity; like the tumor promoter Okadaic acid, it inhibits protein phosphatase types 1 and 2 A [29]. A cyanobacterial toxin, nodularin, which also inhibits protein phosphatases 1 and 2 A with the same potency as does MCLR has been recognized as rat liver carcinogen rather than a tumor promoter [30]. The genotoxicity of nodularin, however, has not been clear. Matsushima et al. [31] demonstrated that MCLR promotes rat liver cancer initiated with diethyl-nitrosamine. The tumor promoting activity of MCE has been also shown in a two-stage transformation assay in vitro using Syrian hamster embryonic cells [32]. The induction of LOH by MCLR through recombination may be associated with its tumor promoting activity. It is reported that Okadaic acid induces minisatellite mutation in NIH3T3 cells probably through recombination events [33]. The potent hepatocarcinogen aflatoxin B1 also preferably induces LOH through HR in TK6 cells and mouse lymphoma L5178Y cells [34,35].

In conclusion, MCLR was clastogenic in human cells in the present study. It induced LOH, but not point mutations. The genotoxic activity may have been associated with the inductions of DSBs and/or its promoting activity. The association between a high incidence of primary liver cancer and drinking of pond and ditch water polluted by high level of cyanobacteria producing MCLR [3,36,37] suggests that liver is a target organ for MCLR carcinogenisity. Further studies using liver cells and tissues are required to clarify the mechanisms of MCLR genotoxicity in the liver.

Acknowledgements

This study was supported by a grant from the Japan–China Sasagawa Medical Fellowship.

References

[1] W.W. Carmichael. Toxins of freshwater algae. In: A.T. Tu (Ed.), Handbook of natural Toxins. Marcel Dekker, New York, 1988, pp. 121–147.
[2] G.A. Codd, S.G. Bell, P. Brooks, Cyanobacterial toxins in water, Water Sci. Technol. 21 (1989) 1–13.
[3] W.W. Carmichael, The toxins of cyanobacteria, Sci. Am. 270 (1994) 78–86.
[4] F.D. Galey, V.R. Beasley, W.W. Carmichael, G. Kleppe, S. Hooser, W.M. Haschek, Blue-green algae (*Microcysis aeruginosa*) hepatotoxicosis in dairy cows, Am. J. Vet. Res. 48 (1987) 1415.
[5] I.R. Falconer, A.M. Bersfoed, M.T.C. Runnegar, Evidence of liver damage by toxin from a bloom of blue-green algae, *Microcystis aeruginosa*, Med. J. Aust. 1 (1983) 511–514.
[6] I.R. Falconer, Effects on human health of some toxic cyanobacteria (blue-green algae) in reservoirs, Toxicity Assess. 4 (1989) 175–184.
[7] W.P. Brooks, G.A. Codd, Distribution of *Microcystis aeruginosa* peptide toxin and interactions with hepatic microsomes in mice, Pharmacol. Toxicol. 60 (1987) 187–191.

[8] A.S. Dabholker, W.W. Carmichael, Ultrastructural changes in the mouse liver induced by hepatotoxin from the freshwater cyanobacterium Microcystis aeruginosa strain 7820, Toxicon 25 (1987) 285–292.
[9] G.A. Codd, S.G. Bell, P. Brooks, Cyanobacterial toxins in water, Water Sci. Technol. 21 (1989) 1–13.
[10] S.B. Hooser, V.R. Beasley, R.A. Lovell, W.W. Carmichael, W.M. Haschek, Toxicity of microcystin-LR, a cyclic heptapeptide hepatotoxin from *Microcystis aeruginosa* to rats and mice, Vet. Pathol. 26 (1989) 246–252.
[11] P.V. Rao, R. Bhattacharya, S.C. Pant, A.S.B. Bhaskar, Toxicity evaluation of in vitro cultures of freshwater cyanobacterium *Microcystis aeruginosa*. Part 1. Hepatotoxic and histopathological effects in rats, Biomed. Environ. Sci. 8 (1995) 254–264.
[12] R. Bhattacharya, P.V.L. Rao, A.S.B. Bhaskar, S.C. Pant, S.N. Dube, Liver slice culture for assessing hepatotoxicity of freshwater cyanobacteria, Hum. Exp. Toxicol. 15 (1996) 105–110.
[13] W.X. Ding, H.M. Shen, H.G. Zhu, B.L. Lee, C.N. Ong, Genotoxicity of microcystic cyanobacteria extract of a water source in China, Mutation Res. 442 (1999) 69–77.
[14] P.V. Rao, R. Bhattacharya, The cyanobacteral toxin microcystin-LR induced DNA damage in mouse liver in vivo, Toxicology 114 (1996) 29–36.
[15] H. Suzuki, M.F. Watanabe, Y. Wu, T. Sugita, K. Kita, T. Sato, X.L. Wang, H. Tanzawa, S. Sekiya, N. Suzuki, Mutagenicity of microcystin-LR in human RSa cells, Int. J. Mol. Med. 2 (1998) 109–112.
[16] H.L. Liber, W.G. Thilly, Mutation assay at the thymidine kinase locus in diploid human lymphoblasts, Mutat. Res. 94 (1982) 467–485.
[17] C.Y. Li, D.W. Yandell, J.B. Little, Molecular mechanism of spontaneous and induced loss of heterozygosity in human cells in vitro, Somat. Cell Mol. Genet.18 (1992) 77–87.
[18] M. Honma, M. Hayashi, T. Sofuni, Cytotoxic and mutagenic responses to X-rays and chemical mutagens in normal and p53-mutated human lymphoblastoid cells, Mutat. Res. 374 (1997) 89–98.
[19] M. Honma, M. Momose, H. Tanabe, H. Sakamoto, Y. Yu, J.B. Little, T. Sofuni, M. Hayashi, Requirement of wild-type p53 protein for maintenance of chromosomal integrity, Mol. Carcinogenesis 28 (2002) 203–214.
[20] L.S. Zhang, M. Honma, A. Matsuoka, T. Suzuki, T. Sofuni, M. Hayashi, Chromosome painting analysis of spontaneous and methylmethanesulfonate, induced trifluorothymidine-resistant L5178Y cell colonies, Mutat. Res. 370 (1996) 181–190.
[21] E.E. Furth, W.G. Thilly, B.W. Penman, H.L. Liber, W.M. Rand, Quantitative assay for mutation in diploid human lymphoblasts using microtiter plates, Anal. Biochem. 110 (1981) 1–8.
[22] K. Tsuji, T. Watanuki, F. Kondo, M.F. Watanabe, S. Suzuki, H. Nakazawa, H. Uchida, K.I. Harada, Stability of microcystins from cyanobacteria—II. Effect of UV light on decomposition and isomerization, Toxicon 33 (1995) 1619–1631.
[23] K. Tsuji, T. Watanuki, F. Kondo, M.F. Watanabe, H. Nakazawa, S. Suzuki, H. Uchida, K.I. Harada, Stability of microcystins from cyanobacteria. Part IV. Effect of chlorination on decomposition, Toxicon 35 (1997) 1033–1041.
[24] W.M. Repavich, W.C. Sonzongni, J.H. Standridge, R.E. Wedephohl, L.F. Meisner, Cyanobacteria (blue-green algae) in Wisconsin waters: acute and chronic toxicity, Water Res. 24 (1990) 225–231.
[25] M. Runnegar, N. Berndt, N. Kaplowitz, Microcystin uptake and inhibition of protein phosphatases: effects of chemoprotectants and self-inhibition in relation to known hepatic transporters, Toxicol. Appl. Pharmacol. 134 (1995) 264–272.
[26] M. Honma, M. Momose, H. Sakamoto, T. Sofuni, M. Hayashi, Spindle poisons induce allelic loss in mouse lymphoma cells through mitotic non-disjunction, Mutat. Res. 493 (2001) 110–114.
[27] J.M. Stark, M. Jasin, Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks, Mol. Cell. Biol. 23 (2003) 733–743.
[28] M. Honma, J.B. Little, Recombinagenic activity of the phorbol ester 12-O-tetradecanoylphorbol-13-acetate in human lymphoblastoid cells, Carcinogenesis 16 (1995) 1717–1722.
[29] S. Yoshizawa, R. Matsushima, M.F. Watanabe, K.I. Harada, W.W. Carmichael, H. Fujiki, Inhibition of protein phosphatases by microcystins and nodularin associated with hepatotoxicity, J. Cancer Res. Clin. Oncol. 116 (1990) 609–614.
[30] T. Ohta, E. Sueoka, N. Iida, A. Komori, M. Suganuma, R. Nishiwaki, M. Tatematsu, S.J. Kim, W.W. Carmichael, H. Fujiki, Nodularin, a potent inhibitor of protein phosphatases 1 and 2A, is a new environmental carcinogen in male F344 rat liver, Cancer Res. 54 (1994) 6402–6406.
[31] R. Matsushima, T. Ohta, S. Nishiwaki, M. Suganuma, K. Koyama, T. Ishikawa, W.W. Carmichael, H. Fujiki, Liver tumor promotion by the cyanobacterial cyclic peptide toxin microcystin-LR, J. Cancer Res. Clin. Oncol. 118 (1992) 420–424.
[32] H.B. Wang, H.G. Zhu, Promoting activity of microcystins extract from waterblooms in SHE cell transformation assay, Biomed. Environ. Sci. 9 (1996) 46–51.
[33] H. Nakagama, S. Kaneko, H. Shima, H. Inamori, H. Fukuda, R. Kominami, T. Sugimura, M. Nagao, Induction of minisatellite mutation in NIH 3T3 cells by treatment with the tumor promoter okadaic acid, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 10813–10816.
[34] P.M. Stettler, C. Sengstag, Liver carcinogen aflatoxin B1 as an inducer of mitotic recombination in a human cell line, Mol. Carcinogen 31 (2001) 125–138.
[35] V. Preisler, W.J. Caspary, F. Hoppe, R. Hagen, H. Stopper, Aflatoxin B1-induced mitotic recombination in L5178Y mouse lymphoma cells, Mutagenesis 15 (2000) 91–97.
[36] S.Z. Yu, G. Chen, Blue-green algae toxins and liver cancer, Chin. J. Cancer Res. 6 (1994) 9–17.
[37] Y. Ueno, S. Nagata, T. Tsutsumi, A. Hasegawa, M.F. Watanabe, H.D. Park, G.C. Chen, G. Chen, S.Z. Yu, Detection of microcystins, a blue-green algal hepatotoxin, in drinking water sampled in Haimen and Fusui, endemic areas of primary liver cancer in China, by highly sensitive immunoassay, Carcinogenesis 17 (1996) 1317–1321.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Mutation Research 540 (2003) 127–140
MR
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Mouse Lymphoma Thymidine Kinase Gene Mutation Assay: International Workshop on Genotoxicity Tests Workgroup Report—Plymouth, UK 2002[1]

Martha M. Moore[a,*], Masamitsu Honma[b], Julie Clements[c], George Bolcsfoldi[d], Maria Cifone[e], Robert Delongchamp[a], Michael Fellows[f], Bhaskar Gollapudi[g], Peter Jenkinson[h], Paul Kirby[i], Stephan Kirchner[j], Wolfgang Muster[j], Brian Myhr[e], Michael O'Donovan[f], Jo Oliver[k], Takashi Omori[l], Marie-Claude Ouldelhkim[m], Kamala Pant[i], Robert Preston[n], Colin Riach[o], Richard San[p], Leon F. Stankowski, Jr.[e], Ajit Thakur[e], Shinobu Wakuri[q], Isao Yoshimura[r]

[a] *National Center for Toxicological Research, Food and Drug Administration, HFT-120, 3900 NCTR Road, Jefferson, AR, USA*
[b] *National Institute of Health Sciences, Division of Genetics & Mutagenesis, Tokyo, Japan*
[c] *Covance Laboratories, Ltd., Harrogate, North Yorkshire, UK*
[d] *Safety Assessment, AstraZeneca R&D, Södertälje, Sweden*
[e] *Covance Laboratories Inc., Vienna, VA, USA*
[f] *Safety Assessment UK, AstraZeneca R&D, Alderley Park, Macclesfield, UK*
[g] *The Dow Chemical Company, TERC, Midland, MI, USA*
[h] *Safepharm Laboratories Ltd., Shardlow, Derbyshire, UK*
[i] *Sitek Research Laboratories, Rockville, MD, USA*
[j] *F. Hoffmann-La Roche Ltd., Basel, Switzerland*
[k] *GlaxoSmithKline, Ware, Hertforshire, UK*
[l] *National Institute of Health Sciences, Pharmaceuticals and Medical Devices Evaluation Center, Tokyo, Japan*
[m] *Aventis Pharma, Paris, France*
[n] *Johnson & Johnson Pharmaceutical Research & Development, Spring House, PA, USA*
[o] *Inveresk Research, Tranent, Scotland, UK*
[p] *BioReliance Corporation, Rockville, MD, USA*
[q] *Hatano Research Institute, Food and Drug Safety Center, Kanagawa, Japan*
[r] *Faculty of Engineering, Tokyo University of Science, Shinjuku-ku, Tokyo, Japan*

Abstract

The Mouse Lymphoma Assay (MLA) Workgroup of the International Workshop on Genotoxicity Tests (IWGT) met on June 28th and 29th, 2002, in Plymouth, England. This meeting of the MLA group was devoted to discussing the criteria for assay acceptance and appropriate approaches to data evaluation. Prior to the meeting, the group conducted an extensive analysis,

* Corresponding author. Tel.: +1-870-543-7050; fax: +1-870-543-7393.
E-mail address: mmmoore@nctr.fda.gov (M.M. Moore).

[1] The views expressed are those of the authors and not necessarily of their respective government, industrial or university employers. Where needed, this manuscript has been reviewed by appropriate institutions and approved for publication. Approval does not signify that the contents necessarily reflect the views or policies of governmental or other national agencies, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

1383-5718/$ – see front matter
doi:10.1016/j.mrgentox.2003.07.003

of data from both the microwell and soft agar versions of the assay. For the establishment of criteria for assay acceptance, 10 laboratories (6 using the microwell method and 4 using soft agar) provided data on their background mutant frequencies plating efficiencies of the negative/vehicle control, cell suspension growth, and positive control mutant frequencies. Using the distribution curves generated from this data, the Workgroup reached consensus on the range of values that should be used to determine whether an individual experiment is acceptable. In order to establish appropriate approaches for data evaluation, the group used a number of statistical methods to evaluate approximately 400 experimental data sets from 10 laboratories entered into a database created for the earlier MLA Workshop held in New Orleans [Environ. Mol. Mutagen. 40 (2002) 292]. While the Workgroup could not, during this meeting, make a final recommendation for the evaluation of data, a general strategy was developed and the Workgroup members agreed to evaluate this new proposed approach using their own laboratory data. This evaluation should lead to a consensus global approach for data evaluation in the near future.

Keywords: Mouse lymphoma; Genotoxicity; Mutant frequency

1. Introduction

The Mouse Lymphoma Assay (MLA) using the thymidine kinase (*tk*) gene is the most widely used of the various in vitro mammalian cell gene mutation assays. There are currently two equally acceptable methods for performing the assay, one using soft agar medium for cloning and enumeration of mutants [2], and the other using liquid medium and 96-well microwell plates [3,4].

While updated OECD Guideline 476 [5] for in vitro gene mutation assays and the International Committee for Harmonization (ICH) S2B guidance [6] describe performance conditions for the MLA, neither document provides sufficient guidance to assure consistency in results and data interpretation from all laboratories. Because of the importance of the MLA in genotoxicity assessment, it is clear that there is a requirement for an internationally harmonized guideline for the conduct of the assay and the interpretation of data. A group of MLA experts have been working to establish this international harmonization and, as consensus is reached on the important issues, to produce a series of guidance documents.

The first MLA Workgroup meeting was a part of the International Workshop on Genotoxicity Test Procedures held in Washington, DC, in the spring of 1999. The MLA Workgroup, comprised of experts from Japan, Europe, and the United States, reached consensus on a number of important issues and also developed a strategy for data analysis, follow-up discussion and the development of further recommendations [7]. At that meeting, the panel identified three main areas requiring further evaluation and discussion. These included: (1) the conduct of a data-based analysis to result in a final recommendation for the cytotoxicity measure, (2) the issues related to the ICH-recommended use of a 24-h treatment time (including the ability of the assay to detect aneugens), and (3) the criteria for data acceptance and appropriate approaches to data evaluation.

The Workgroup met again in April of 2000, in New Orleans, Louisiana, USA, and reached consensus that the relative total growth (RTG) should be the standard cytotoxicity measure. The use of a 24-h treatment time was discussed and, based on the available information, some recommendations were made. In addition, recognizing the importance of dose selection in the conduct of optimal mouse lymphoma assays, the Workgroup discussed and reached consensus on the important aspects of dose selection. The consensus reached during the New Orleans meeting is reported in Moore et al. [1].

The Workgroup has been working since the 2000 meeting to address the third issue: criteria for data acceptance and appropriate approaches to data evaluation. In this report, we present the Workgroup consensus reached at the June 2002 IWGT meeting in Plymouth, England.

2. Steps for proper assay evaluation

In approaching the analysis of MLA experiments the Workgroup recommends four steps.

(1) Determine whether an experiment meets all assay acceptance (quality control) criteria. If it does not, the experiment should be declared invalid and

another experiment must be conducted. This is particularly important if the data appears to be negative. "Negative" experiments that do not meet the assay acceptance criteria should not be included in the overall evaluation of the test chemical. If the assay meets the acceptance criteria, one can proceed to step 2.

(2) Determine whether the appropriate cytotoxicity range is adequately covered. Generally this requires doses that cover the range between 100 and 10% RTG. A more detailed set of recommendations is reported in Moore et al. [1].

(3) Only after completing steps 1 and 2 should one apply appropriate evaluation criteria to determine whether the observed response is positive, negative, or equivocal.

(4) As a final step, determine whether the data from a single experiment is definitive or whether there is a need to conduct an additional experiment to confirm the first experiment. It is generally necessary to adjust the dose selection for the confirmatory experiment(s) (see Moore et al. [1]).

As further guidance, the Workgroup recommends that all performed experiments be carefully considered in their entirety before reaching a final evaluation of a test chemical.

3. Assay acceptance criteria

In the years prior to the 1999 initiation of the IWGT MLA Workgroup deliberations, various MLA experts held informal discussions and made recommendations for the appropriate values for various assay parameters [8,9]. These recommendations were based on the general experience and opinions of the participants. The current MLA Workgroup decided to approach the establishment of assay acceptance criteria by conducting an evaluation of data generated by laboratories routinely conducting the MLA. All Workgroup participants were asked to submit data for this evaluation.

4. Workgroup data-based approach to setting assay acceptance criteria

MLA data were submitted by 10 laboratories, which included 4 laboratories conducting the agar version of the assay and 6 laboratories performing the microwell method. Each laboratory was asked to provide vehicle and positive control data from at least 50 recent experiments that they considered to provide valid data. Data submitted for positive controls included 7 different positive control chemicals at a number of different concentrations, such that 32 separate treatment groups were represented. The positive control chemicals included methyl methanesulfonate (MMS), ethyl methanesulfonate (EMS), and 4-nitroquinoline oxide (NQO) for experiments conducted without exogenous activation, and 3-methylcholanthrene (MCA), benz[a]pyrene (BP), cyclophosphamide (CP), and dimethylbenzanthracene (DMBA) for S9 activation experiments.

The first approach to establishing appropriate assay acceptance criteria was to determine how the induced mutant frequency (IMF) and RTG for the 32 different positive control treatments depended on the associated negative/vehicle control parameters. These negative/vehicle control parameters were the mutant frequency (MF), the cloning efficiency (CE), and the cellular suspension growth (SG) over the 2-day expression period of each experiment.

The positive control-treated cultures represent the only data available for describing the assay response to a given mutagenic treatment in multiple experiments over the course of time in a laboratory. It is known and accepted that there is variation in the positive control response (and the historical range and standard deviation are monitored and reported according to OECD Guideline 476 [5]). However, the contribution of variations in the concurrent negative/vehicle control parameters to the positive control responses has not been characterized. It was hoped that it might be possible to find consistent linkages between the negative/vehicle control parameters and the IMF for the positive control treatments that would help define appropriate assay acceptance criteria.

Scatter-plots were constructed by plotting the IMF obtained for the positive control treatment in each experiment against each of the three negative/vehicle control parameters. In addition, the RTG of the positive control treatment in each experiment was plotted against the negative/vehicle cloning efficiency and suspension growth. Among a total of 160 scatter-plots so constructed, essentially no relationships between

the negative/vehicle control parameters and the positive control responses were apparent. Visually, the scatter-plots appeared randomly distributed, and statistically, the coefficients of determination (r^2) were consistently small (about 0.16 or less). The highest r^2 value was only 0.444 (meaning that only 44.4% of the data points were consistent with a linear correlation—in this case for one lab only that showed a tendency for the RTG for a CP treatment to decrease as the negative/vehicle control CE increased). Thus, while a few labs may be experiencing some influence of the negative/vehicle control parameters on the IMF or RTG, the overall conclusion from this analysis was that, within the range of control data submitted, there was no dependence of the positive control IMF or RTG on the negative/vehicle control behavior.

In interpreting this first approach to the data evaluation, it is important to note that the Workgroup used only data from experiments considered to be valid by these experienced laboratories. The range of negative/vehicle control parameters available was limited, and the lack of correlation could therefore be considered as affirmation that, for the most part, adequate negative/vehicle control acceptance criteria are already in use. The observed variations in IMF and RTG for each positive control treatment remain a subject for further investigation, but the analysis conducted in this study conclusively showed that assay response did not depend on the negative/vehicle control parameters generally used by the 10 laboratories.

The second approach to setting appropriate negative/vehicle control acceptability criteria was to construct distribution plots for the MF, CE, and SG over the 2-day expression period. The distributions were prepared for each lab, and the data were also combined from the labs to obtain multi-laboratory distributions for both the agar and microwell assay methods. The multi-laboratory distribution plots for negative/vehicle control MF, CE, and SG for the two assay methods are shown in Fig. 1. These distribution plots were reviewed by the Workgroup for the purpose of selecting assay acceptance criteria based on actual data and biological expectations for healthy L5178Y cell cultures. These values should be used to determine if individual experiments are acceptable for the evaluation of a test chemical.

5. Recommended assay acceptance criteria

The Workgroup used the distribution plots (Fig. 1) for guidance in establishing a consensus for the negative/vehicle control acceptance criteria. The recommended criteria from the consensus meeting held in Portland in 1994 and reported by Clive et al. [8] were used as a reference. It should be noted that most (but not all) of the Workgroup members agreed with the newly established values shown below. It should also be noted that with the collection of additional data sets that these values may be modified in the future by the Workgroup.

The new IWGT recommended criteria and the Portland consensus are listed as follows:

5.1. IWGT Workgroup recommendations

Agar method	Microwell method
MF: 35–140 × 10^{-6}	MF: 50–200 × 10^{-6}
CE: 65–120%	CE: 65–120%
SG: 8–32	SG: 8–32

For both methods the CE referred to in the criteria is the absolute cloning efficiency obtained at the time of mutant selection.

5.2. Portland consensus

Agar method	Microwell method
MF: 20 × 10^{-6} and greater	MF: 60 × 10^{-6} and greater
CE: 70–120%	CE: 70–130%
SG: none	SG: none

The Portland consensus did not recommend an upper limit for MF and did not consider acceptable SG values for the negative/vehicle control during the expression period.

5.3. Negative controls

Each laboratory must establish a database for their negative/vehicle controls. This database should include a description of the distribution of observed responses, including the mean/median and 95% confidence interval. In general, the definition of the

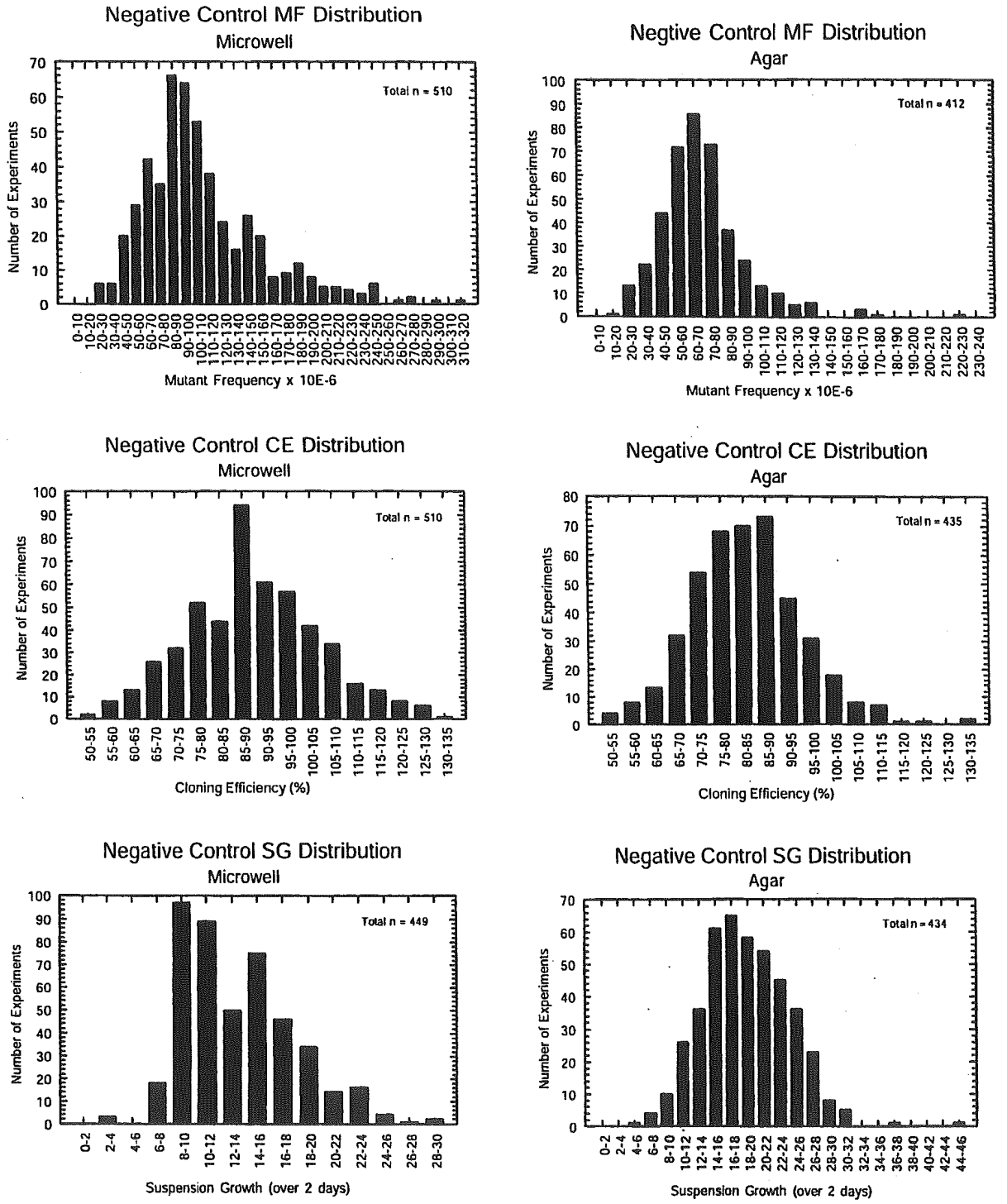

Fig. 1. Distribution plots for negative/vehicle controls. Data were submitted by six laboratories conducting the microwell version of the assay and by four laboratories conducting the agar version.

shape of the distribution curve does not occur until about 50 experimental values are collected for each parameter.

5.3.1. Background (spontaneous) mutant frequency

A low negative/vehicle control MF is often an indicator of poor mutant colony recovery, especially for the small-colony mutants. As soft agar techniques have steadily improved over the years, small-colony recovery in soft agar has improved. The microwell method is even better at detecting small-colony mutants. Thus, the Workgroup agrees that the Portland consensus of a minimum MF of 20×10^{-6} for agar experiments is not acceptable and at this point in the history of the assay, actually represents only a very small proportion of the data submitted for our evaluation. The Workgroup determined that a minimum MF of 50×10^{-6} for microwell experiments is appropriate. The Workgroup recommends raising the minimum MF to 35×10^{-6} for agar experiments, in order to provide as much assurance as possible that small colonies are detected. At the high end, few experiments are rejected if MF values over 140 (agar) or 200 (microwell) are not accepted. The microwell method appears to have a wider dispersion of MF values above the median. By narrowing the acceptable range of negative/vehicle control MF values as suggested, a comparison of test compound data to historical negative/vehicle control data for each lab will likely be facilitated.

5.3.2. Cloning efficiency for the negative control

It has been previously shown that very poor plating efficiency can impact the detected MF, particularly impacting small-colony recovery in the treated cultures [10]. Therefore, the attainment of proper CE is critical to the conduct of the MLA. The newly recommended CE ranges are very similar to the Portland ranges. Few experiments would be eliminated from the overall distributions for both the agar and microwell labs included in this data analysis. Mouse lymphoma cells are generally hardy and, with proper culturing conditions, should always have CEs of 65% or higher. Values above 100% will occur because of errors in cell enumeration and dilution. By controlling this error, the RTG calculations for test material treatments should be better controlled between labs.

5.3.3. Negative/vehicle control suspension growth during expression period

An acceptable range for this negative/vehicle control parameter has not been set in the past, but the Workgroup agreed that limits should be set for this important parameter. The suspension growth contributes to the RTG calculated for the test material treatments, and slowly growing, practically static cultures, could be expected to react differently to test material treatments compared to optimally (logarithmically) growing cultures. The theoretical optimum suspension growth is about 5-fold per day, or 25-fold over the 2-day expression period. As the distribution plots show, there is a wide range of growth rates for laboratories using both methods. Furthermore, the distribution plots are different for the agar and microwell versions, with slower growth generally obtained in the microwell labs. This is surprising because the two methods should be identical in the 2-day expression phase of the experiment. This may result from the fact that most laboratories using the agar method keep cells in motion (i.e. in culture tubes on roller drums) while the microwell laboratories tend to use stationary cultures. The Workgroup is very concerned with growth rates below 8. It is unclear why this should occur and whether this could cause significant non-repeatability in both MF and RTG calculations for weakly positive test compounds. Therefore, a suspension growth range that exceeds 8 was chosen to put a limit on slow population growth. A high limit of 32 allows for reasonable errors in cell counting and dilution and does not eliminate many experiments. The acceptable range should be the same for both assay methods.

5.3.4. Positive controls

Positive control cultures should be included in every MLA experiment. Each laboratory must establish a database for their positive controls. This database should include a description of the distribution of observed responses, including the mean/median and 95% confidence interval. This is done to assure that the assay is working within historical experience and that small-colony *tk* mutants are being adequately enumerated in each experiment. Unfortunately, a number of different chemicals (at various dose levels) are currently used as positive controls by laboratories routinely conducting the MLA. An analysis of the positive control data from the participating laboratories did

not provide an objective approach to setting absolute minimum and maximum MFs for positive controls.

With the goal of providing better use of the positive control information the Workgroup agreed to identify two chemicals that would become the standard positive controls, one with and one without S9 activation. The Workgroup recommends the following properties for these two positive controls: (1) induces primarily small-colony *tk* mutants, (2) is non-volatile, (3) is soluble in water at the required concentrations, (4) is chemically stable thus allowing for freezing aliquots, (5) provides a toxicity curve that is not too steep, and (6) for the S9-dependent positive control, the chemical is only mutagenic with metabolic activation.

6. Appropriate approaches to data evaluation (mutagenicity evaluation)

After satisfying the criteria for experimental acceptance and determining the adequacy of the dose selection, the individual experiment data can be evaluated to determine whether the response is positive, negative, or equivocal.

7. Summary of Workgroup approaches to evaluating the suitablity of various statistical methods

The Workgroup agreed that none of the previously used methods are entirely satisfactory for data analysis. In order to investigate a broad array of possible statistical methods, the group utilized actual experimental data collected from 10 laboratories (6 microwell labs and 4 agar labs). Twenty-nine statistical methods were applied to these 398 data sets (see Table 1). What follows is a brief summary of the analysis conducted. A complete description of the analysis can be found in Delongchamp et al. (in preparation).

In each experiment, MFs are estimated at several doses. With data of this type, a dose–response can be evaluated by testing for a linear trend (TREND) and/or a quadratic trend (QUAD). In those studies where measurements of the MF are replicated in at least some of the doses (368 studies), an overall dose-effect can be tested by analysis of variance (ANOVA). For our evaluation, a total of 27 versions of these generic statistical methods were computed by varying the dependent variable, the independent variable, and weight associated with the residual variance. Table 1 defines these versions. They represent a cross-section of statistical analyses that could be applied to these studies. All of these methods are linear models that can be estimated by any weighted least-squares regression program. For this analysis, the Workgroup used SAS procedures, PROC GLM for the ANOVA methods, and PROC REG for the TREND and QUAD methods. Methods #1, #2, #11, #12, #21, and #28 are ANOVA methods, methods #3, #4, #7–10, #13, #14, #17–20, #22, #24, #25, and #29 are TREND methods, and methods #5, #6, #15, #16, and #23 are QUAD methods. Methods #4, #12, #28, and #29 are based on the methods developed by Robinson et al. [11] for the United Kingdom Environmental Mutagen Society (UKEMS). The exact methods of Robinson et al. [11] were not implemented because they calculated heterogeneity factors, H_m and H_s, which require historical data. For our analysis, we used arbitrary values for the heterogeneity factors: methods #4 and #12 assume $H_m = H_s$, while methods #28 and #29 assume $3.0H_m = 1.8H_s$. Otherwise, methods #4 and #29 are weighted regressions of the mutant fraction on dose and methods #12 and #28 are weighted analyses of variance on the logarithm of mutant fraction (we used the ANOVA P-value instead of a series of Dunnett's tests). This approach is quite similar to the statistical methods of Omori et al. [12].

The major difference between the methods reported here and those of Robinson et al. [11] or Omori et al.[12] is the manner in which extra-binomial variation is modeled (extra-Poisson variation with agar method). Apart from their heterogeneity factors, our weights are equivalent to the weights derived in Robinson et al. [11]. The weights, which we used, are derived for the MF and its logarithm from the appropriate likelihood functions: ratio of binomial random variables in the case of microwell studies or the ratio of Poisson random variables in the case of agar studies. We model over-dispersion by assuming that the dependent variable has a normal distribution with variance proportional to the variance of the MF (binomial-based variance in microwell studies/Poisson-based variance in agar studies). Hence, we judge statistical significance based upon the F-distribution, not the chi-squared distribution [11,12].

Table 1
Statistical methods are defined by the dependent variable {MF, log(MF), or rank(MF)}, independent variable(s) {dose, dose categories, dose + dose × dose, log(dose), or rank(dose)} and the weight {none, variance(dependent variable)}

Method	Dependent variable	Independent variable	Weight
1	MF	Dose category	No
2			Yes
3		Dose	No
4			Yes
5		Dose + dose × dose	No
6			Yes
7		log(dose)	No
8			Yes
9		Rank(dose)	No
10			Yes
11	log(MF)	Dose category	No
12			Yes
13		Dose	No
14			Yes
15		Dose + dose × dose	No
16			Yes
17		log(dose)	No
18			Yes
19		Rank(dose)	No
20			Yes
21	Rank(MF)	Dose category	No
22		Dose	No
23		Dose + dose × dose	No
24		log(dose)	No
25		Rank(dose)	No
26	Induced MF, see Mitchell et al. [13]		
27	Two-fold change		
28	log(MF)	Dose category	Yes
29	MF	Dose	Yes

Several conclusions can be drawn from these analyses. When a *P*-value less than 0.05 is used to judge a data set as positive, all of the statistical methods agreed on the positive or negative call in approximately 40% of the data sets. However, for 60% of the data sets, the various statistical methods gave divergent positive/negative results. Graphical representations were made for all of the data sets. This provided a means to visually analyze the shapes of the dose–response curves and to determine that the data sets include all possible shapes of dose–response curves. To facilitate understanding how the various statistical methods deal with various dose–response curve shapes, three members of the Workgroup evaluated the curve shapes and placed the data sets into categories based on shape. A full description of this analysis can be found in Delongchamp et al. (in preparation). The two extremes of the curve shapes were those showing very clear, large positive increases in MF with increasing dose (Fig. 2), and those showing patterns where the test cultures gave MFs that varied around the background MF in that particular experiment (Fig. 3). Approximately 10% of the 398 experiments gave curves similar to Fig. 2, which should be evaluated as positive, and about 50% gave curves similar to Fig. 3, which should be evaluated as negative. Another approximately 20% yielded dose–response curves that showed very small upward trends with increasing dose. The remaining experiments fell into curves showing MF increases that would generally be classified as weak positives or curves that showed a positive upward trend followed by a downturn. Thus, this database includes a fairly representative sample of "real world" MLA test data.

It is widely recognized that it is easy to classify as positive chemicals inducing large increases in MF (such as seen in Fig. 2). The difficulty arises in those

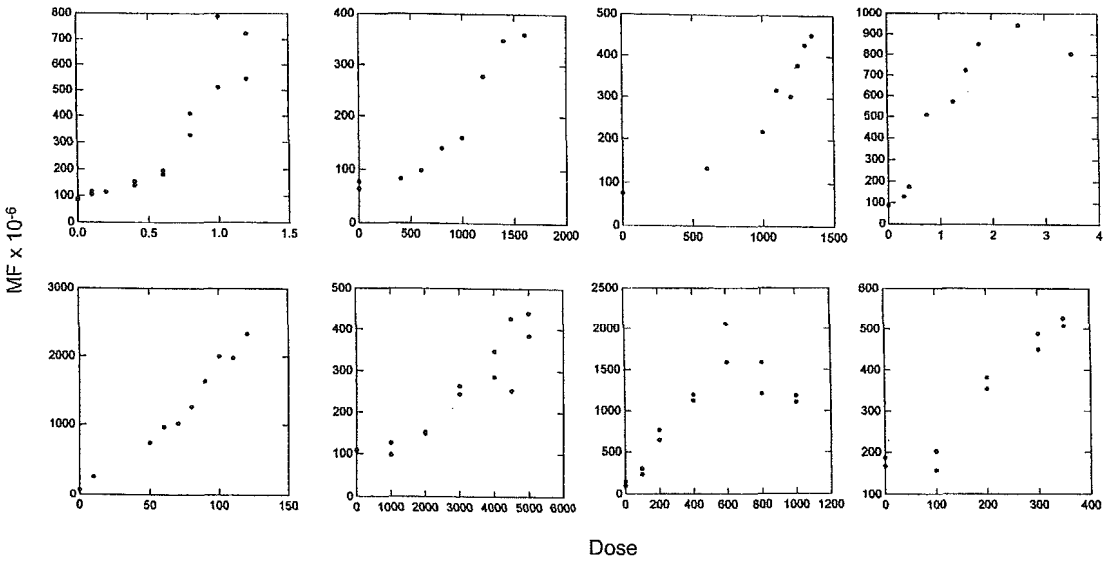

Fig. 2. Examples from the data set of 398 experiments showing clearly positive responses.

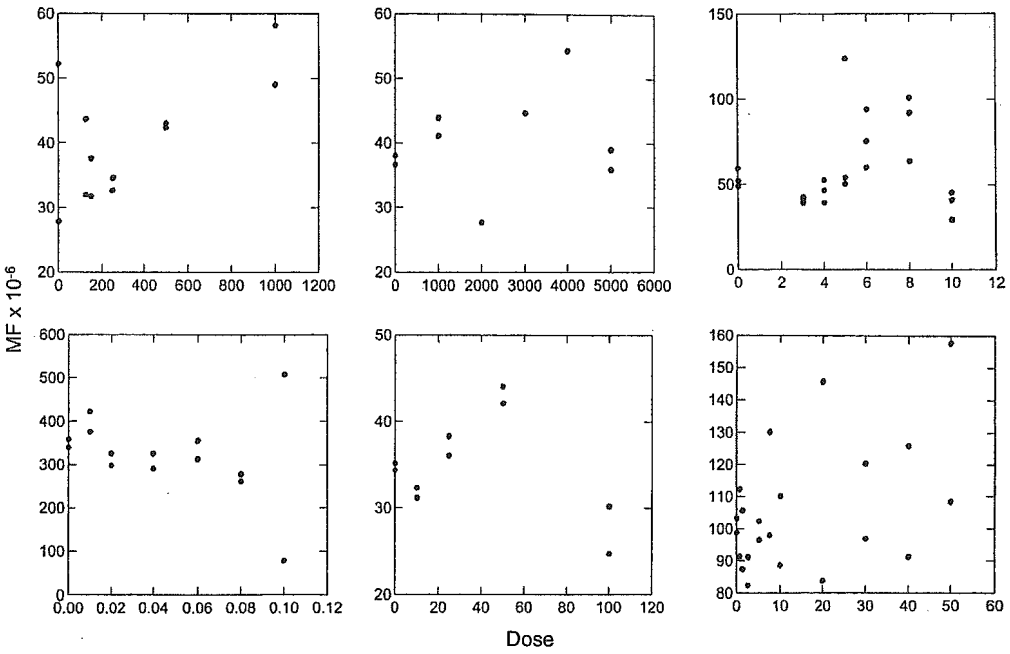

Fig. 3. Examples from the data set of 398 experiments where at least one statistical method judged the response to be positive. It is, however, clear that the responses actually represent the inherent variability of the negative/vehicle control.

situations where the increase in MF is very small. Statistical methods are designed to estimate the probability that a particular response occurred by chance. Therefore, those experiments having little variability between duplicate cultures or data points fitting tightly on a linear (or quadratic) curve are more likely to be judged positive (by statistical analysis) than experiments showing large variability.

We know from experience that the background MF shows a broad range of values (Fig. 1). All laboratories generate some experimental data where a "statistically" positive dose–response curve can fit within the normal range of the background MF. In some of these experiments, particularly if the variability between data points is small, some statistical methods will determine that the response is positive. In other instances, and by other statistical methods, the response will be declared negative. Our evaluation of the 398 data sets (from the database created for the New Orleans Meeting) by these 29 methods confirmed this. In fact, all experiments with very slight to moderately increasing dose–responses were determined positive by some methods and negative by others.

As an illustration of this, Fig. 3 shows some curve shapes that were declared to be positive by one or more of the statistical methods. The Workgroup members feel that these dose–response curve shapes clearly represent negative MLA responses.

8. Biological relevance of assay responses

Using the results from the analysis of the various statistical methods, the Workgroup concluded that there is no single, "correct" statistical method for evaluating MLA data.

However, recognizing the importance of a standardized approach to evaluating MLA data, the Workgroup proceeded to consider the general issue of biological relevance and other properties of the MLA in arriving at their final recommendation.

(1) First, the MLA is capable of detecting large increases in MFs (see examples in Fig. 2). This means that chemicals capable of inducing high MFs will be readily identified as such by the MLA.
(2) Second, very small increases in MF can often be seen at high cytotoxicity levels. While the Workgroup recommends that experiments should adequately cover the entire dose range (for negative or weakly positive chemicals), they also recognize that greater variability in both the RTG and the MF can occur under these highly toxic exposures. It has long been recognized for all in vitro systems that high toxicity may lead to secondary effects and that mutations induced under these circumstances would not normally occur in vivo. Thus, these responses, while perhaps statistically significant by some methods, are not considered to be biologically relevant.
(3) Third, mutations are induced in an additive manner rather than by fold-increases of the background MF. Statistical methods for trend evaluate the slope of dose–response curves rather than the significance of absolute increases. This fact becomes important when an experiment has a very low background. Several experiments in the data set had background MFs of 20–30 $\times$ 10^{-6}. While experiments with such low background MFs would be eliminated by applying the new IWGT MLA Workgroup acceptance criteria, an evaluation of these dose–response curves makes an important point. Fig. 4 shows two experiments with low background MFs and a positive trend with increasing dose. While the highest IMF is very low (approximately 20 $\times$ 10^{-6}), several statistical methods call this response positive. In fact, statistical methods view these responses to be the same as the responses seen in Fig. 5, where the background is approximately 200 and the highest IMF is approximately 200 $\times$ 10^{-6}. When one considers the biological relevance of the two responses, it is clear that an IMF of approximately 200 $\times$ 10^{-6} is a much larger response than an IMF of approximately 20 $\times$ 10^{-6}.
(4) Finally, the assay response to a test chemical is not well defined in a single experiment in any laboratory. It is clear from the wide range of positive control responses, which occurs for all positive control treatments examined and for all laboratories, that the same behavior will occur for test chemicals. This means that the MF range for many test chemicals (that would occur if many experiments were performed) would overlap or be contained within the background MF range. Thus, a positive statistical test, while appropriate for an

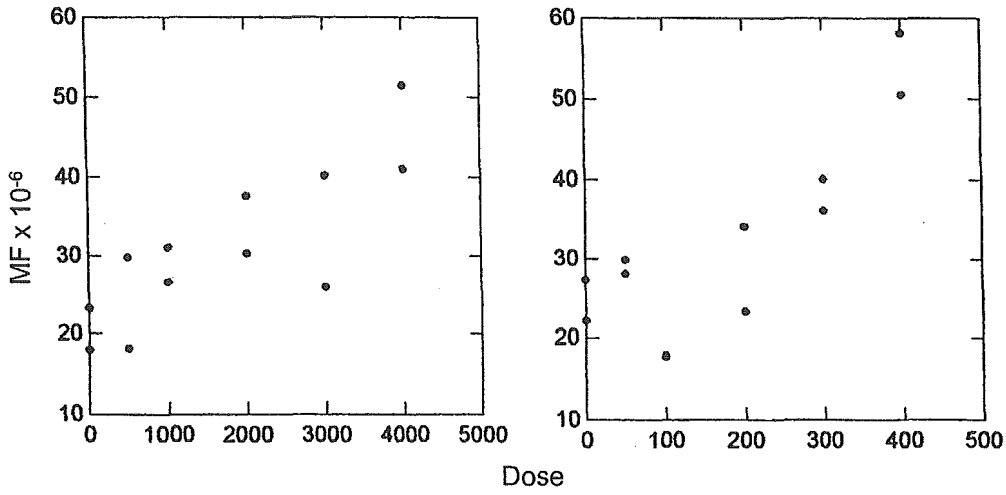

Fig. 4. Examples from the data set of 398 experiments where the background MF of approximately 20×10^{-6} is doubled at the highest dose.

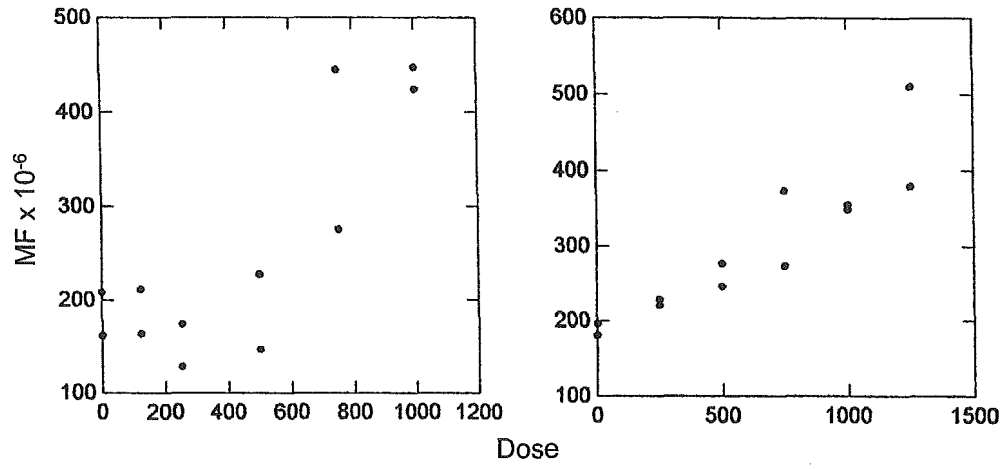

Fig. 5. Examples from the data set of 398 experiments where the background MF of approximately 200×10^{-6} is doubled at the highest dose.

individual experiment, is inadequate for the evaluation of a test chemical in a global sense (repeatability within- and between-laboratories).

9. Recommendation for combining biological relevance with statistical analysis

The majority of the Workgroup agreed that a biologically relevant approach to MLA data evaluation might be achieved by requiring that the IMF exceed some value based on the historical background MF. Once a response reaches the required IMF level, statistical analysis should be applied to determine whether there is a dose-related increase in MF. Such a procedure would disregard small increases in IMF, judged by experts in the assay to be of no biological significance, but would retain the objectivity of a statistical method.

The choice of an appropriate minimum increase in MF over the concurrent background MF in order to regard a response as significant was previously debated and arbitrarily selected for the EPA Genetox Committee review of published data [13]. Now, however, the

acceptable range of MFs for the vehicle controls can be used to provide a rationale in selecting the minimum required IMF. In the future, all acceptable assays will fall within the acceptable ranges for negative/vehicle control MFs. The mean of the low and high limits defines the mid-value MF for each method. The Workgroup proposed that this mid-value MF might be used as a data-based estimate of a global evaluation factor for each method.

The majority of the Workgroup agreed that it would be desirable to use the multi-laboratory distributions shown in Fig. 1 to calculate the global evaluation factor and that factor would be used for evaluating MLA experiments in all laboratories. Based on the current background MF distribution curves, the global evaluation factors are approximately 100×10^{-6} (88×10^{-6} for the agar method and 125×10^{-6} for the microwell method). A few Workgroup members felt that each laboratory should calculate and use an evaluation factor based exclusively on their own laboratory historical data.

In either case, once it is established that the IMF in one or more treated cultures of an experiment equals or exceeds the global evaluation factor, then appropriate statistics should be applied to determine if there is a dose-related increase in MF. For example, if the negative/vehicle control MF in an agar experiment is 50×10^{-6}, then one of the test cultures must have a MF of at least $50 + 88$ (the agar global evaluation factor) $= 138 \times 10^{-6}$ in order to trigger the application of statistics.

10. Follow-up evaluation to develop final recommendations

Because of the importance of the recommendations for data evaluation, the Workgroup agreed that this newly proposed approach should be thoroughly tested prior to its acceptance. The approach will be applied to both the 398 data sets and also to individual laboratory data sets. Once these evaluations have been conducted, final recommendations will be made. This new approach is summarized below:

(1) Experiments found to meet the acceptance criteria (as defined above) and also to properly cover the cytotoxicity range (see Moore et al. [1]) are examined to determine if the IMF (MF minus the concurrent negative/vehicle control MF) for any treatment (whether single or replicate dosing is used) equals or exceeds the global evaluation factor. The global evaluation factor values, based on the accepted ranges for the negative/vehicle control MF are currently approximately 100×10^{-6} but will be established during the next year both for agar and for microwell. If the global evaluation factor is exceeded, the evaluation moves to step 2. If the factor is not exceeded the test is negative.
(2) The second step is to evaluate the data for the presence of a positive dose-related trend by an appropriate, one-sided statistical test. To avoid false negative evaluations due to low degrees of freedom, for optimal statistical analysis an experiment should include at least four dose levels. Either single or replicate cultures may be used for each test chemical dose group. The negative/vehicle control should always be performed in replicate, which will allow a goodness of fit test to be used in the statistical analysis.
(3) A test agent response is positive if *both* a positive trend test is obtained and the IMF meets or exceeds the global evaluation factor.

11. Guidance for confirmatory experiments

In compliance with ICH and OECD Guidelines, if all the test results are clearly negative under all the conditions defined by the MLA Workgroup (appropriate coverage of the recommended dose range, acceptable values for the negative/vehicle control, limited by solubility or 10 mM, etc.), or clearly positive in any of the conditions, then an acceptable study would comprise a single test with 3–4 h exposure with and without S9 and 24 h exposure without S9 (for negatives at the short treatment time). Additional experiments are necessary to clarify results that are not clearly positive or negative. It is generally necessary to adjust the concentrations for the additional experiments.

References

[1] M.M. Moore, M. Honma, J. Clements, K. Harrington-Brock, T. Awogi, G. Bolcsfoldi, M. Cifone, D. Collard, M. Fellows, K. Flanders, B. Gollapudi, P. Jenkinson, P. Kirby, S.

Kirchner, J. Kraycer, S. McEnaney, W. Muster, B. Myhr, M. O'Donovan, J. Oliver, M.-C. Ouldelhkim, K. Pant, R. Preston, C. Riach, R. San, H. Shimada, L. Stankowski Jr., Mouse lymphoma thymidine kinase gene mutation assay: follow-up international workshop on genotoxicity test procedures, New Orleans, Louisiana, Environ. Mol. Mutagen. 40 (2002) 292–299.

[2] N.T. Turner, A.G. Batson, D. Clive, Procedures for the L5178Y/TK$^{+/-}$ to TK$^{-/-}$ mouse lymphoma assay, in: B.J. Kilbey, M. Legator, W. Nichols, C. Ranel (Eds.), Handbook of Mutagenicity Test Procedures, 2nd ed., Elsevier, Amsterdam, 1984, pp. 239–268.

[3] J. Cole, C.F. Arlett, M.H.L. Green, J Lowe, W. Muriel, A comparison of the agar cloning and microtitration techniques for assaying cell survival and mutation frequency in L5178Y mouse lymphoma cells, Mutat. Res. 111 (1983) 371–386.

[4] J. Cole, W.J. Muriel, B.A. Bridges, The mutagenicity of sodium fluoride to L5178Y (wild-type and TK$^{+/-}$ 3.7.2C) mouse lymphoma cells, Mutagenesis 1 (1986) 157–167.

[5] OECD Guideline 476, In Vitro Mammalian Cell Mutation Test, adopted July 21, 1997.

[6] ICH Guidance S2B, Genotoxicity: a standard battery for genotoxicity testing of pharmaceuticals, Fed. Reg. 62 (1997) 62472.

[7] M.M. Moore, M. Honma, J. Clements, T. Awogi, G. Bolcsfoldi, J. Cole, B. Gollapudi, K. Harrington-Brock, A. Mitchell, W. Muster, B. Myhr, M. O'Donovan, M.-C. Ouldelhkim, R. San, H. Shimada, L. Stankowski Jr., The Mouse Lymphoma Thymidine Kinase Locus (*Tk*) Gene Mutation Assay: International Workshop on Genotoxicity Test Procedures (IWGTP) workgroup report, Environ. Mol. Mutagen. 35 (2000) 185–190.

[8] D. Clive, G. Bolcsfoldi, J. Clements, J. Cole, M. Honma, J. Majeska, M. Moore, L. Muller, B. Myhr, T. Oberly, M-C. Ouldelhkim, C. Rudd, H. Shimada, T. Sofuni, V. Thybaud, P. Wilcox, Consensus agreement regarding protocol issues discussed during the mouse lymphoma workshop: Portland, Oregon, May 7, 1994, Environ. Mol. Mutagen. 25 (1995) 165–168.

[9] T. Sofuni, P. Wilcox, H. Shimada, J. Clements, M. Honma, D. Clive, M. Green, V. Thybaud, R.H.C. San, B.M. Elliott, L. Muller, Mouse lymphoma workshop: Victoria, British Columbia, Canada, March 27, 1996. Protocol issues regarding the use of the microwell method of the mouse lymphoma assay, Environ. Mol. Mutagen. 29 (1997) 434–438.

[10] M. Meyer, K. Brock, K. Lawrence, B. Casto, M.M. Moore, Evaluation of the effect of agar on the results obtained in the L5178Y mouse lymphoma assay, Environ. Mutagen. 8 (1986) 727–740.

[11] W.D. Robinson, M.H.L. Green, J. Cole, M.J.R. Healy, R.C. Garner, D. Gatehouse, Statistical evaluation of bacterial/mammalian fluctuation tests, in: D.J. Kirkland (Ed.), Statistical Evaluation of Mutagenicity Test Data, Cambridge University Press, Cambridge, 1989, pp. 102–140.

[12] T. Omori, M. Honma, M. Hayashi, Y. Honda, I. Yoshimura, A new statistical method for evaluation of L5178Y *tk*$^{+/-}$ mammalian mutation data using microwell method, Mutat. Res. 517 (2002) 199–208.

[13] A.D. Mitchell, A.E. Auletta, D. Clive, P.E. Kirby, M.M. Moore, B.C. Myhr, The L5178Y/*tk*$^{+/-}$ mouse lymphoma specific gene and chromosomal mutation assay. A phase III report of the U.S. Environmental Protection Agency Gene-Tox Program, Mutat. Res. 394 (1997) 177–303.