Unclassified **English - Or. English**

19 September 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Detailed Review Paper on the miniaturised versions of the bacterial reverse gene mutation test

Series on Testing and Assessment
No. 358

JT03502634

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

OECD Environment, Health and Safety Publications
Series on Testing & Assessment
No. 358

Detailed Review Paper on the miniaturised versions of the bacterial reverse gene mutation test

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment: link

For this and many other Environment,

Health and Safety publications, consult the OECD's

World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,

Environment, Health and Safety Division

2 rue André-Pascal

75775 Paris Cedex 16

France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

© OECD 2022 Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

Foreword

This document presents the Detailed Review Paper (DRP) on the miniaturised versions of the bacterial reverse gene mutation test.

The DRP was prepared by a dedicated OECD Expert Group formed to develop and provide input into the document, as part as a project led by Belgium, the Netherlands and the US. The DRP provides an overview of the available miniaturised bacterial reverse gene mutation tests and presents the results of a retrospective analysis conducted on existing data. The objective of the retrospective analysis was to better understand the performance of each of the miniaturised tests compared to the standard OECD Test Guideline 471. The DRP also discusses the outcome of the data analysis, taking into account the limitations of the retrospective validation study and provides recommendations on the future use and/or the additional work needed for each of the miniaturised bacterial reverse gene mutation tests.

Draft versions of the DRP were circulated twice for comments to the Working Party of the National Coordinators of the Test Guidelines Programme (WNT), in September and December 2021. The DRP was approved by the WNT in April 2022. This document is published under the responsibility of the Chemicals and Biotechnology Committee.

Table of Contents

Background information/Scope 12

Glossary 13

Introduction 16

Executive summary 17

1 Overview of the available bacterial reverse gene mutation tests 24
1.1. Introduction 24
1.1.1. Standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997) 24
1.1.2. Grouping of miniaturised bacterial reverse gene mutation tests 31
1.2. Miniaturised agar plating tests using standard tester strains 32
1.2.1. Introduction 32
1.2.2. Miniaturised agar plating test in 6-well plates 33
1.2.3. Miniaturised agar plating test in 24-well plates 37
1.3. Microfluctuations tests using standard tester strains 39
1.3.1. History 39
1.3.2. Most important characteristics 40
1.3.3. Important differences with standard bacterial reverse gene mutation test 42
1.3.4. Perceived advantages 43
1.3.5. Challenges 43
1.3.6. Historical performance 44
1.4. Miniaturised agar plating tests using non-standard tester strains 45
1.4.1. Miniaturised version of the microsuspension method 45
1.4.2. Miniaturised agar plating tests using bioluminescent tester strains ('Bioluminescent Agar test') 45
1.5. Microfluctuation test using non-standard tester strains 46
1.5.1. Ames II test 46
1.5.2. Microfluctuation tests using bioluminescent tester strains 52

2 Applications of the miniaturised bacterial reverse gene mutation tests 54
2.1. Current applications of the miniaturised bacterial reverse gene mutation tests 54
2.1.1. Testing of drug impurities 54
2.1.2. Screening of chemicals 54
2.1.3. Other applications 55
2.2. Possible scenarios for future use of the miniaturised bacterial reverse gene mutation tests 56

3 Retrospective data analysis 58
3.1. Data collection 59
3.1.1. Exploratory survey 59
3.1.2. First Call for data 60

3.1.3. Second Call for data 62
3.2. Comparison of protocol variations among the laboratories 63
3.2.1. 6-well plate assay 65
3.2.2. 24-well plate assay 66
3.2.3. Ames II and Ames MPF™ 67
3.3. Data curation and preparation of the master dataset for analysis 68
3.3.1. Uniformity of assessment calls 69
3.3.2. Handling of 'multiple assessments' and 'mixed calls' 69
3.3.3. Master dataset description 70
3.4. Retrospective data analysis 80
3.4.1. Data analysis approach 80
3.4.2. Results of Assay correspondence analyses 84
3.4.3. Summary of Assay Correspondence Analyses 108

4 Discussion 115
4.1. Interpretation of results 115
4.2. Limitations of the Retrospective Performance Analyses 117
4.3. Next steps 119

5 Recommendations 122

6 References 123

Annex I. OECD Expert group Survey on Miniaturised Ames Tests 132

Annex 2: Supplementary tables 135

Table 1. Genotypic information for the bacterial strains recommended in the standard OECD TG 471 bacterial reverse gene mutation test (OECD, 1997). 26
Table 2. Most important characteristics of the standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997) and different miniaturised versions (vc: vehicle control; pc: positive control). 30
Table 3. Volumes per well (in µL) for the Ames MPF™ assay (o.n. culture: overnight culture). 41
Table 4. Genotypes of TA98 and TAMix S. typhimurium strains (from Flückiger-Isler and Kamber, 2014). 47
Table 5. Concordance analysis by Gee et al (1998) for the Ames II assay. 52
Table 6. Overview of the application of miniaturised bacterial reverse gene mutation test in different regulatory domains. 56
Table 7. Overview of the data fields included in the standardised Excel data collection spreadsheet. 60
Table 8. Overview of relative sizes of some parameters and solvent control counts as reported by laboratories participating in the retrospective validation. 63
Table 9. Strains used and strain-specific outcome criteria provided by the laboratories. 64
Table 10. Summary of the protocol variations for the 6-well plate assay. 66
Table 11. Summary of the protocol variations for the 24-well plate assay. 67
Table 12. Summary of the protocol variations for the Ames II and Ames MPF™ assays. 68
Table 13. Descriptive Overview of Collected Data By Assay Type 71
Table 14. Descriptive Overview of Collected Data – Strain Use Comparison by Assay Type 71
Table 15. Format of the 2x2 contingency tables used for correspondence analyses. 81
Table 16. Agreement categorization based on Kappa/PABAK statistic values. 83
Table 17. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values are based on comparisons of individual assessments (e.g., combinations of test chemical, strain and S9-mix). Equivocal responses excluded from the analyses. 86
Table 18. Performance of miniaturised bacterial mutations assays based on comparisons of individual assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits. 86
Table 19. Number of assessments, number of substances, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on individual assessments; equivocal responses excluded. 87

Table 20. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values based on comparisons of 5-strain overall calls. 5 strains included S. typhimurium TA100, TA98, TA1535, TA1537 or 97 or 97a, and E. coli WP2 strain or S. typhimurium TA102. Strains for Ames II assay included only S. typhimurium TA98 and TAMix. Equivocal responses excluded from the analyses. 90
Table 21. Performance of miniaturised bacterial mutations assays based on comparisons of 5-strain overall call assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits. 90
Table 22. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on 5-strain overall call assessments; equivocal responses excluded. 91
Table 23. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values based on comparisons of 2-strain overall calls. With the exception of the Ames II assay, calls based on responses for S. typhimurium TA98 and TA100. For the Ames II assay, calls based on S. typhimurium TA98 and TAMix. Equivocal responses excluded from the analyses. 91
Table 24. Performance of miniaturised bacterial mutations assays based on comparisons of 2-strain overall call assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits. 91
Table 25. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on 2-strain overall call assessments; equivocal responses excluded. 92
Table 26. 2 x 2 contingency tables showing call frequency correspondence values for each the Ames MPF^{TM} and Ames II assays versus the preincubation version of the standard assay. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 95
Table 27. Performance of the Ames MPF^{TM} and Ames II assays based on comparisons of individual assessments with the preincubation version of the standard assay. Equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 95
Table 28. Number of assessments, number of substances, PI and BI for correspondence analyses between the Ames MPF^{TM} and Ames II assays and the preincubation version of the standard bacterial reverse gene mutation test based on individual assessments. Equivocal responses not included in the analyses. 95
Table 29. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the plate incorporation version of the standard assay (i.e., without preincubation). The results for the 6-well and 24-well plate assays are also plate incorporation versions of the assay (i.e., without preincubation). Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 96
Table 30. Performance of the miniaturised assays based on comparisons of individual assessments with the plate incorporation version of the standard assay (i.e., without preincubation). Equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 97
Table 31. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the plate incorporation version of the standard assay (i.e., without preincubation) based on individual assessments. Equivocal responses not included in the analyses. 97
Table 32. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA98 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 100
Table 33. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA98 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 100
Table 34. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA98 only; equivocal responses not included in the analyses. 101
Table 35. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA100 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 101
Table 36. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA100 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 101
Table 37. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene

mutation assay based on comparisons of individual assessments . Strain S. typhimurium TA100 only; Equivocal responses not included in the analyses. 102
Table 38. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA1535 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 103
Table 39. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA1535 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 103
Table 40. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA1535 only; equivocal responses not included in the analyses. 103
Table 41. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA97 or TA1537 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 104
Table 42. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA97 or TA1537 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 105
Table 43. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA97 or TA1537 only; equivocal responses not included in the analyses. 105
Table 44. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA102 - E. coli WP2 combination. Frequency values are based on individual assessments. Equivocal responses not included in the analyses. 106
Table 45. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. S. typhimurium TA102 - E. coli WP2 combination; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits. 106
Table 46. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. S. typhimurium TA102 - E. coli WP2 combination; equivocal responses not included in the analyses. 106
Table 47. Characteristics of test chemicals that adequately represent the applicability domain of the bacterial reverse gene mutation test 120
Table 48. Test chemicals assessed and included in the retrospective analyses. An additional 188 coded test chemicals were assessed and included in the analyses. 135
Table 49. Test chemicals that elicited equivocal responses on the standard bacterial reverse gene mutation test. All entries reflect the call provided by the data submitter. 145
Table 50. Test chemicals that elicited equivocal responses on the 6-well plate assay. All entries reflect the call provided by the data submitter. 147
Table 51. Test chemicals that elicited equivocal responses on the Ames MPF™ assay. All entries reflect the call provided by the data submitter. 148
Table 52. Test chemicals that elicited equivocal responses on the 24-well plate assay. All entries reflect the call provided by the data submitter. 152
Table 53. Test chemicals that elicited equivocal responses on the Ames II assay. All entries reflect the call provided by the data submitter. 152
Table 54. Summary of correspondence mismatches; 6-well plate assay versus standard assay, individual assessments. Equivocal calls not included in the analysis. 153
Table 55. Summary of correspondence mismatches; 24-well plate assay versus standard assay, individual assessments. Equivocal calls not included in the analysis. 154
Table 56. Summary of correspondence mismatches; Ames MPF™ assay versus standard, individual assessments. Equivocal calls not included in the analysis. 154
Table 57. Summary of correspondence mismatches; Ames II assay versus standard, individual assessments. Equivocal calls not included in the analysis. 156
Table 58. Results of the test chemicals assessed in the 6-well plate assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold. 157

Table 59. Results of the test chemicals assessed in the 24-well plate assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold. 162
Table 60. Results of the test chemicals assessed in the Ames MPF^{TM} assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold. 162
Table 61. Results of the test chemicals assessed in the Ames II assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold. 164
Table 62. Overall summary of performance metrics and agreement statistics 167

Figure 1. Schematic overview of the different types of miniaturised bacterial reverse gene mutation tests.32
Figure 2. Experimental set-up of the Ames MPF^{TM} assay. C-: negative control; C+: positive control; D1-6: test chemical dose 1 to 6. 41
Figure 3. Timeline summarising the different steps of the data collection and data analysis for the retrospective validation study. 58
Figure 4. Visualization of the chemical space of the test chemicals used in the different bacterial reverse gene mutation tests as represented by the first and second principal component. 73
Figure 5. Projection of the chemical space of the bacterial reverse gene mutation tests onto the REACH and DrugBank databases: Physicochemical descriptor-based PCA (upper panels) and fingerprint-based PCA (lower panels). 75
Figure 6. 3D representation of the chemical space (distribution based on structure similarity) of (left) DrugBank approved drugs (n = 2509) and (right) REACH registered industrial chemicals with high quality SMILES-structure-Name relationship (downloaded from the OECD Toolbox n=7404). 76
Figure 7. Details of the chemical space covered by chemicals tested either in the standard (left) or in all the miniaturised versions (right) of the bacterial reverse gene mutation assay. 77
Figure 8. Details of the chemical space covered by chemicals tested either in 6-well (left) or 24-well (right) plate versions of the bacterial reverse gene mutation assay. 77
Figure 9. Details of the chemical space covered by chemicals tested either in the Ames II (left) or Ames MPF^{TM} (right) versions of the bacterial reverse gene mutation assay. 78
Figure 10. Chemical space covered by the chemicals tested in the various versions of bacterial reverse gene mutation assays (n=227) with REACH registered chemicals with high quality SMILES-structure-Name relationship (downloaded from the OECD Toolbox n=7404) and DrugBank approved drugs (n = 2509). 79

Background information/Scope

1. Bacterial gene mutation tests can be subdivided in back/reverse mutation tests and forward mutations tests. However, because the bacterial tests detecting forward mutations (Miller et al., 2005; Hamel et al., 2016) are rarely used, the scope of the present detailed review paper was limited to the back/reverse mutation assays. These tests are further referred to as ‘bacterial reverse gene mutation tests’.

Glossary

2-strain overall call: overall call based on responses for *Salmonella* strains TA98 and TA100. A 2-strain overall call was deemed possible only if the results of 4 assessments for each test were available, i.e., assessments in each of the 2 strains with and without S9, for both standard and miniaturises tests. For the Ames II assay, the 2-strain overall calls were based on results for TA98 and TAMix.

5-strain overall call: overall call based on responses for all 5 bacterial strains specified in OECD TG471 i.e., (i) TA98, (ii) TA100, (iii) TA1535, (iv) TA97, 97a or 1537, and (v) *E. coli* WP2 or TA102. A 5-strain overall call was deemed possible only if the results of 10 assessments for each test were available, i.e., assessments in each of the 5 strains with and without S9, for both standard and miniaturises tests. For the Ames II assay, the 5-strain overall calls were based on results for TA98 and TAMix only.

Agreement Statistic: a statistic (e.g., Cohen's Kappa) that is used to assess agreement between qualitative determinations (e.g., positive or negative) by two assessment methods e.g., miniaturised assay and standard assay.

Assay: a protocol for the bacterial reverse gene mutation test using a specific method or procedure which can be one of the standard procedures (pre-incubation or plate incorporation using 90-100 mm agar plates) or one of the miniaturised procedures. It should be noted that the terms 'assay' and 'test' are used interchangeably in the DRP.

Assessment: mutagenicity evaluation for any combination of assay, test chemical, bacterial strain and S9.

Bias Index: difference in the relative proportions of positive and negative calls for each of the individual assay formats being compared.

Call: the qualitative outcome of a mutagenicity evaluation, e.g., positive, negative, equivocal.

Dose or concentration: amount of test chemical used in the assay. Both terms are used for agar plate assays (standard, 6-well, 24-well) and are numerically identical and use the same units: µg/plate. Some experts avoid "dose", reserving that term for *in vivo* animal experiments, other experts avoid "concentration" reserving that term for units of µg/ml which is not useful unit for agar plate assays. Concentration is used for the fluctuation assays and refers to the concentration at the exposure step in µg/ml. Throughout the DRP, the term 'concentration' has mostly been used.

Dynamic range: the dynamic range of an assay is the ratio between the smallest and largest number of revertants which can be counted. A larger (wider) dynamic range often corresponds to an improved ability to detect small increases.

Equivocal: a call that cannot be deemed clearly positive nor clearly negative in one assay, i.e. the outcome of the experiment does not permit a call of positive or negative.

False Negative: a miniaturised assay negative call that incorrectly corresponds to a positive call on the standard assay.

False Positive: a miniaturised assay positive call that incorrectly corresponds to a negative call on the standard assay.

Mismatch: instances where a positive or negative call for an assessment based on a

miniaturised assay experiments is opposite to the result from the corresponding assessment in the standard assay results.

Mixed call: a series of assessments that reflect discordant results among multiple experiments in the same laboratory and/or multiple experiments in different laboratories, i.e., no consensus.

Mutagenicity: mutagenicity results in events that alter the DNA and/or chromosomal number or structure that are irreversible and, therefore, capable of being passed to subsequent cell generations if they are not lethal to the cell in which they occur, or, if they occur in germ cells, to the offspring. The bacterial gene mutation assays detect changes in the base sequence of certain genes; more specifically, substitution, addition or deletion of one or a few DNA bases that result in base-pair substitution or frameshift mutations.

Negative Predictive Value (NPV): the proportion of miniaturised assay negatives that are correct relative to the standard assay.

Overall call: qualitative evaluation based on responses for a series of specified test conditions (i.e., strains and S9).

Performance metrics: calculated values that reflect the performance of the miniaturised tests evaluated, i.e., Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Positive Rate and False Negative Rate.

Positive control: a chemical known to induce a positive response or the plate(s)/well(s) in an experiment containing that chemical as well as the bacteria and other components of the assay.

Positive Predictive Value (PPV): the proportion of miniaturised assay positives that are correct relative to the standard assay.

Prevalence Index: difference in the incidence of positive versus negative calls for the combination of the two assessment methods being compared.

Revertant: bacteria genetically reverted from histidine or tryptophan auxotrophy (i.e. only able to grow in a minimal culture medium that is supplemented with that amino acid) to wild-type, i.e., histidine or tryptophan prototrophy (i.e. able to grow in a minimal growth medium depleted in histidine or tryptophan)

S9 liver fraction: supernatant of liver homogenate after 9000g centrifugation, i.e., raw liver extract.

S9-mix: the mixture required to confer exogenous metabolic activity, i.e. the combination of the S9 liver fraction and the required enzyme cofactors.

Sensitivity: the proportion of standard assay positive responses that were correctly detected, i.e., also positive in the miniaturised assay.

Solvent/Vehicle control: both terms were used to refer to plates or wells receiving only the solvent or vehicle used to dissolve or suspend the test chemical, i.e., without any test chemical.

Solvent/Vehicle control counts: number of revertant colonies in the solvent/vehicle control cultures.

Specificity: the proportion of standard assay negative responses that were correctly detected, i.e., also negative in the miniaturised assay.

Standard assay: the bacterial reverse gene mutation assay using 90-100 mm agar plates and as described in TG 471

Standard bacterial strains: The 10 *Salmonella typhimurium* or *E. coli* strains listed in OECD TG471 and recommended for use in choosing the panel of 5 bacterial strains used to evaluate a chemical for mutagenicity. The TG notes that, with provided justification, other strains may be appropriate for use as well.

TAMix: an equimolar mixture of six *Salmonella typhimurium* strains (TA7001, TA7002, TA7003, TA7004, TA7005 and TA7006) used in the Ames II assay.

Test chemical: the term test chemical is generally used in OECD documents to refer to the substance being tested.

Top dose/concentration: the highest dose/concentration of test chemical used or recommended to be used in the assay.

Untreated control: cultures receiving no treatment (i.e., neither test chemical nor solvent) but which are processed the same way as the cultures receiving the test chemical.

Weak Responder: a mutagen that is difficult to detect because the maximum induction of revertants in any of the standard strains using the standard protocols is relatively small, albeit large enough to be reproducibly identified as a mutagen.

Introduction

2. The bacterial reverse gene mutation test (described in OECD Test Guideline 471; OECD, 1997) is the most widely used test for the detection of mutagenicity. The standard format consists in the agar plate incorporation method and/or the preincubation method. Several miniaturised versions of the bacterial reverse gene mutation test have been developed and are already used, particularly for early screening of new products in industry research and development. The potential advantages of these miniaturised versions include a significant reduction of the amount of test chemical needed to conduct the experiments including a reduced amount of animal-derived exogenous metabolic systems (i.e., S9-mix), and the possibility for simultaneous analyses of large number of samples, allowing an increase in throughput and a reduction of the necessary resources and cost. However, since the miniaturised bacterial reverse gene mutation tests are not described in any existing OECD Test Guideline, they do not benefit from Mutual Acceptance of Data (MAD). Therefore, the results from these miniaturised bacterial reverse gene mutation tests are currently not widely accepted by the regulatory agencies. In 2016, an OECD project was initiated to compile a detailed review paper (DRP) to address the following questions:

- What miniaturised bacterial reverse gene mutation tests are available?
- What are the most important characteristics of the different miniaturised bacterial reverse gene mutation tests compared to the standard OECD TG471 bacterial reverse gene mutation test?
- What are the knowledge gaps which would prevent acceptance of the methods for some or all regulatory uses of the assay?
- How can these knowledge gaps, if any, be addressed?
- What recommendations can be made with respect to the future use of the miniaturised bacterial reverse gene mutation tests?

3. In order to address the different questions, this DRP provides at first an overview of the available miniaturised bacterial reverse gene mutation tests. Next, the outcome of a retrospective analysis on existing data is presented. The aim of this analysis was to better understand the performance of each of the miniaturised tests compared to the standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997). To this extent, a call for data was made and the collected information was included in one central database. Using this database, the qualitative results obtained with the miniaturised versions were compared to those obtained with the standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997). A critical discussion of the outcome of the data analysis is also included in the DRP taking into account the limitations of the retrospective validation study.

4. It is important to note that, in contrast to other DRPs, the aim was not to directly revise OECD TG471 based on the outcome of the DRP, but instead, to provide recommendations on the future use and/or the additional work needed for each of the miniaturised bacterial reverse gene mutation assays.

Executive summary

5. The bacterial reverse gene mutation test described in OECD Test Guideline 471 is the most widely used *in vitro* test for the detection of mutagenicity. The standard format utilizes plate incorporation and/or preincubation method; a test mixture containing bacteria, test chemical, S9-mix, when required, and agar, is plated on 90- to 100-mm plates. Several miniaturised versions of the assay have been developed and are already in use, particularly for early screening of new products, as during research and development, large numbers of chemicals have to be tested that are often only available in low amounts. The primary advantage of these miniaturised versions is a significant reduction of the amount of test material needed to conduct the experiments; some may also allow simultaneous analyses of large number of samples, increasing throughput and reducing resources and cost. The aim of this Detailed Review Paper (DRP) was to evaluate the performance of several types of miniaturised assays, relative to the standard bacterial reverse gene mutation test specified in OECD TG471. Performance evaluations were based on a retrospective analysis of data provided by well-established genotoxicity testing laboratories. The purpose of the evaluations was to provide recommendations regarding future use of miniaturised assays, and/or to define additional work required to appropriately characterize the performance of each of the assays examined.

6. A survey was conducted to provide an overview of available miniaturised bacterial reverse gene mutation test data. A miniaturised assay was defined as an assay that requires less test chemical, and is characterised by a reduction in vessel size and/or format for treatment and scoring; by definition, the miniaturised assay uses less bacteria. The different miniaturised assays were subdivided based on their technical features; including the use of standard versus non-standard tester strains, the size of the plate for agar-based assays, and/or the use of liquid suspension (i.e., microfluctuation) assay formats. Based on these criteria, the following miniaturised assays were identified:

- Miniaturised tests using standard bacterial strains:
  - Miniaturised agar-plating assays (e.g., 6- and 24-well assays);
  - Miniaturised fluctuation assays (e.g., Ames MPF™)

- Miniaturised versions using non-standard bacterial strains:
  - Miniaturised agar-plating assays (e.g., test employing bioluminescent strains);
  - Miniaturised fluctuation assays (e.g., Ames II and bioluminescent tests in liquid format).

7. For each of these assays, the most important characteristics and differences relative to the standard format were discussed, as well as their perceived advantages, challenges, and reported historical performance. The survey revealed that four miniaturised assay versions are commonly used: the 6- and 24-well agar-based assays, the Ames MPF™ assay, and the Ames II assay.

8. At present, miniaturised assays are primarily used to screen substances in early stages of product research/discovery, and to test impurities at later stages; compound synthesis challenges often contribute to limited test chemical availability. In these circumstances, the standard bacterial reverse gene mutation test may not be feasible; consequently, a miniaturised test version is warranted as an alternative. For example, the ICH guideline for evaluating the mutagenic potential of impurities in pharmaceuticals (ICH, 2017), identifies limited circumstances under which a bacterial assay with proven high concordance to the standard assay could be justified. However, since the miniaturised assays are not described in any existing OECD Test Guideline, they do not benefit from the Mutual Acceptance of Data (MAD) agreement. Accordingly, the results from these miniaturised assays are currently not widely accepted by regulatory authorities.

9. Three primary scenarios for future use of miniaturised assay versions were identified:

- i) Use as fully-accepted alternatives to the standard assay;
- ii) Use as alternative tests when the standard format is not feasible;
- iii) Use only for test chemical screening.

In order to obtain insight regarding the most appropriate use context for each miniaturised assay, a retrospective analysis was performed using data provided by well-established genotoxicity testing laboratories. More specifically, to obtain the data for the retrospective performance analyses, two targeted calls were sent out to well-established laboratories worldwide, encouraging them to submit data generated with a miniaturised version of the bacterial reverse gene mutation assay using a standardised Excel spreadsheet. Data had to be entered in a separate row for each treatment concentration, but not for each replicate (i.e., only mean values for each concentration were entered). Submitters also had to include a single assessment call (i.e., negative, positive, equivocal) for each experiment for which, in one assay, various concentrations of the same test chemical were applied to a single strain-S9-mix combination. Data were received from over 20 laboratories; the data primarily represented four different miniaturised assays, i.e., the 6- and 24-well agar-based assays, and the Ames MPF™ and Ames II microfluctuation assays. After a thorough data curation process, a master dataset was compiled for the retrospective performance analysis. The final master dataset comprised 8727 assessments; including 429 test chemicals, of which 188 were coded. As the evaluation of the correspondence between the miniaturised versions of the bacterial reverse gene mutation assay and the standard assay (i.e., the point of reference) requires a single entry for each individual assessment (i.e. assay-test chemical-strain-S9-mix combination), assessments with multiple entries were isolated, and subsequently screened to determine a single consensus call. In cases where all calls were in agreement across the replicated entries, a single consensus call was inserted (e.g., positive, negative, equivocal). In cases where there was disagreement amongst the multiple calls for a given individual assessment, the term 'mixed' was inserted. Almost all assessments with disagreements amongst the multiple calls were permanently labeled as mixed, and the assessments subsequently excluded from the correspondence analyses described below. Exceptions were made for some mixed calls for positive controls listed in OECD TG471 (OECD, 1997).

10. The curated and appropriately-formatted dataset was used to examine binary correspondences between the qualitative outcomes (i.e., mutagenic or non-mutagenic) of the

selected miniaturised assays and the outcome of the standard assay. Standard assay data included data generated using the preincubation and plate incorporation test versions.

11. Different types of retrospective performance analyses were conducted:

- Correspondence analyses based on **individual assessments**, where an *assessment* was defined as test for a given assay-test chemical-strain-S9 combination. The analysis did not require that the assessment data were generated by the same laboratories, or that the miniature and standard assays were conducted concurrently.

- Correspondence analyses based on two types of **overall calls** i.e., response calls determined using the results obtained across a series of test conditions:
    - Overall calls determined using the responses in **all 5 bacterial strains** recommended in OECD TG471, i.e., (i) *S. typhimurium* TA98, (ii) *S. typhimurium* TA100, (iii) *S. typhimurium* TA1535, (iv) *S. typhimurium* TA97, TA97a or TA1537, and (v) *E. coli* WP2 or *S. typhimurium* TA102. The overall call designation did not differentiate between the different genotypes of *E. coli* WP2. A 5-strain overall call was deemed possible only if the results of 10 assessments were available, i.e., assessments in each of the 5 strains with and without S9;

    - Overall calls determined using responses on (i) *S. typhimurium* TA98 and (ii) *S. typhimurium* TA100 only. A **2-strain overall call** was deemed possible only if the results of 4 assessments were available, i.e., assessments in each of the 2 strains with and without S9.

  For an overall positive call, a positive response for any of the relevant strain-S9-mix combinations was deemed sufficient; designation of a negative overall call required a negative response for all of the relevant strain-S9-mix combinations. Overall calls for the Ames II assay were based on results for *S. typhimurium* TA98 and TAMix only.

- Additional analyses investigated strain-specific correspondences, correspondences based only on standard preincubation assessments, and correspondences based only on standard plate incorporation assessments (i.e., without preincubation).

12. For all initial analyses, equivocal calls provided by the data submitters were excluded. Follow-up analyses investigated changes in the results if the equivocal calls provided by the data submitters were alternatively denoted positive or negative. Also excluded were *mixed* calls, i.e., instances whereby the results of multiple assessments were discordant and a consensus call was not possible.

13. For each analysis, 2x2 contingency tables were constructed, and a set of performance metrics and agreement statistics were employed to evaluate correspondence between the miniaturised assay outcome and the standard assay outcome.

14. The outcome of the different analyses can be summarised as follows:

- Overall, there was a good agreement between the miniaturised assay calls and the standard assay calls.
- Although the **individual assessment** analyses provide useful insights into the performance of the miniaturised bacterial reverse gene mutation assays examined, results based on the 5-strain overall call comparisons might be deemed most relevant with respect to the use of the assays for mutagenicity screening. Each test chemical included in the 5-strain overall call comparison was necessarily analysed in a miniaturised assay with the complete set of bacterial strains described in OECD TG471 (OECD, 1997), both with and without S9-mix metabolic activation. Unfortunately, there were too few 5-strain overall data to conduct meaningful analyses for the 24-well plate, Ames MPF™ and Ames II assays. However, with the exception of the 24-well assay, sufficient data was available to examine the correspondences associated with 2-strain overall calls for the miniaturised assays investigated.
- **6-well plate assay:** As compared with the other miniaturised assays examined, the dataset included 6-well assay data from more laboratories and for more test chemicals. It was the only assay for which enough data were available to conduct a 5-strain overall call analysis. Overall agreement for the correspondence analysis, as measured by Prevalence-and-Bias-Adjusted–Kappa (PABAK), was almost perfect for the 62 test chemicals for which data were available. Performance measures such as sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) were high, with concomitantly low values levels of false positives and false negatives. Compared with the 5-strain overall call analysis, agreement and performance were as good or better for the 2-strain overall call analyses on the data of 97 test chemicals; the narrower confidence intervals are not surprising since the 2-strain overall call dataset is larger. However, it should be noted that the additional test chemicals included in the 2-strain overall call analyses were not mutagenic in either the 6-well plate or standard assays. Thus, some performance measures (e.g., false negatives, sensitivity) were necessarily based on essentially the same overall call data. With respect to agreement and performance, the overall call analyses were similar to the individual assessment results; although, for the overall call assessments, there was a small but consistent improvement in each measure. The individual assessment results were based on 913 assessments of 121 test chemicals. Repeating the correspondence analysis using only standard assay plate incorporation results had no appreciable effect on any of the measures of agreement or performance. Strain specific analyses resulted in increased sensitivity, and fewer false negatives, for *S. typhimurium* strains TA98, TA100, and the combined *E. coli* WP2 and *S. typhimurium* TA102 strains, with a corresponding decrease in sensitivity for *S. typhimurium* TA1537 and TA1535. However, PABAK and the other performance measures did not vary among the strains except for an increase in false positive rate for the combined *E. coli* WP2 and *S. typhimurium* TA102 strains.
- **Ames MPF™ assay:** The second largest portion of the dataset is associated with the Ames MPF™ assay. For the 2-strain overall call analysis (47 test chemicals), most performance measures were similar to that obtained for the 2-strain overall call analysis with the 6-well plate assay, except for a notable decrease in sensitivity, and

corresponding increase in false negatives. Although PABAK was in the "almost perfect" range, it was at the lower end of that range. Marginal overlap with the PABAK 95% confidence interval for the 6-well assay evaluation suggests a significantly lower level of agreement. For the **individual assessment** analysis (674 assessments of 125 test chemicals), there was a similar pattern of agreement and performance, except for sensitivity and false negatives. However, these findings should be interpreted in the context of the noted dataset limitations (see Paragraphs 15 and 152). When standard assay data were segregated into preincubation and plate incorporation results, there were small but consistent improvements in all measures of agreement and performance. The strain-specific analyses did not reveal any consistent patterns of differences relative to the individual assessment results.

- **Ames II assay:** Overall agreement for the Ames II assay was "almost perfect" for the 2-strain overall call analyses based on data for 32 test chemicals. However, as for Ames MPF™ assay, the sensitivity was somewhat lower, and associated false negatives higher, relative to the miniaturised agar-plating assays. Patterns such as improvement in the overall call analysis relative to the **individual assessments** analysis (173 assessments of 97 test chemicals), and better correspondence with standard assay preincubation data, were similar to that observed for the Ames MPF™ assay. However, this comparison should be interpreted with caution in light of the wider confidence intervals around the measures of agreement and performance; the wider intervals are consistent with the smaller number of test chemicals in the dataset. For example, there were 32 test chemicals for the Ames II assay 2-strain overall call analysis, compared with 47 test chemicals for the Ames MPF™ assay analyses.
- **24-well plate assay:** Overall call analyses could not be conducted for the 24-well plate assay, the miniaturised assay with the smallest dataset. In the analysis of **individual assessments** (152 assessments of 42 test chemicals), overall agreement was at the lower end of the "almost perfect" range, and lower than agreement for the 6-well plate assay. Most measures suggested performance similar to that observed for the 6-well plate assay, with the exception that specificity and NPV were slightly lower, and false positives higher. There was a consistent decline in agreement and performance when the 24-well plate assay assessments were compared to standard assay plate incorporation data only. However, it should be noted that the differences were greatest for measures with the widest confidence intervals.
- There were relatively few equivocal calls in any of the datasets. Denoting equivocal calls as positive or negative did not reveal any meaningful changes in the performance metrics and agreement statistics.

15. **The most important limitations** of the retrospective performance analyses arise because each correspondence analysis was necessarily based on a **different set of test chemicals**; thus, the agreement statistics and performance metrics for each of the analyses reflect test performance for the set of test chemicals investigated. Extrapolation of the results from one set of chemicals to any other set, including the set used for another assay format within this DRP should thus be approached with extreme caution. Characteristics of the datasets which limit the robustness of the conclusions can be summarised as follows:

- Uncertainty in the metrics and statistics increases when fewer chemicals are tested. For example, there were fewer chemicals and thus more uncertainty in results for Ames II and the 24-well assays relative to the results for 6-well and Ames MPF™ assays.
- Several datasets were asymmetric with respect to the standard assay calls, i.e., a **high prevalence of non-mutagens**. This contributed to increased NPV values, reduced PPV values, and increased false positive rates.
- For some miniaturised assays, the collected data included **low numbers of revertants for solvent controls**; indeed, for some experiments there were no revertants in any of the solvent control wells. Such low solvent control counts were noted for strains TA1535, TA1537 and TA98 in the Ames MPF™ and 24-well assays. This may contribute to a restricted ability to detect weak responses.
- ***Mixed*** **call instances were excluded from the analyses**. Mixed calls likely indicate poor intra- and inter-laboratory reproducibility, thus removal can contribute to an overestimate of miniaturised assay performance. Mixed calls might be expected for some test chemicals on some strains; indeed, a panel of 5 bacterial strains is required since some are designed to preferentially detect certain classes of mutagens. There is also a possibility that differential detection of mutagens by the standard assay and the miniaturised assays could contribute to over-estimates of agreement.
- **Experimental protocols and test interpretation criteria for each of the assays considered may have varied between laboratories**; each standard or miniaturised test was considered an independent experiment, whether or not they were conducted at the same time or even in the same laboratory.
- In some instances, sample sizes are small (<100); consequently, some assay performance statistics have wide confidence intervals. This means that care is warranted when drawing conclusions from the retrospective performance analyses; indeed, the revealed patterns should not be over-interpreted. Nevertheless, the results obtained are useful for exploring the degree of correspondence between the standard assay and the miniaturised assays examined; moreover, discordant results provide important insights regarding assay performance.
- In some instances, a large fraction of the substances included in the retrospective performance analyses were coded compounds (i.e., unidentified proprietary compounds). Coded compounds are useful because they represent a diversity of chemicals likely to be evaluated using the assays. However, the absence of structural information limits the ability to use these data to investigate further the spectrum of the test chemicals included in the analyses or to evaluate in more depth performance metrics (e.g. sensitivity and false negatives) and to compare them across assay formats.

16. Considering the outcome of the retrospective correspondence analyses, and their associated limitations, the following recommendations are made:

- Several miniaturised versions of the bacterial reverse gene mutation assay have been developed and are already in use, particularly for early screening of new products. Despite the current use of the methods for specific reasons, there is currently no scientific consensus on the general regulatory use of these methods. However, under specific conditions (e.g. when limited test chemical is available such as for some impurities or metabolites), data from one or more of the miniaturised versions of the bacterial reverse gene mutation test may be among the alternatives to contribute to an overall safety assessment.
- The retrospective analysis described in this DRP provides valuable insights into qualitative aspects of the performance of these miniaturised assays compared to the standard Ames test. This may open the way for conducting further confirmatory work in order to facilitate the regulatory acceptance of these methods in the future.
- The concordance analysis shows a good agreement between the miniaturised assay calls and the standard assay calls. However, it presents several limitations and does not consider quantitative aspects, (particularly those) related to identification of mutagens which produce weakly positive responses in the test. For these reasons, none of the four evaluated miniaturised bacterial reverse gene mutation tests can at this stage be recommended as an alternative for the standard OECD TG471 bacterial gene mutation test (OECD, 1997), nor can any of the four miniaturised bacterial reverse gene mutation tests be explicitly included in the current OECD TG 471.
- However, because there are areas of application for miniaturised bacterial reverse gene mutation assays and because the retrospective study results are promising, it is recommended that, if additional resources allow, further work be conducted in the future, including the quantitative exploitation of the data collected in the context of the retrospective analyses and, as appropriate, a prospective validation study.
- Follow-up quantitative analyses of the collected concentration-response data could thus provide insight regarding sensitivity, call criteria, and inter- and intra-laboratory variability for each of the assays examined and would provide the strong basis to determine the relevance to conduct prospective validation studies.‘

1 Overview of the available bacterial reverse gene mutation tests

1.1. Introduction

1.1.1. Standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997)

17. In order to facilitate comparison of the different miniaturised bacterial reverse gene mutation tests with the standard OECD TG471 bacterial reverse gene mutation test, the most important characteristics of the latter are first briefly discussed.

18. The current OECD TG471 bacterial reverse gene mutation test (OECD, 1997) describes the use of *Salmonella typhimurium* and *Escherichia coli* tester strains. Several testing systems to detect mutagenicity have been developed in parallel. Bridges (1972) reported a system for screening chemical mutagens based on their ability to cause base-substitution mutations in an ochre stop codon, reverting a tryptophan-requiring bacterial strain to prototrophy. He used *E. coli* B/r WP2, which Witkin isolated from *E coli* UV B/r with and without inactivation of the DNA repair gene *uvrB*, to increase sensitivity (Witkin, 1956). His original method described a qualitative “spot test”, in which mutant colonies were observed along the concentration gradient caused when the test chemical migrated from a piece of filter paper at the center of an agar plate. He also described a “quantitative” version in which exposure in liquid medium was followed by dilution and plating on both selective plates (no tryptophan) and non-selective plates to measure the true mutation frequencies. Sensitivity of the quantitative version was later increased by distributing aliquots of the bacteria into 50 tubes, an adaptation of the original fluctuation test of Luria and Delbrük (Green et al. 1976).

19. Almost at the same time, Ames and his colleagues at the University of California in Berkeley described a similar screen using *Salmonella* (Ames, 1971). Ames’ initial method employed strains *S. typhimurium* TA1535 and TA1537 as part of a panel of strains designed to be sensitive to either base substitution or the frame-shift mutations, respectively. The strains in the panel were selected from among over 1000 *Salmonella typhimurium* LT-2 mutants used by Ames and Hartman during a collaborative effort to map the histidine operon (Hartman *et al.* 1971). Ames further developed the assay by adding a metabolic activation step using S9-mix, the supernatant obtained by centrifuging at 9,000 x g homogenised livers from either a human cadaver or from rodents pre-treated with phenobarbital to induce metabolic enzymes (Ames, 1973). Both systems employed agar plates to screen for bacteria with reverting mutations. In 1983, the OECD published the initial test guidelines for the *S. typhimurium* (TG471) and *E. coli* (TG472) tests. The two tests were combined into the current version of TG471 in 1997. Although both *S. typhimurium* and *E. coli* tester strains are thus described in OECD TG471, the TG is often referred to as the ‘Ames test’. The TG also mentions fluctuation assays but describes the plate-based assays in more detail. In this DRP, the term ‘the standard bacterial reverse gene mutation test’ will be used to refer to the bacterial reverse

gene mutation test using 90—100 mm agar plates described in the current OECD TG471 (OECD, 1997).

20. The pre-existing mutation in bacterial tester strains selected for use in the test inactivates a gene involved in the synthesis of an essential amino acid, either histidine (*S. typhimurium*) or tryptophan (*E. coli*). Consequently, these tester strains can only grow in a minimal culture medium that is supplemented with that amino acid (auxotrophy). A limiting amount of histidine or tryptophan allows all exposed bacteria to replicate. After exposure to a mutagen, a mutation can be induced during replication which reverts the strain back to the wildtype phenotype and to grow in a minimal growth medium depleted in histidine or tryptophan (prototrophy). The initial rounds of replication exhaust the supply of the required amino acid, after which only revertants can continue to replicate. Depending on the strain used, the original mutations responsible for the gene inactivation are substitutions of individual base pairs or frameshift mutations by base pair addition or deletion (Hamel et al., 2016). Prototrophy can be restored by a variety of mutations (**Table 1**). The most common mutations are in (or near) the site of the original mutation although mutations elsewhere in the bacterial genome (e.g., tRNA suppressor mutations which allow the tRNA to recognize the stop codon and insert an amino acid in its place, resulting in a functional protein (Kupchella et al., 1994)) can also lead to the prototrophic phenotype. Sensitivity to mutagens is enhanced by alterations of DNA repair pathways via inactivation of endogenous genes (*uvrA* or *uvrB*) and/or addition of exogenous genes on plasmid pKM101.

21. To screen for a broad selection of potential mutagens, OECD TG471 (OECD, 1997) recommends conducting the test using a panel of strains with varying types of inactivating mutations and mechanisms of DNA repair disruption. The OECD TG471 (OECD, 1997) recommends a panel of 5 strains chosen from among 9 commonly used strains. The most important characteristics of the bacterial strains used in the standard bacterial reverse gene mutation test are summarised in **Table 1**. The strains are further referred to as ‘standard tester strains’.

Table 1. Genotypic information for the bacterial strains recommended in the standard OECD TG 471 bacterial reverse gene mutation test (OECD, 1997).

Assembled from various reviews, DeMarini et al. (1998), DeMarini (2000), Mortelmans and Zeiger (2000), Ohta et al. (2002).

	Strain	Genotype			Reverting Mutations[1] (most common)
		DNA repair, etc.	*Target gene*	*Original mutation*	
1	*S. typhimurium* TA1535	Δ*uvrB*, *rfa*	*hisG46*	GAG → GGG	G:C → T:A, G:C → A:T, G:C → C:G, T:A → G:C
2	*S. typhimurium* TA100	Δ*uvrB*, *rfa*, pKM101	*hisG46*		
3	*S. typhimurium* TA1537 or	Δ*uvrB*, *rfa*	*hisC3076*	CCCC→ CCCCC	Δ C:G
	S. typhimurium TA97 or TA97a	Δ*uvrB*, *rfa*, pKM101	*hisD6610* *hisO1242*	CCCCC→ CCCCCC	Δ C:G
4	*S. typhimurium* TA98	Δ*uvrB*, *rfa*, pKM101	*hisD3052*	CCGCGCGCGG → CCGCGCGG	+ C:G or Δ CG:GC
5	*E. coli* WP2 uvrA pKM101 or	Δ*uvrA*, *uvrB*$^+$, pKM101	*trpE65*	CAA→ TAA	T:A → C:G, T:A → C:G, T:A → A:T, A:T → T:A, A:T → C:G, GC → TA, GC → AT
	E. coli WP2pKM101 or	*uvrA*$^+$, *uvrB*$^+$ pKM101			
	E. coli WP2 uvrA or	Δ*uvrA*$^+$, *uvrB*$^+$			
	E. coli WP2 or	*uvrA*$^+$, *uvrB*$^+$			
	S. typhimurium TA102	Δ*hisG*, *rfa*, *uvrB*$^+$, pKM101, pAQ1	*hisG428* (on pAQ1)		

[a] (rfa): This mutation leads to a defective lipopolysaccharide (LPS) layer that coats the cell surface, making the bacteria more permeable to bulky chemicals

[b] (uvrB): The uvrB deletion mutation eliminates the accurate excision repair mechanism, thereby allowing more DNA lesions to be repaired by error-prone DNA repair mechanisms. The deletion through the biotin gene makes the bacteria biotin dependent

[c] (pKM101): This R factor plasmid enhances chemical and UV-induced mutagenesis via an error-prone recombinational DNA repair pathway. The plasmid also confers ampicillin resistance

22. The standard bacterial reverse gene mutation test can be performed according to the plate incorporation method or the preincubation method. Both methods are described briefly in OECD TG471 and the most important characteristics of the assays are summarised in **Table 2** (OECD, 1997). In the plate incorporation method, bacteria are exposed to the test chemical[2] by mixing suspensions generally consisting of 100 µL of bacterial overnight culture

[1] Reverting mutations: mutations causing reversion to the wild type phenotype observed by DNA sequencing of the target genes as well as tRNA genes from revertant colonies.

[2] The term "test chemical" is used in this Detailed Review Paper to refer to what is being tested and is

at about 10^9 bacteria per mL, 50 or 100 µL test chemical and 500 µL exogenous metabolic activation system (experiment with S9-mix) or buffer (experiments without S9-mix) with 2.0 mL of molted overlay agar followed by immediate plating onto minimal agar medium. In the preincubation method, the suspensions consisting of bacterial cells, test chemical and S9-mix or buffer are first incubated for usually 20 or 30 min (in a volume of generally 0.65 or 0.7 mL) before being mixed with a molted overlay agar (generally resulting in a total volume of 2.7 mL) and plated onto minimal agar medium. The recommended top concentration for soluble non-cytotoxic test chemicals is 5 mg/plate or 5 µL/plate for liquid test chemicals. For non-cytotoxic test chemicals that are not soluble at 5 mg/plate or 5 µL/plate, one or more concentrations tested should be insoluble in the final treatment mixture but the precipitate should not interfere with the scoring (e.g. precipitation of material on the plate may look like a colony during automatic counting and contribute to a false positive call). Test chemicals that are cytotoxic already below 5 mg/plate or 5 µL/plate should be tested up to a cytotoxic concentration. For both the plate incorporation and the preincubation method, after two or three days of incubation at 37°C, revertant colonies are counted and the number of revertant colonies on the plates treated with the test chemical are compared to the number of spontaneous revertant colonies on solvent control plates. Positive control plates using known mutagenic compounds appropriate to the strain (e.g., inducing base pair substitutions or frameshift mutations) as well as promutagens requiring metabolic activation by S9-mix, are included in each experiment to check that the test system is performing as expected. To ensure that the integrity of the bacterial strains has not been compromised, OECD TG471 (OECD, 1997) recommends that the phenotypic characteristics of the stock culture of bacteria and the amino acid requirements of the tester strains are checked.

23. The protocols described in OECD TG 471 can detect a wide variety of test chemicals, i.e., the test has a broad applicability domain. Chemicals within or outside the applicability domain can be classified a variety of techniques including chemicals defined by chemical structural classes per se or due to properties of the chemicals such as physico-chemical properties (e.g. poor solubility) or based on the response when tested using the standard protocols. For example, the test guideline notes some “special cases”, (azo-dyes and diazo compounds, gasses and volatile compounds, and glycosides) for which adaptations of the protocol are recommended. The test guideline lists also certain chemical structures that may be better assessed using the pre-incubation protocol. Some chemical classes generally considered to be within the applicability domain may be difficult to detect when using protocols other than the standard protocols. This may include chemicals that are “weak responders”, i.e., the maximum induction of revertants in any of the standard strains using the standard protocols is relatively small. Nitrosamines and aromatic rings with amino or nitro substituents are examples of structural classes containing many weak responders. In many cases these compounds are promutagens which are activated by metabolic enzymes with weak activity (or not present) in standard rat S9 preparations. Note that some weak responders (low mutagenic potency) induce cancer in rodents at low doses (high carcinogenic potency). There is no comprehensive list of structural alerts for weak responders. However, ensuring adequate sensitivity to known weak responders from a variety of chemical structural classes will be an important criterion in the overall performance assessment of each miniaturised assay. These

not related to the applicability of the test methods to the testing of mono-constituent substances, multi-constituent substances and/or mixtures.

are an important subset of mutagens identified by the bacterial reverse gene mutation test using the standard protocol. In this DRP they will be called “weak responders”.

24. In the OECD TG471 (OECD, 1997) describing the standard bacterial reverse gene mutation test, while there is general language concerning data interpretation, no specific criteria for a positive result are included. In practice, many laboratories employ empirically derived “fold rules” in which a positive result is determined when the fold increase (the ratio of mean counts at once concentration versus the mean counts in the concurrent solvent control) exceeds a certain value. The most common empirical rules are a two-fold increase for most strains, ranging up to three-fold for strains with lower background counts. Ames and Bridges each mentioned the use of fold rules in early descriptions of the tests but also indicated that statistical methods needed to be considered. Until today, the use of fold rules has endured in spite of critical analyses challenging their applicability (Claxton et al., 1987; Cariello and Piegorsch, 1996; Hamada et al., 1994; McCann *et al.* 1984; Levy et al., 2019). One impediment to developing statistical tests is that count data in the bacterial reverse gene mutation test is often reported to deviate from normal or binomial distributions, deviations which are strain-specific (e.g., see Kim and Margolin, 1999). The most popular statistical tests assume the data fit one of those distributions. A second impediment is the large number of comparisons built into the protocol (triplicate plates for ≥5 doses in each of 5 strains compared with 5 solvent controls in each of the two metabolic conditions). With so many comparisons within a single experiment, random variation may be expected to result in relatively frequent instances of count data exceeding statistical limits/intervals of 5 or even 1%. Whatever the causes, statistical tests often have resulted in descriptions of statistical significance for data of the bacterial reverse gene mutation test which were perceived to be at odds with expert judgements regarding biological significance. Statistical tests that correct for multiple comparisons, addressing the second impediment have been proposed, among which the Dunnetts’s test have been recommended (e.g., Kirkland, 1994) but not adopted widely. OECD TG471 (OECD, 1997) recommends evaluating the data to determine whether there is a concentration-related increase and emphasizes that first consideration should be biological relevance. A more recent publication proposes considering criteria similar to that proposed during recent updates of other genetic toxicology test guidelines (Thybaud et al., 2017) including using the distribution of historical solvent control data in evaluating the study outcome (Levy et al., 2019).

25. Besides the plate incorporation and the preincubation method included in OECD TG471 (OECD, 1997), other procedures have been described to perform the bacterial reverse gene mutation test. In the early eighties, Kado and colleagues developed a modified version of the preincubation test to detect mutagenic metabolites in small sample volumes such as urinary samples obtained from animals treated with test chemicals or non-concentrated human urine (Kado et al., 1983). In this procedure, a concentrated suspension containing 10 times more bacterial cells (~10^9 cells/incubation tube) as compared to the standard bacterial reverse gene mutation test is preincubated with 3 times less volume of test chemical and S9-mix compared to the standard preincubation bacterial reverse gene mutation test. The concentrated bacterial suspension is obtained by centrifugation of the overnight bacterial cultures. With the exception of the higher concentration of bacteria and the smaller volumes used to perform preincubation, all steps are comparable to the standard bacterial reverse gene mutation test. For this reason, the method is also referred to as the ‘micropreincubation’ or the ‘microsuspension assay’. Several authors reported an increased sensitivity of the microsuspension assay compared to the standard bacterial reverse gene mutation test (Kado

et al., 1983; Agurell & Stensman, 1992; Watanabe et al., 1995; Cerná et al., 1999). Due to the higher sensitivity, the microsuspension assay has been extensively used for environmental sample testing, especially when only small sample volumes are available (Kado et al., 1986; Umbuzeiro et al., 2004; Crebellia et al., 2005; Di Giorgio et al., 2011; Alves et al., 2016).

Table 2. Most important characteristics of the standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997) and different miniaturised versions (vc: vehicle control; pc: positive control). [3]

Test method	Bacterial strains	Initial number of bacteria	Metabolic activation (S9 % v/v)	Method	Recommended top test concentration for soluble non-cytotoxic test chemicals	Scoring	Number of replicate plates or wells
Standard OECD TG471 bacterial reverse gene mutation test (OECD, 1997)							
Standard	Panel of 5 strains, 1 strain per plate	10 x 10^7/plate	5-30%	(Pre) incubation: 0.5-0.7 mL [bacteria + S9-mix + test chem.] Final top agar: 2.0-3.7 mL Surface plate: 64-80 cm^2 (with d = 9-10 cm)	5 mg or 5 µL/plate	Colony counts/plate *(automation possible)*	3 2-3 vc 2-3 pc
Miniaturised agar plating tests using standard strains							
6-well plate	Panel of strains, 1 strain per well	2 x 10^7/well	5-10%	(Pre) incubation: 0.14 mL [bacteria + S9-mix + test chem.] Final top agar: 0.54-0.65 mL Surface well: 9.5 cm^2 (with d = 3.5 cm)	1 mg/well	Colony counts/well *(automation possible)*	3 2-6 vc 2-6 pc
24-well plate	Panel of strains, 1 strain per well	0.5 x 10^7/well	5-10%	(Pre) incubation: 0.04 mL [bacteria + S9-mix + test chem.] Final top agar: 0.13-0.35 mL Surface well: 1.9 cm^2 (with d = 1.6 cm)	0.25 mg/well	Colony counts/well	2-3 2-6 vc 2-6 pc
Microfluctuation tests using standard tester strains							
Ames MPF™	Panel of strains, 1 strain in 48 wells per replicate	1-2 x 10^7/replicate well[a]	10-30%	Exposure: 0.25 mL/well Final volume: 0.05 mL/well (*automation possible*)	5 mg/mL	Fraction of wells displaying pH change due to growth (*automation possible*)	3[b] 3 vc 3 pc
Microfluctuation tests using non-standard tester strains							
Ames II test	1 or a mixture of strains in 48 wells per replicate	1-2 x 10^7/replicate well	10-30%	Exposure: 0.25 mL/well Final volume: 0.05 mL/well (*automation possible*)	5 mg/mL	Fraction of wells displaying pH change due to growth (*automation possible*)	3[b] 3 vc 3 pc

[a] According to the Instructions for Use of the Ames MPF™ assay, cells are not counted: the overnight culture is added in fixed volume ratio if OD-based criteria are met. The initial number of cells for the Ames MPF™ assay should be in the same order of magnitude of the Ames II assay.
[b] The number of replicates for the microfluctuation tests refers to the number of replicate wells in the exposure plate (see also Figure 2).

---

[3] The characteristics included in the table are mostly based on the protocols provided by the labs that submitted data for the retrospective validation study. Consequently, some parameters (e.g. S9 %) may vary when these miniaturised assays are performed in other labs.

1.1.2. Grouping of miniaturised bacterial reverse gene mutation tests

26. Several miniaturised bacterial reverse gene mutation tests have been developed, mainly to reduce the amount of test chemical. These miniaturised tests are further characterised by a reduction in the size of the format for scoring and consequently, also in the number of bacteria exposed to the test chemical. The Expert Group (EG) did not include the microsuspension assay as described above in this DRP with the miniaturised versions evaluated here because plating is still done in the standard format. Depending on the type of the miniaturised bacterial reverse gene mutation test, one or more of the following additional differences compared to the standard bacterial reverse gene mutation test can be identified:

- the size of the agar plate;
- the use of liquid media instead of agar for measurement of revertant bacteria (microfluctuation tests);
- the use of non-standard tester strains.

27. In this DRP, the available miniaturised versions were subdivided based on their technical aspects (**Figure 1**). First, a distinction was made between miniaturised bacterial reverse gene mutation tests using a panel of standard tester strains described in OECD TG 471 (OECD, 1997) and those using non-standard or mixtures of strains. Because it is easier to compare the performance of the standard tester strains using altered protocols to the performance of the same strains using the standard protocol, emphasis was placed on these comparisons. Both the miniaturised versions using standard and non-standard tester strains can be further subdivided in two large groups:

- **Miniaturised agar plating tests**: these tests are based on the same principle as the standard plate incorporation or preincubation bacterial reverse gene mutation test, but use smaller agar plates or wells, resulting in a significant (and approximately proportional) reduction of the number of bacteria and the amount of test chemical needed, without substantially modifying the other experimental parameters. Revertant frequencies are calculated based on the number of progeny bacteria which grow into a colony on the plate or well.

- **Microfluctuation tests**: in these tests, bacteria are maintained in suspension throughout the whole experiment from treatment to the detection of revertant bacteria. Bacteria are exposed to the test chemical in liquid medium (with or without S9-mix) in a multi-well format and/or tubes before dilution and transfer into microtiter plates for detection of revertant bacteria. Mutant frequencies are calculated based on the fraction of wells in which the pH of the medium (and colour of the wells) has changed as a result of the prototrophic bacterial growth.

28. Although this subdivision is convenient for the most commonly used miniaturised bacterial reverse gene mutation tests, it is important to note that some assays are difficult to assign to one of the subgroups as multiple parameters differ from the standard bacterial reverse gene mutation test.

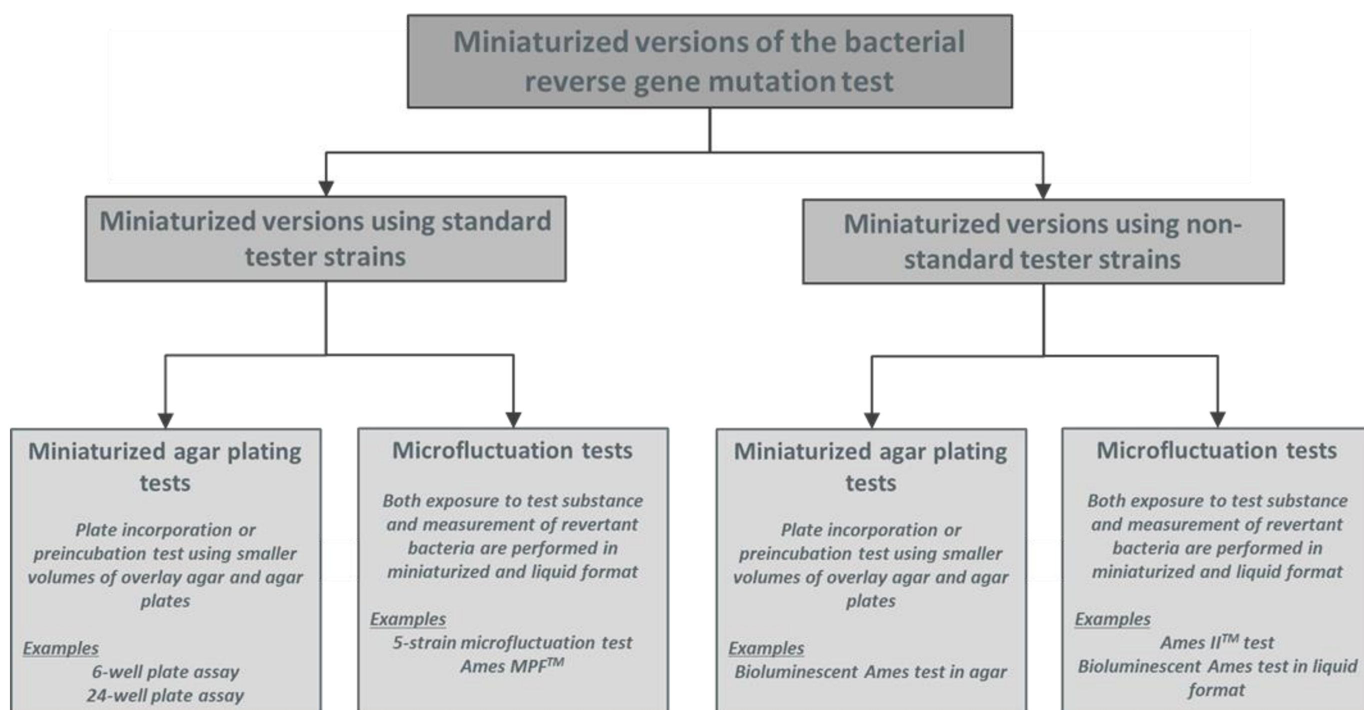

Figure 1. Schematic overview of the different types of miniaturised bacterial reverse gene mutation tests.

1.2. Miniaturised agar plating tests using standard tester strains

1.2.1. Introduction

29. In order to be able to screen test chemicals for which only small quantities are available early in the research and development process, a scaled-down version of the standard bacterial reverse gene mutation test, with no modifications of experimental conditions other than a proportional reduction of the volumes of all components, was developed in the mid-nineties. The assay, also called the ‘Miniscreen’, was performed in square 25-well Petri dishes[4] of which each compartment was filled with 2 mL Vogel-Bonner minimal medium. Two versions using two (i.e., *S. typhimurium* TA98 and TA100) (Brooks, 1995) or three (i.e., S. typhimurium TA98, TA100 and TA102[5]) (Burke et al., 1996) standard tester strains have been described. Due to their low spontaneous mutation rate, *S. typhimurium* TA1535 and TA1537 were considered less appropriate for the test. Some differences other than the number of strains existed between the versions of Brooks and Burke et al. (e.g., different top concentration and incubation time), but overall, the tests were carried out similarly. In short, 100 µL of S9-mix (or buffer for tests in the absence of S9-mix), 20 µL of test chemical or control in a suitable solvent and 500 µL of supplemented top agar (containing 25 µL of bacterial overnight culture) were added to each well of the 25-well Petri dishes. The concentrations tested ranged from 0.25 to 200 µg/well. After they had solidified, plates were inverted and

[4] A 25-well Petri dish is a sterile, square Petri dish with 25 compartments, each measuring 1.8 cm" with a maximum capacity of 5 mL.

[5] *S. typhimurium* TA102 was afterwards replaced by *E. coli* WP2(pKM101) and WP2uvrA(pKM101).

incubated for 48 (Brooks, 1995) or 72 (Burke et al., 1996) hours at 37°C, followed by manual counting of the number of revertant colonies. Colony counts were initially evaluated statistically using Dunnett's test, but the “two-fold rule” was used in some later applications of the test. Both the statistical analysis and the two-fold rule also relied on evidence of a concentration-dependent response. Overall, a good concordance was observed between the results obtained with the Miniscreen and the standard bacterial reverse gene mutation test. However, the number of test chemicals (i.e., 14 known mutagens and 9 novel test chemicals) was limited. One important technical limitation of this original ‘Miniscreen’ assay was that to obtain quantitative results, counting of the colonies had to be done manually (by eye) under a magnifying glass because of the greatly reduced size of the revertant colonies and the square shape of the wells. Subsequent modifications to this 25-well plate assay were made by several laboratories such as a switch to 6-well or 24-well plates with round wells, the addition of more standard tester strains and use of the preincubation method. At present, the original ‘Miniscreen’ is no longer used but the 6-well and 24-well plate assays derived from this assay are still widely applied. Data from the original Miniscreen have not been included in this DRP.

1.2.2. Miniaturised agar plating test in 6-well plates

History

30. A modified version of the original ‘Miniscreen’ was developed in 6-well cell-culture dishes, allowing automated scoring (Diehl et al., 2000). This ‘Modified Miniscreen’ assay used the bacterial strains *S. typhimurium* TA98, TA100 and TA102, and *E. coli* WP2 *uvrA* and was conducted in standard 6-well culture dishes. Diehl et al. only used *S. typhimurium* TA1535 for test chemicals that showed a positive result in this strain in the standard bacterial reverse gene mutation test. Test chemicals were studied up to 2000 µg/well. As in the standard bacterial reverse gene mutation test, culture dishes were incubated for 48 hours at 37°C after hardening of the top agar. An important improvement compared to previous miniaturised versions of the bacterial reverse gene mutation test consisted of the automatic counting of revertant colonies, except when precipitation was present. Diehl et al. considered an increase in the number of revertant colonies higher than 2-fold indicative of a positive result.

31. Further modifications to this 6-well plate assay were made by several laboratories such as the inclusion of additional standard tester strains and a further reduction of the top concentration to 1000 µg/plate. For example, Flamand et al (2001) described a 6-well plate assay performed in all *S. typhimurium* strains of the standard bacterial reverse gene mutation test (TA98, TA100, TA102, TA1535, TA1537 and TA1538) including those with a low spontaneous rate of revertant colonies and referred to it as the ‘Mini Mutagenicity Test (MMT)’. According to the authors, the MMT was performed strictly according to the protocol of the standard bacterial reverse gene mutation test except that all reagents were divided by a factor of five.

32. Since their development, 6-well plate assays have been used mainly as screening test in pharmaceutical industry (Escobar et al., 2013) and have been offered by contract research organizations (Pant et al., 2016).

Most important characteristics

33. Overall, the test is performed in the same way as the standard bacterial reverse gene mutation test, but in a smaller well format. For screening purposes, testing is often only done in *S. typhimurium* TA98 and TA100, but in principle, the 5–strain battery recommended in TG471 can be used. The metabolic activation system and the top and bottom agar mixes are the same as in the standard test, although reduced to smaller amounts. Like in the standard bacterial reverse gene mutation test, plates are incubated for 48 hours to allow DNA damage to be fixed and revertant colonies to grow. Colony counting is used as endpoint. Scoring of the revertant colonies can be done manually or with aid of colony counters. Both the plate incorporation and the preincubation method can be performed in the 6-well plate format. More details of the assay are included **Table 2**.

Important differences with standard bacterial reverse gene mutation test

- Fewer bacteria exposed: total number of initial bacteria in contact with the test chemical: in the 6-well plate assay between 2.0 - 2.5 x 10^7 bacteria are used versus 1 x 10^8 bacteria in the standard bacterial reverse gene mutation test;
- Lower colony counts: the number of spontaneous revertant colonies is lower in the different bacterial strains due to reduction in the initial number of bacteria plated (Pant et al., 2016);
- Top concentration for soluble non-cytotoxic test chemicals: 1000 µg/well instead of 5000 µg/plate.

Perceived advantages

- Reduced amount of test chemical needed: the extent of the reduction will depend on the characteristics of the 6-well plate assay (i.e. number of strains, number of replicates, top concentration,…). In general, the amount of test chemical used is 5-fold lower compared to a standard bacterial reverse gene mutation test with the same design and a top concentration of 5000 µg/plate.
- Reduced reagent expenses: reduced proportionately to 1/5th of that required for the standard assay assuming the same basic design such as strains, concentrations, replicates, activation conditions, etc. (e.g. 0.1 mL S9-mix vs 0.5 mL S9-mix in standard plates).
- Less space needed: one 6-well plate can contain all the replicates for a concentration level with and without S9-mix. Consequently, the assay requires less counter and hood space. Furthermore, less space is taken up in the incubators.

Challenges

- The small size of the 6-well plates can complicate enumerating colonies:
  - ✓ Crowding may reduce the accuracy for plates with many colonies. The plate surface area is reduced ~8-fold vs 5-fold reduction in viable bacteria plated so multiple superimposed colonies may be counted as a single colony. In addition,

the dynamic range (range from the smallest to largest number of colonies that can be counted) may also be reduced;
  - ✓ Identifying and restreaking to differentiate revertants from wild-type microcolonies or precipitate may be more difficult, although these are rarely encountered by most laboratories.
- Because of the smaller format, sensitivity may be altered:
  - ✓ The number of replications occurring before depletion of histidine (or trp) may have an impact on the sensitivity. The impact of miniaturisation on bacterial growth dynamics has not been studied. Growth is slower further from the plate surface and may be altered near plate edges. There are small but potentially significant variations in the relative ratios. Compared to the standard assay the 5-fold lower amounts of bacteria and test chemical are distributed in a volume of top agar which is 5.7-6.9-fold lower (3.7 mL vs 0.54-0.65 mL, see Table 2). In addition the surface area of 6-well plates is 7- or 8-fold lower that the plates used for the standard assay (usually 90 or 100 mm plates, respectively).
  - ✓ Reduced volumes, especially in preincubation version, may complicate the testing of insoluble chemicals or selectively reduce metabolic activation by enzymes sensitive to interference by DMSO.
- Lower spontaneous revertant colony plate counts may alter sensitivity:
  - ✓ For the *S. typhimurium* TA1535 and TA1537 strains, the low end of the 95% tolerance intervals may be as low as 1-2 colonies. Consequently, test chemical wells can have revertant counts within tolerance intervals representing multiple fold increases over concurrent controls. This wider dynamic range, in theory, may increase sensitivity.
  - ✓ On the other hand, the ability to detect a small increase in revertants may be reduced when there are too few revertants in the solvent control. Increasing the number of replicate wells for the solvent controls provides some compensation for this loss. As discussed in the review of Escobar et al. (2013), it is important to note that different experiences have been reported regarding the use of strains with low spontaneous rates of revertant colonies. Some laboratories considered that these strains can be successfully used, provided that the results are interpreted based on sound scientific judgment and experience, taking into account the reproducibility of the data and the typical control intervals. In contrast, other laboratories pointed out that the use of these strains is not practicable due to the too low solvent control values, even lower than those already observed in the standard bacterial reverse gene mutation test. Especially, the *S. typhimurium* TA1537 strain has low spontaneous revertant colony counts in the 6-well plate assay. For this strain, *S. typhimurium* TA97 has been proposed as an alternative based on its comparable or even higher sensitivity together with its higher background.
  - ✓ Criteria for positive results when using *S. typhimurium* TA98, *E. coli* WP2 *uvrA* may need to be adapted. Most labs apply the 2-fold increase criterion to results in these strains in both the 6-well and standard assays. Based on lower counts, a larger increase (e.g., 3-fold) may be more appropriate.

- May not be suited for volatile test chemicals: the volatility of several compounds can cause an increase in the reversion frequencies of adjacent, untreated wells (Wilson & Cariello, 1997). However, this problem can be overcome by using adequate plate sealers.

Historical performance

34. Diehl et al. (2000) compared the results obtained with their 6-well plate assay with those obtained in the standard bacterial reverse gene mutation test for known mutagenic chemicals and proprietary test chemicals. In total, 91 test chemicals were included in this comparative study. Importantly, only one standard bacterial reverse gene mutation test and one 6-well plate assay were conducted for each test chemical with only one plate or well for each concentration and activation condition. Furthermore, some of the test chemicals were only tested using *S. typhimurium* TA98 and TA100 whereas for others the *S. typhimurium* TA98, TA100 and TA102, and *E. coli* WP2*uvrA* strains were used. Based on the obtained results, the authors concluded that there was a 100% concordance between both tests when taking into account the overall call thus based on the results obtained in the different strains. Also, according to the authors, strain *S. typhimurium* TA1535 was considered to allow detection of moderate and strong mutagenic responses. In the review of Escobar et at. (2013), results of two comparative studies between the 6-well plate assay and the standard bacterial reverse gene mutation test were included. The first analysis was done by Novartis based on the results obtained with 63 proprietary test chemicals. Testing in the 6-well plate assay was only done in *S. typhimurium* TA98 and TA100. Also in this study, the predictivity of the 6-well plate assay for the standard bacterial reverse gene mutation test was found to be 100%. Furthermore, with the exception of two cases, both tests were concordant with respect to metabolic activation condition. Another comparative study was done by AbbVie and included 56 test chemicals. Only for one out of the 56 test chemicals, the outcome of the standard bacterial reverse gene mutation test was not predicted by the 6-well plate assay. Later, another 12 test chemicals were tested at AbbVie in both the 6-well plate assay and the standard bacterial reverse gene mutation. Concordance for these 12 test chemicals was shown to be 100% (Nicolette et al., 2015 + personal communications). It should be noted that these comparative analyses are heavily weighted towards test chemicals that are negative in the 6-well plate assay, as the miniaturised version has mainly been used as a screening test. In a more recent study (Nicolette et al. 2018), the 6-well plate assay and the standard bacterial reverse gene mutation test were compared at two different laboratories using a more balanced set of 24 test chemicals. Eighteen out of the 24 chemicals were tested concurrently in the 6-well assay and the standard bacterial reverse gene mutation test whereas for the remaining 6 test chemicals, standard data were collected from the literature. Testing across the two formats resulted in 100% concordance in overall mutagenicity judgement. Additionally, the strain-to-strain concordance was also high, i.e., 94%. In 2020, Egorova et al. published the results of a comparative study with 14 technical grade active ingredients of pesticides using the standard bacterial reverse gene mutation test and the 6-well plate assay. Comparison of the standard bacterial reverse gene mutation test and the 6-well plate assay in this study resulted in 98% of concordance across five strains and conditions (± S9-mix).

1.2.3. Miniaturised agar plating test in 24-well plates

History

35. A 24-well plate assay using ~10 mg of test chemical and the five-strain set as required by OECD TG471 [*S. typhimurium* TA97a, TA98, TA100, TA1535 and WP2*uvrA*(pKM101)] was described by Schlosser et al. (2007) and Wells et al. (2008). According to the authors, these studies were essentially performed according to OECD TG471 (OECD, 1997) using the preincubation method at a 1/20th scale. Since its development, this 24-well plate assay has been used primarily as screening tool in pharmaceutical industry (Escobar et al., 2013; Proudlock and Evans, 2016) and has been offered by contract research organizations (Pant et al., 2016).

Most important characteristics

36. Overall, the test is performed in the same way as the standard bacterial reverse gene mutation test, but in a smaller well format. Screening may be done only in *S. typhimurium* TA98 and TA100, but it generally is more common to use five standard tester strains required by OECD TG471 (OECD, 1997). The metabolic activation system and the top and bottom agar mixes are the same as in the standard bacterial reverse gene mutation test, although they are proportionately reduced. Like in the standard assay, the plates are incubated for 48 – 72 hours to allow DNA damage to be fixed in reverse mutations and revertant colonies to grow. However, due to the small well size, the scoring of the revertant colonies has to be done manually. Both the plate incorporation and the preincubation method can be performed in the 24-well plate format. More details of the assay are included in **Table 2**.

Important differences with standard bacterial reverse gene mutation test

- Fewer bacteria exposed: total number of initial bacteria in contact with test chemical: in the 24-well plate assay approximately 0.5 x 10^7 bacteria are used versus 1 x 10^8 per plate in the standard bacterial reverse gene mutation test;
- Lower colony counts: the number of spontaneous revertant colonies is lower in the different bacterial strains due to reduction in the initial number of bacteria plated.
- Top concentration for soluble non-cytotoxic test chemicals: 250 µg/well instead of 5000 µg/plate.
- Number of replicates: the test chemical is usually evaluated in duplicate wells. A higher number of replicates may be used for the solvent controls. In the standard assay, all treatments are performed in triplicate plates.

Perceived advantages

- Reduced amount of test chemical needed: the extent of the reduction will depend on the characteristics of the 24-well plate assay (i.e., number of strains, number of replicates, top concentration,…). In general, the amount of test chemical used is 20-fold lower compared to the standard bacterial reverse gene mutation test with a top concentration of 5000 µg/plate.

- Reduced reagent expenses: reduced proportionately to 1/20th of that required for the standard assay (assuming the same basic design – strains, concentrations, replicates, activation conditions, etc).
- Less space needed: two 24-well plates can contain all the replicates for all concentration levels and controls, with and without S9-mix, for a single tester strain. Consequently, the assay requires less counter, hood, and incubator space.

Challenges

- The small size of the 24-well plates can complicate enumerating colonies:
  - ✓ Crowding may reduce the accuracy for plates with many colonies. The plate surface area is reduced ~40-fold vs 20-fold reduction in viable bacteria plated so multiple superimposed colonies may be counted as a single colony. In addition, the dynamic range (range from the smallest to largest number of colonies that can be counted) may also be reduced;
  - ✓ Identifying and restreaking to differentiate revertants from wild-type microcolonies or precipitate may be more difficult, although these are rarely encountered by most laboratories.
- Because of the smaller format, sensitivity may be altered:
  - ✓ The number of replications occurring before depletion of histidine (or trp) may have an impact on the sensitivity. The impact of miniaturisation on bacterial growth dynamics has not been studied. Growth is slower further from the plate surface and may be altered near plate edges. There are small but potentially significant variations in the relative ratios. Compared to the standard assay the 20-fold lower amounts of bacteria and test chemical are distributed in a volume of top agar which is 10.5-28-fold lower (3.7 ml vs 0.13-0.35 ml, see Table 2). In addition, the surface area of 24-well plates is 34- or 42-fold lower that the plates used for the standard assay (usually 90 or 100 mm plates, respectively).
  - ✓ Reduced volumes, especially in the preincubation version, may complicate testing of insoluble chemicals or selectively reduce metabolic activation by enzymes sensitive to interference by DMSO or other solvents.
- The preincubation version of the miniaturised agar plating test in 24-well plates method is complicated by the small volumes used.
- Lower spontaneous revertant colony plate counts may alter sensitivity:
  - ✓ The ability to detect a small increase in revertants may be reduced when there are too few revertants in the solvent control. Increasing the number of replicate wells for the solvent controls provides some compensation for this loss.
  - ✓ For the *S. typhimurium* tester strain TA1535, the low end of the 95% tolerance interval may be <1 colony/well. Consequently, the evaluation criteria for the various strains have been adjusted to rely on a minimum 2- or 3-fold increase AND a defined threshold or net increase in revertants/well to account for sampling error and biological relevance. These criteria may not be equivalent to the criteria used in the standard assay for the same strains
  - ✓ Tester strain *S. typhimurium* TA97 or TA97a have been substituted for *S. typhimurium* TA1537 to overcome the problem of even lower spontaneous revertant frequencies in that tester strain. Again, as discussed in the review of

Escobar et al. (2013), it is important to note that different laboratories report different experiences regarding the use of strains with low spontaneous rates of revertant colonies.

- May not be suited for volatile test chemicals: the volatility of several compounds can cause an increase in the reversion frequencies of adjacent, untreated wells (Wilson & Cariello, 1997). However, this problem can be overcome by using adequate plate sealers.

Historical performance

37. Several laboratories have compared the results obtained using the 24-well plate assay with those obtained in the standard bacterial reverse gene mutation test for known mutagenic/non-mutagenic test chemicals (Schlosser et al., 2007; Wells et al., 2008; Sawant et al., 2016; Smith et al., 2017). The different groups evaluated various numbers of test chemicals (11 – 20 test chemicals per publication, with some overlap), using a relatively standardised design that included at least five tester strains, multiple concentration levels ±S9-mix, and the appropriate positive and solvent controls. In some cases, the comparisons were concurrent and direct, while in others the 24-well data were compared to previously published results. In addition, limited concentration-response comparisons were performed for 31 test chemicals using only the tester strain/activation combinations that were expected to elicit a positive response (Proudlock and Evans, 2016). All authors reported almost 100% concordance for the overall positive/negative calls using the two assay formats. The only outlier was oxazepam (a known *in vitro* clastogen, *in vivo* mutagen, and rodent carcinogen), which was positive in the 24-well plate assay +S9-mix but variously reported to be negative or weakly positive in the standard bacterial reverse gene mutation test. Some studies also made strain/activation and proportional concentration-by-concentration comparisons and reported approximately 90 – 95% concordance for those analyses as well.

38. In the review of Escobar et al. (2013), results of one comparative study between the 24-well plate assay and the standard bacterial reverse gene mutation test were included. This analysis was done by Amgen based on the results obtained with 11 known mutagens/non-mutagens. Testing in the 24-well plate assay was only done in *S. typhimurium* TA97a, TA98, TA100 and TA1535, and the first phase using four known strong mutagens was only done with or without S9-mix as required. In the second phase, compounds with low or moderate mutagenic activity were studied both in absence and in presence of a metabolic activation system. The predictivity of the 24-well plate assay for the standard bacterial reverse gene mutation test was found to be 100% for all 11 test chemicals evaluated, and the concentration-by-concentration concordance was 81 – 100% for the four strong mutagens tested (e.g., the results at 5000 µg/plate in the standard assay were compared those at 250 µg/well in the 24-well plate assay).

1.3. Microfluctuations tests using standard tester strains

1.3.1. History

39. The original version of the fluctuation test for bacterial reverse mutation screening has already been described in the 1970s (Green et al., 1976). The principle of the test is the same

as in the liquid preincubation standard test. However, instead of counting colonies on an agar plate, the frequency of revertant bacteria is obtained by adding diluted fractions to a fixed number of wells and counting the fraction of wells in which growth is observed. Initially, growth was detected by observation of turbidity in the wells. To reduce time, costs and the amount of test chemical needed, and to allow the test to be adapted to high throughput screening, the original fluctuation assay was modified to allow automation of plating the exposed cells in selective media. In these tests, bacteria are incubated with the test chemical in medium with S9-mix (or buffer) in 24-well plates before inoculation of the content of each well into 48 wells of the 384-well microtiter plates for the detection of revertant bacteria. The frequency of revertant bacteria is determined by the number of wells that have bacterial growth. Bacterial respiration releases acidic metabolites which lower the pH of the growth medium which can be detected using a pH indicator dye. Several commercial kits containing ready-to-use media and performance-tested *Salmonella* tester strains are available for performing microfluctuation tests; some kits provide only a selection of standard tester strains, whereas in others, all five standard tester strains are included. Among the different commercial kits, the Ames fluctuation protocol of Xenometrix (Ames MPF™ assay) is most widely used, in particular for early mutagenicity screening of early drug candidates, chemicals, cosmetic products, herbal extracts, environmental samples, food contact material, and medical devices (*e.g.*, implants). Furthermore, it has also been adapted to test native non-concentrated water samples using of a 10-fold concentrated exposure medium (Flückiger-Isler and Kamber, 2014). Concordance with standard assay response has been evaluated using various sets of chemicals (Gee et al., 1998, Flückiger-Isler and Kamber, 2012, Spiliotopoulos and Koelbert, 2020).

1.3.2. Most important characteristics

40. The microfluctuation assay Ames MPF™ is a liquid 384-well microplate format modification of the standard bacterial reverse gene mutation test. Standard tester strains required by OECD TG471 (OECD, 1997) can be run in the format. Both the quantity of test chemical and the volume of the metabolic activation system needed are significantly reduced. Like in the standard bacterial reverse gene mutation test, the plates are incubated for 48 hours to allow DNA damage to be fixed in reverse mutations and revertant bacteria to grow. Instead of colony counting, the read-out is colorimetric by using indicator medium. Both handling and analysis can be automated.

41. The experimental protocol, detailed by Flückiger-Isler and Kamber (2014), can be summarised as follows:

- Exposure is performed in triplicate in 24-well plates. The 24-well exposure plates include negative and positive controls (2 wells per replicate) and 6 concentrations of the test chemical (6 wells per replicate) (see **Figure 2**). Bacteria of freshly prepared overnight cultures are exposed to of the test chemical, the positive or solvent control in the 24-well plate in medium containing sufficient histidine (or tryptophan, for the *E. coli* strains) to support approximately two cell divisions in the volumes per well mentioned in **Table 3**. In each well, 0.010 mL of test chemical, the solvent and positive controls is added. Due to the 25-fold dilution in the exposure medium, a 125 mg/mL stock solution must be prepared to achieve a top concentration of 5 mg/mL. Addition of Aroclor 1254 or phenobarbital/ naphthoflavone-induced rat liver S9-fraction is generally performed with a 1.5% or 4.5% S9-fraction.

- The 24-well exposure plates are incubated at 37°C for 90 min, with shaking at 250 rpm. After the exposure step, 2.6–2.8 mL indicator medium is added to each well of the exposure plates. The content of each well of the 24-well plate is subsequently transferred into 48 wells of a 384-well plate aliquoting 50 µL/well. Therefore, 48 wells of a 384-well plate correspond to one well of the 24-well exposure plate (where exposure has occurred).

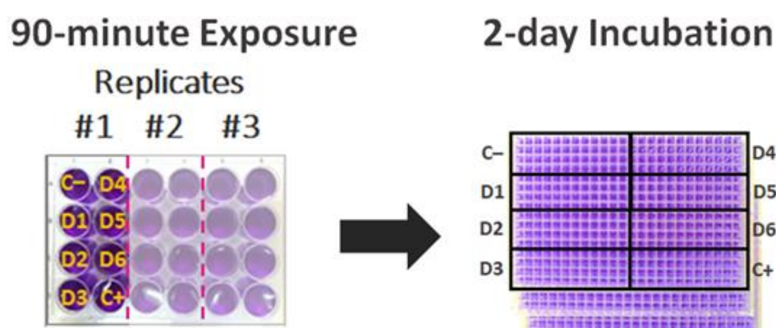

Figure 2. Experimental set-up of the Ames MPF™ assay. C-: negative control; C+: positive control; D1-6: test chemical dose 1 to 6.

- Within two days, cells which have undergone the reversion to *His* (or *Trp*, for the *E. coli* strains) will grow into colonies. Positive wells are those that have an indication of colour change from purple to yellow (Flückiger-Isler and Kamber, 2014). Metabolism by the bacterial colonies reduces the pH of the medium, changing the colour of that well. This colour change can be detected visually or by microplate reader. The number of wells containing revertant colonies are counted for each concentration (manually or with aid of a microplate reader) and compared to a solvent (negative) control.
- Analysis of the results: The fold increase of revertants relative to the solvent control is determined by dividing the mean number of positive wells at each concentration by that of the solvent control baseline. The solvent control baseline is derived from the mean number of positive wells in the solvent control plus 1 standard deviation. If the baseline is less than 1, the value is set to 1 for calculation. A fold increase greater than two times the baseline level is generally considered as an alert. Multiple concentrations with a positive (concentration–dependent) response will lead to the test chemical being classified as a clear positive. A test chemical is classified negative when no response greater than two times the baseline is recorded (Flückiger-Isler and Kamber, 2014).

Table 3. Volumes per well (in µL) for the Ames MPF™ assay (o.n. culture: overnight culture).

		In the absence of S9-mix		In the presence of S9-mix		
	Dilution	o.n. culture	Exp. Medium[a]	o.n. culture	Exp. Medium[a,b]	S9-mix
TA98	1:10	25.0	215.0	25.0	177.5	37.5
TA100	1:20	12.5	227.5	12.5	190.0	37.5
TA1535	1:10	25.0	215.0	25.0	177.5	37.5
TA1537	1:10	25.0	215.0	25.0	177.5	37.5
TA97	1:45	5.6	234.4	5.6	196.9	37.5

EC Combo[c]	*E. coli* WP2 *uvrA* *E. coli* WP2 [pKM101]	1:14.3 1:33.3	17.5 7.5	215.0	17.5 7.5	177.5	37.5
	E. coli WP2 *uvrA*	1:15	16.7	223.3	16.7	185.8	37.5
	E. coli WP2 [pKM101]	1:20	12.5	227.5	12.5	190.0	37.5
	E. coli WP2 uvrA[pKM101]	1:30	8.3	231.7	8.3	194.2	37.5

[a]: The exposure medium for the Salmonella and E. coli strains have different compositions.
[b]: As a result of the toxicity of some batches of S9, Xenometrix AG provides the S9 together with a "S9 booster solution", which is mixed with the Exposure Medium at a ratio 1:667, to protect the strains S. typhimurium TA100 and TA1537.
[c]: "EC Combo" is the mixture of E. coli WP2 uvrA and E. coli WP2 [pKM101], grown individually and mixed just prior the exposure step.

1.3.3. Important differences with standard bacterial reverse gene mutation test

- Total number of initial bacteria in contact with test chemical: 1-2 x 10^7 versus 1 x 10^8 per plate in the standard bacterial reverse gene mutation test.
- Lower counts: spontaneous revertant well counts tend to be low (1 to 10 wells, depending on the strain). These numbers are not directly comparable to background colony counts in the standard bacterial reverse gene mutation test.
- Top concentration for soluble non-cytotoxic test chemicals: 1250 µg/well in the 24-well exposure plate instead of 5000 µg/plate.
- Number of replicates: the test chemical is usually evaluated in triplicate. A higher number of replicates may be used.
- Bacteria are exposed in liquid phase in a relatively small volume (250 µL), resulting in a relatively higher effective concentration of test chemical. This is similar to the preincubation version of the standard assay.
- Colorimetric readout: mutant frequencies are calculated based on the fraction of wells in which the pH of the medium changed as a result of the progeny bacteria growth.
- Toxicity of the test chemical is usually measured directly by checking an increase of the brilliance of the purple medium as compared to the solvent control due to cell lysis and absence of bacterial cells, or by inspection of the wells with a microscope. In the standard bacterial reverse gene mutation test, toxicity is inferred from reductions in revertant counts or qualitative observations of background lawns. Precise toxicity measurements can be useful for evaluation of concentration-response results that include a small increase in revertants, as well as a decline in revertants due to toxicity. Although not commonly done, toxicity can be precisely measured by post-exposure sampling and culturing of bacteria; either from a liquid suspension or agar plates. In some cases, these more precise measurements can help distinguish a mutagen from a non-mutagen. Methods for sampling bacteria from bacterial lawns on agar plates have been described in the literature; however, they are technically challenging (Prival, 2001). With respect to the liquid cultures used in fluctuation assays, similar measurements, although not reported in the literature, would likely be technically easier.
- The entire experiment takes place in a liquid phase eliminating the effects of the semi-solid agar. It is possible that these effect (e.g., alterations in migration of the test chemical or nutrients, bacterial growth or S9 metabolism) may be beneficial or adverse.

1.3.4. Perceived advantages

- Reduced amount of test chemical needed: the extent of the reduction will depend on the characteristics of the microfluctuation assay (i.e., number of strains, number of replicates, top concentration,…). In case tests are done in triplicate with a top concentration of 5000 µg/mL in five strains, the amount of test chemical used is 4-fold lower compared to the standard bacterial reverse gene mutation test with a top concentration of 5000 µg/plate.
- Higher throughput: although when not automated, the Ames MPF™ is still a 'low throughput' assay, the throughput is approximately twice as fast as compared to the standard bacterial reverse gene mutation test. Importantly, the assay can be partly or fully automated and processed by pipetting stations, thus requiring less hands-on time and resulting in a higher throughput.
- Read-out: the assay has a readout in 48-well sections of a 384-well plate. Typically, triplicate plates with six sample dilutions, negative and positive controls are scored by eye in approximately 5 min. The read-out is automatable for high throughput screening.
- Reduced reagent expenses: considerably less S9-mix is used (13× less) and plastic ware (reduced disposal costs).
- An advantage of the fluctuation assay methodology is that unlike agar plate tests it allows determination of the true mutation frequency. This is because non-revertant bacteria cannot be quantitatively extracted from agar pates after exposure to the test chemical.
- Less space needed: two 24-well exposure plates and 6 x 384-well plates contain all the replicates for 6 concentration levels and controls, with and without S9-mix, for a single tester strain. Consequently, the assay requires less incubator space and causes less plastic waste.
- No microcolonies are formed which can be wrongly interpreted as revertants.
- Although commercial kits like the Ames MPF™ are available, the microfluctuation tests using standard tester strains can also be performed without these kits.

1.3.5. Challenges

- Difficulty to compare the test concentrations with those used in the plate incorporation version of the standard bacterial reverse gene mutation test. Comparison with the pre-incubation assay is relatively easy (Flückiger-Isler and Kamber, 2012).
- The exposure concentrations, while higher than in the plate incorporation assay, are lower than in the preincubation assay which was developed because those high concentrations were thought to be needed for some types of test chemicals.
- Toxicity assessment by the decrease of positive wells relative to the solvent control and by an increase of the brilliance of the purple medium as compared to the solvent control (cell lysis, absence of bacterial cells) is not directly comparable to toxicity assessment for the agar plate assays.

- Colored test chemicals: Interference of colored test chemicals with the colorimetric read-out is possible.
- High spontaneous revertant counts may result in an invalid result due to the narrow dynamic range based on the top limit of wells considered per replicate (48).
- Low spontaneous revertant well counts may alter sensitivity:
    - ✓ Use of strains with very low spontaneous revertant colony plate counts: the spontaneous revertants for the tester strains *S. typhimurium* TA98, TA1535 and TA1537 can be as low as 0–1.
    - ✓ The ability to detect a small increase in revertants may be reduced when there are no or too few revertants in the solvent control wells. Increasing the number of replicate wells for the solvent controls provides some compensation for this loss.
    - ✓ The fold-increase-based evaluation criteria cannot be directly compared to colony counts on agar plates and may need to be adapted. Statistical criteria might have to be considered.
- Commercial assay including proprietary elements (e.g., amount of histidine present in the media).
- Like the miniaturised agar plating assays, microfluctuation assays may not be suited for volatile test chemicals: the volatility of several compounds can cause an increase in the reversion frequencies of adjacent, untreated wells (Wilson & Cariello, 1997). However, this problem can be overcome by using adequate plate sealers.

1.3.6. Historical performance

42. Several validation studies comparing the Ames MPF™ with the standard bacterial reverse gene mutation test have been performed.

- In the first validation study of 1998 with base-specific tester strains, Gee et al. (1998) included also the traditional frameshift strains *S. typhimurium* TA98 and TA1537 for comparing the liquid microplate format with published data of the standard preincubation method. There was an overall concordance of 84 % (21/25) and 94 % (18/20) in the *S. typhimurium* TA98 and TA1537 strains, respectively. The liquid format appeared to be more sensitive to pick up mutagenicity.
- In a later study, Flückiger-Isler and Kamber (2012) tested 15 equivocal to weakly positive chemicals selected from the National Toxicology Program (NTP) database concurrently in the Ames MPF™ and the standard preincubation method. Thirteen of the 15 test chemicals showed concordant results in both tests.
- Recently, 61 test chemicals selected from the updated recommended list of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity assays (Kirkland et al., 2016) were tested in the Ames MPF™ assay using up to five OECD TG471-compliant (four *S. typhimurium* and *E. coli*) strains (OECD, 1997). The data generated with the *S. typhimurium* strains showed a strain-by-strain concordance higher than 90% with the standard bacterial reverse gene mutation data found in the scientific literature. For the *E. coli* strains, a thorough comparison with the standard assay was not possible due to the lack of literature data (Spiliotopoulos and Koelbert, 2020).

- Rainer et al. (2021) evaluated the concordance between the Ames MPF™ and the standard bacterial reverse gene mutation test, each performed in only two tester strains (i.e., *S. typhimurium* TA98 and TA100), for 21 test chemicals based on overall calls. Discordant results were obtained for two test chemicals (i.e., sodium azide and benzo[a]anthracene), resulting in a concordance 90% or higher.

1.4. Miniaturised agar plating tests using non-standard tester strains

1.4.1. Miniaturised version of the microsuspension method

43. In order to improve the concordance with the standard bacterial reverse gene mutation test and to simplify the dosing procedure, modified versions of the method developed by Kado et al. (1983) have been described. In these modified versions, the preincubation is performed in multi-well plates (e.g. 96-well plates) instead of tubes and parameters such as bacterial cell number and S9-mix concentration have been optimised (Muster et al., 2000; Escobar et al., 2013). In these tests, plating is still done in the conventional Petri dishes (90 x 15 mm), and consequently, these assays are not considered as miniaturised versions of the bacterial reverse gene mutation test within the context of this DRP. However, in 2018, Zwarg et al. described a miniaturised version of the microsuspension assay using 12-well microplates for the plating instead of the conventional Petri dishes. Overall, the principle of the miniaturised bacterial reverse gene mutation test of Zwarg et al. (2018) is the same as for the microsuspension assay described by Kado et al. (1983), but 12-well plates are used for plating instead of 90-100 mm diameter plates. As a result, the volume of test sample needed can be further reduced. The assay was developed with bacterial strains with low (TA1538), medium (TA98), and high (YG1041) spontaneous frequencies, although other strains may be used as well. The total number of bacteria treated is approximately 0.3-1.25 x 10^8 per well, thus comparable to 1 x 10^8 per plate in the standard bacterial reverse gene mutation test. Plates are incubated for 66 hours to allow revertant colonies to become visible and colony counting is used as endpoint. Scoring of the revertant colonies is done manually (Zwarg et al., 2018).

44. As for the other miniaturised versions of the bacterial reverse gene mutation test, the test chemical requirements can be reduced (i.e., 20-40 times) compared to the standard assay (assuming the same basic design – strains, concentrations, activation conditions, etc). Also, the volume of S9-mix can be decreased significantly (by a factor 40) compared to the standard bacterial reverse gene mutation test (when same concentration of S9-mix is used). However, it should be noted that so far, there is only limited experience with the assay. Hence, the amount of data received in response to the data call was too limited for this miniaturised bacterial reverse gene mutation test to be included in the retrospective data analysis.

1.4.2. Miniaturised agar plating tests using bioluminescent tester strains ('Bioluminescent Agar test')

45. The bioluminescent Ames test in agar is not only performed in a smaller format (i.e., 24-well format) than the standard bacterial reverse gene mutation assay, it also uses bioluminescent derivatives of the standard tester strains. These genetically engineered standard *Salmonella* tester strains TA98, TA100, TA1535, TA1537 and *E. coli* strain WP2*uvrA*pKM101 express the *lux(CDABE)* operon from *Xenorhabdus luminescence* and are

therefore able to produce luciferase and the fatty acid reductase substrate. Cleavage of the substrate by luciferase results in emission of bioluminescence which is measured as a sensor for the detection of revertant colonies. Histidine-dependent cells will starve over time due to the lack of histidine in the medium, which will make them incapable to maintain the bioluminescent phenotype. In contrast, the histidine-independent revertant cells will, under the same conditions, be capable of maintaining luminescence forming small luminescent colonies, easily detectable via a photon counting camera (Aubrecht et al., 2007; Escobar et al., 2013).

46. Advantages are comparable to those reported for the 24-well plate assay. Moreover, revertant colonies are easy to detect by their phosphorescence and can be scored using a custom-built system resulting in a higher throughput. Small bioluminescent colonies that are not visible macroscopically can be easily visualised using a photon counting camera. Furthermore, association of the state of starvation or energy depletion with the bioluminescent phenotype enables a relatively easy assessment of cytotoxicity of test chemicals and simplified differentiation of histidine-dependent microcolonies (non-revertants) arising as a consequence of cytotoxicity and colonies of histidine-independent cells (true-revertants) (Aubrecht et al., 2007).

47. Similar challenges as those reported for the 24-well plate assay (e.g., sensitivity, use of strains with low spontaneous revertant colony plate,…) apply for the bioluminescent agar test. Furthermore, the assay uses bacterial strains that are not compliant with the standard bacterial reverse gene mutation test. In addition, whereas *S. typhimurium* TA98 and TA100 have been well-validated in the assay, strains *S. typhimurium* TA1535, TA1537 and *E. coli* WP2uvrApKM101 have not often been used and should be utilised with more caution (Escobar et al., 2013). Moreover, at present, this miniaturised bacterial reverse gene mutation test is not often used anymore. During the calls for data, no information was received for the bioluminescent agar test and consequently, this assay was not included in the retrospective data analysis.

1.5. Microfluctuation test using non-standard tester strains

1.5.1. Ames II test

History

48. The Ames II™ assay was developed as a predictive screening assay for genotoxicity in Dr. Bruce Ames' laboratory at the University of California-Berkeley. Several modifications to the standard bacterial reverse gene mutation test were made to increase the throughput of the assay and to reduce the amount of test chemical needed including the use of new tester strains and the application of the fluctuation format (see 1.3). The new set of six *S. typhimurium* (TA7001 to TA7006) strains, each carrying a unique missense mutation in the histidine biosynthetic operon, was constructed to more easily detect basepair substitutions (Gee et al, 1994). The resulting microfluctuation test was designated 'Ames II' (Xenometrix Inc., Boulder, CO, USA). The tester strain TA98, which detects small deletions, additions and suppressions of the frameshift mutations, has also been included for routine use in the Ames II test (Gee et al., 1998). Since its development, the Ames II assay has been used mainly as screening test

in pharmaceutical industry (*e.g.*, Braun, 2001; Gervais et al., 2003; Lorge et al., 2007). Concordance with standard assay response has previously been evaluated (Flückiger-Isler et al., 2004; Kamber et al., 2009).

Most important characteristics

49. The Ames II assay is a 384-well microplate liquid format modification of the standard bacterial reverse gene mutation test. In this assay, the frameshift mutations are detected by the traditional *S. typhimurium* strain TA98 and base-pair substitutions by a mixture of the six *S. typhimurium* strains specifically engineered for the assay (**Table 4**). Each strain carries a different missense mutation in the histidine operon that is designed to revert uniquely to one of the six possible base substitution combinations causing transitions or transversions in base sequence. Additional modifications have been made to these mutants to improve the sensitivity of their reversion (*uvrB*, *rfa*, plasmid pKM101). The six strains (TA7001, TA7002, TA7003, TA7004, TA7005 and TA7006) are combined into an equimolar mixture called TAMix and treated as if they were an individual strain. The TAMix strain has a lower spontaneous reversion frequency compared to the standard *S. typhimurium* tester strains which makes them more compatible with a microfluctuation format. These two strains (TA98 and TAMix) are plated in triplicate 384-well plates. Wells are determined to be revertants if the indicator medium undergoes a colour change from purple to yellow or a colony is clearly visible in the well. In case a positive result is obtained, the type of mutation can be identified by performing tests with the individual strains to obtain more detailed information about the mutation spectrum, if that information is desired.

Table 4. Genotypes of TA98 and TAMix S. typhimurium strains (from Flückiger-Isler and Kamber, 2014).

Strain	Mutation	Type	Target	Cell wall[a]	Repair[b]	pKM101[c]
TA98	*hisD3052*	Frameshift	GC	*rfa*	*uvrB*	Yes
TAMix contains						
TA7001	*hisG1775*	b.p. subst.	T:A>C:G	*rfa*	*uvrB*	Yes
TA7002	*hisC9138*	b.p. subst.	T:A>A:T	*rfa*	*uvrB*	Yes
TA7003	*hisG9074*	b.p. subst.	T:A>G:C	*rfa*	*uvrB*	Yes
TA7004	*hisG9133*	b.p. subst.	C:G>T:A	*rfa*	*uvrB*	Yes
TA7005	*hisG9130*	b.p. subst.	C:G>A:T	*rfa*	*uvrB*	Yes
TA7006	*hisC9070*	b.p. subst.	C:G>G:C	*rfa*	*uvrB*	Yes

[a] (rfa): This mutation leads to a defective lipopolysaccharide (LPS) layer that coats the cell surface, making the bacteria more permeable to bulky chemicals
[b] (uvrB): The uvrB deletion mutation eliminates the accurate excision repair mechanism, thereby allowing more DNA lesions to be repaired by error-prone DNA repair mechanisms. The deletion through the biotin gene makes the bacteria biotin dependent
[c] (pKM101): This R factor plasmid enhances chemical and UV-induced mutagenesis via an error-prone recombinational DNA repair pathway. The plasmid also confers ampicillin resistance

Protocol

50. The TAMix strains and the microfluctuation test procedure that is used in the Ames II assay were developed by Gee et al. (1994). With the exception of the bacterial strains used,

the experimental protocol, detailed by Flückiger-Isler and Kamber (2014), is identical to the one described for Ames MPF™ (see 1.3).

Important differences with standard bacterial reverse gene mutation test

- Tester strains: Six *Salmonella* tester strains that are not included in the standard bacterial reverse gene mutation test (TA7001–TA7006);
- Total number of initial bacteria in contact with the test chemical: this number ranges between 1-2 x 10^7 in the Ames II test whereas 1 x 10^8 bacteria are used in the standard test. The Instructions for use of the Ames II assay recommend OD-based criteria for bacterial growth. Based on the OECD TG471 recommendations, many (but not all) laboratories confirm adequate bacterial growth by counting.
- Lower counts: spontaneous revertant well counts tend to be low (1 to 10 wells, depending on the strain). These numbers are not directly comparable to background colony counts in the standard bacterial reverse gene mutation test.
- Top concentration for soluble non-cytotoxic test chemicals: 1250 µg/well in the 24-well exposure plate instead of 5000 µg/plate.
- Number of replicates: the test chemical is usually evaluated in triplicate. A higher number of replicates may be used.
- Bacteria are exposed in liquid phase in a relatively small volume (250 µL), resulting in a relatively higher effective concentration of test chemical. This is similar to the preincubation version of the standard assay.
- Colorimetric readout: mutant frequencies are calculated based on the fraction of wells in which the pH of the medium changed as a result of the progeny bacteria growth.
- Toxicity of the test chemical is usually measured directly by checking an increase of the brilliance of the purple medium as compared to the solvent control due to cell lysis and absence of bacterial cells, or by inspection of the wells with a microscope. In the standard bacterial reverse gene mutation test, toxicity is inferred from reductions in revertant counts or qualitative observations of background lawns. Precise toxicity measurements can be useful for evaluation of concentration-response results that include a small increase in revertants, as well as a decline in revertants due to toxicity. Although not commonly done, toxicity can be precisely measured by post-exposure sampling and culturing of bacteria; either from a liquid suspension or agar plates. In some cases, these more precise measurements can help distinguish a mutagen from a non-mutagen. Methods for sampling bacteria from bacterial lawns on agar plates have been described in the literature; however, they are technically challenging (Prival, 2001). With respect to the liquid cultures used in fluctuation assays, similar measurements, although not reported in the literature, would likely be technically easier.
- The entire experiment takes place in a liquid phase eliminating the effects of the semi-solid agar. It is possible that these effects (e.g., alterations in migration of the test chemical or nutrients, bacterial growth or S9 metabolism) may be beneficial or adverse.
- Additional experiments can be conducted to determine true mutation frequencies, toxicity of the test chemical.

Perceived advantages

51. The Ames II procedure has several advantages over the standard bacterial reverse gene mutation test:

- Reduced amount of test chemical needed: the amount of test chemical used in the Ames II assay is one twentieth or less than what is needed for the standard bacterial reverse gene mutation test (e.g. total amount of test chemical needed for two strains with a top concentration of 5 mg/mL and half-log dilution scheme is 25.6 mg).
- Higher throughput: although when not automated, the Ames II is still a 'low throughput' assay, the throughput is faster compared to the standard bacterial reverse gene mutation test. Importantly, the assay can be partly or fully automated and processed by pipetting stations, thus requiring less hands-on time and resulting in a higher throughput.
- Read-out: the assay has an easy colorimetric readout in 48-well sections of a 384-well plate. Typically, triplicate plates with six sample dilutions, negative and positive controls are scored by eye in 5 min. The read-out is automatable for high throughput screening.
- An advantage of the fluctuation assay methodology is that unlike agar plate tests it allows determination of the true mutation frequency. This is because non-revertant bacteria cannot be quantitatively extracted from agar pates after exposure to the test chemical.
- Reduced reagent expenses: considerably less S9-mix and plastic ware (reduced disposal costs) is used.
- Less space needed: two 24-well exposure plates and 6 x 384-well plates contain all the replicates for 6 concentration levels and controls, with and without S9-mix, for a single tester strain. Consequently, the assay requires less incubator space and causes less plastic waste.
- No microcolonies are formed which can be wrongly interpreted as revertants.
- The test protocol of the Ames II assay is highly standardised and parameters optimised. The relatively homogeneous genetic background among the strains in the TAMix makes it easier to develop uniform criteria for bacterial growth and performance (e.g., optimised histidine concentration). The Ames II assay is available as a kit including all necessary ingredients ready-to-use (therefore, no media preparation, no autoclaving or sterility testing necessary no genotype analysis necessary).

Challenges

- Difficulty to compare the test concentrations with those used in the plate incorporation version of the standard bacterial reverse gene mutation test. Comparison with the pre-incubation assay is relatively easy (Flückiger-Isler and Kamber, 2012).
- The exposure concentrations, while higher than in the plate incorporation assay, are lower than in the preincubation assay which was developed because those high concentrations were thought to be needed for some types of test chemicals.

- Toxicity assessment by the decrease of positive wells relative to the solvent control and by an increase of the brilliance of the purple medium as compared to the solvent control (cell lysis, absence of bacterial cells).
- Interference of colored test chemicals with the colorimetric read-out is possible.
- High spontaneous revertant counts may result in lowered sensitivity due to the narrow dynamic range, resulting in an invalid result due to the top limit of wells considered per replicate (48).
- Use of strains with low spontaneous revertant colony counts.
  - ✓ The ability to detect a small increase in revertants may be reduced when there are no or too few revertants in the solvent control wells. Increasing the number of replicate wells for the solvent controls provides some compensation for this loss.
  - ✓ The six base-specific strains are combined to a single culture—the TAMix—and thus diluted 6-fold, further reducing sensitivity to each of the mutation types assessed by the assay.
  - ✓ Laboratories are unable to verify that each of the 6 strains is equally represented after overnight growth;
- Ames II does not detect frameshift mutations that specifically revert *S. typhimurium* TA1537 and thus has one strain specifically sensitive to frameshift mutations compared to two in the standard bacterial reverse gene mutation test;
- Commercial assay including proprietary elements (e.g., amount of histidine present in the media).
- Like the miniaturised agar plating assays, microfluctuation assays may not be suited for volatile test chemicals: the volatility of several compounds can cause an increase in the reversion frequencies of adjacent, untreated wells (Wilson & Cariello, 1997). However, this problem can be overcome by using adequate plate sealers.

Historical performance

52. According to Flückiger-Isler and Kamber (2014), there are several validation studies comparing the Ames II with the standard bacterial reverse gene mutation test using several strains. In all these studies, many chemical classes were evaluated.

- The first Ames II validation study with 25 test chemicals (i.e., 18 mutagens and 7 non-mutagens) was published in 1998 (Gee et al., 1998). In the Ames II test, strains used were 7001, 7002, 7003, 7004, 7005, and 7006 individually, the Mix (i.e., mixture of all these base-specific strains), TA1537 and TA98. In the NTP preincubation test used by Gee et al (1998) for comparison, strains actually used were *S. typhimurium* TA1535, TA1537, TA97 (but not always), TA98, TA100 and in very few times, TA 102/4. According to Gee et al., the overall concordance was 88% when the results of all of the strains TA700x, Mix (the 6 strains mixed), TA1537 and TA98 in the Ames II test

were compared to the NTP results for each of the 25 test chemicals[6] obtained from just TA100 and TA98 in the preincubation assay. However, when comparison is accurately analysed, the concordance is rather "qualitative", i.e., mutagen or not mutagen but not systematically qualitative (i.e., the same strain(s) and experimental conditions lead to the same result, e.g., mutagenic activity exclusively in strain S. typhimurium TA1537 and only with metabolic noted in both standard bacterial reverse gene mutation test and Ames II). See the recapitulative **Table 5** below, from Gee et al. (1998) publication. In any case, the concordances could be considered qualitative. Note that this agreement/disagreement of mutation spectra is often found.

- The authors concluded that the high concordance with the traditional bacterial reverse gene mutation test using *Salmonella* tester strains and the reproducibility among cultures and replicates demonstrate that the Ames II test procedure is an effective screen for identifying *Salmonella* mutagens.
- In an international Ames II round robin study (Flückiger-Isler et al, 2004), 19 coded test chemicals were selected for testing based on a published study with a large data set from the standard plate-incorporation bacterial reverse gene mutation test. The results of both assay systems were compared, and the inter-laboratory consistency of the Ames II test was assessed. Of the eight mutagens selected, six were correctly identified with the Ames II assay by all laboratories, one test chemical was judged positive by five of six investigators and one by four of six laboratories. All seven non-mutagenic samples were consistently negative in the Ames II assay. Of the four test chemicals that gave inconsistent results in the standard bacterial reverse gene mutation test, three were uniformly classified as either positive or negative in Ames II, whereas one test chemical gave equivocal results. Overall, the Ames II gave an 84 % concordance with published traditional results and an almost 90 % interlaboratory consistency.
- Another validation study was performed by Gervais et al. (2003). They tested 42 proprietary test chemicals and obtained a concordance of 83 %. The disagreement in the test results was obtained mostly with test chemicals that specifically revert *E. coli* or *S. typhimurium* TA1535.
- In a study released in 2009 (Kamber et al., 2009), 71 chemicals tested with the Ames II assay were compared with published data for the traditional bacterial reverse gene mutation test with *Salmonella* strains using the NTP database as the reference. There was 84% agreement between the two procedures in identifying mutagens and non-mutagens.
- Otherwise, the Ames II assay is routinely used for the screening of early drug candidates by several pharmaceutical companies or for the investigation of genotoxic impurities. Sanofi -Aventis reports a predictability of ~92 % when comparing the results of the Ames II with the full-scale standard bacterial reverse gene mutation test, and a throughput of 40 test chemicals weekly (Braun K, 2001). The Servier Group uses the Ames II test for early evaluation of in-house test chemicals (Lorge et al., 2007). All test

[6] This was calculated from a total of 22 chemicals (16 NTP mutagens and 6 NTP non-mutagens) + two mutagens in the NTP preincubation test not classified as mutagens, and one NTP non-mutagen (benzaldehyde) in the NTP procedure that was mutagenic in this study

chemicals that were positive in the Servier Ames II screening assay were confirmed positive in the standard bacterial reverse gene mutation test.

Table 5. Concordance analysis by Gee et al (1998) for the Ames II assay.

Chemical	±S9	AMAX									NTP preincubation test					
		7001	7002	7003	7004	7005	7006	Mix	1537	98	1535	1537	97	98	100	102/4
9 Aminoacridine HCl · H_2O	NA	−	−	−	−	−	−	−	+	−	−	+		−	−	
	S9										−	+		−	+w	
2-Amino-5-nitrophenol	NA	−	−	−	−	−	−	−	−	+	−	+		+	+w	
	S9										−	+		+	+w	
2-Amino-5-nitrophenol	NA	−	−	−	−	−	−	−	−	+	−	+		+	+w	
	S9										−	+		+	+w	
5-Azacytidine	NA	−	−	−	+	−	+	+	−	+	+w		−	−	?	+
	S9										+		−	−	?	+
Benzaldehyde	NA	−	−	−	−	−	−	−	−	−	−	−		−	−	
	S9	−	−	−	−	+	−	−	−	−	−	−		−	−	
Benzo[*a*]pyrene	NA	−	−	−	−	−	−	−	−	−	−	−		−	−	
	S9	+	+	−	+	+	+	+	+	+	?	+		+	+	
Benzyl chloride	NA	−	−	−	+	+	−	−	−	−	−	−	−	−	−	
	S9										−	−	−	−	+w	
Benzyl chloride	NA	−	−	−	+	+	−	+	−	−	−	−	−	−	−	
	S9	−	−	−	−	−	−	−	−	−	−	−	−	−	+w	
1-Chloro-2-propanol	NA	−	−	−	−	−	−	−	−	−	+	−	−	−	−?	
	S9	−	−	−	−	−	−	−	−	−	+	−	−	−	−?	
Coumarin	NA	−	−	−	−	−	−	−	−	−	−	−		−	−	
	S9	−	+	−	−	−	−	−	−	−	−	−		−	+	
Coumarin	NA	−	−	−	−	−	−	−	−	−	−	−		−	−	
	S9	−	+	−	−	−	−	−	−	−	−	−		−	+	
Crotonaldehyde	NA	−	+	−	+	+	+	+	−	+	−	−		−	+	
	S9										−	−		−	+	
Cumene hydroperoxide	NA	−	+	−	+	+	+	+	+	+	−	−		−	−	
	S9	+	−	−	+	+	+	+	+	−	−	−		+	+w	
Dicumyl peroxide	NA	−	−	−	−	−	−	−	−	−	−		−	−	−	
	S9	−	−	−	−	−	−	−	−	−	−		−	−	−	
Di(2-ethylhexyl)phthalate	NA	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
	S9	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
Dimethyl sulfoxide	NA	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
	S9	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
1,2-Epoxybutane	NA	−	−	−	+	−	−	−	−	−	+	−		−	+	
	S9	−	−	−	−	−	−	−	−	−	+	−		−	+	
Ethylenediamine	NA	−	−	−	+	+	+	+	−	−	+	−	−	−	+w	
	S9										+	−	−	−	+w	
Ethylenediamine	NA	−	−	−	+	+	−	+	−	−	+	−	−	−	+w	
	S9										+	−	−	−	+w	
8-Hydroxyquinoline	NA	−	−	−	−	−	−	−	−	−	−		−	−	−	
	S9	−	+	−	−	+	−	+	+	+	−		+	−	+	
Isobutyl nitrite	NA	−	−	−	−	−	−	−	−	−	?	−		−	+	
	S9	−	−	−	−	−	−	−	−	−	+w	−		−	+	
Isobutyl nitrite	NA	−	−	−	−	−	−	−	−	−	?	−		−	+	
	S9	−	−	−	−	−	−	−	−	−	+w	−		−	+	
Nitrofurantoin	NA	−	+	−	+	+	+	+	−	+	−	−		+	+	
	S9										−	?		+	+	
4,4′-Oxydianiline	NA	−	−	−	−	+	−	+	−	−	−		−	−	+w	
	S9										−	+w	+	+	+	
Phenol	NA	−	−	−	−	−	−	−	−	−	−	−		−	−	
	S9	−	−	−	−	−	−	−	−	−	−	−		−	−	

NA: not tested with S9; -: negative; +: positive; +w, weakly positive; ?: equivocal.
Note: Five chemicals, i.e. 2-amino-5-nitrophenol, benzyl chloride, coumarin, ethylenediamine, and isobutyl nitrite, were tested in duplicate to assess the reproducibility of the Ames II assay.

1.5.2. Microfluctuation tests using bioluminescent tester strains

53. The bioluminescent Ames test in liquid format has been described by Côté et al. (1995). In the assay, genetically modified *Salmonella* TA98 bacteria expressing lux (*luxA* and *luxB)* genes from *Vibrio harveyi* are exposed to test chemical for 48 hours in 48-well plates containing an appropriate liquid medium. Afterwards, cells are centrifuged and resuspended in buffer. Changes in revertant biomass are measured as an increase in bioluminescence.

54. However, although Côté et al. showed the potential of using bioluminescent *Salmonella* strains to screen compounds for their mutagenic potential, they only performed experiments with the genetically modified TA98 strain in the absence of metabolic activation. Further validation of the assay was thus required, but so far, no well-validated versions of this assay have been described. As also no data were received, this miniaturised bacterial reverse gene mutation test was not further considered in this DRP.

2 Applications of the miniaturised bacterial reverse gene mutation tests

55. This chapter describes the applications of the different miniaturised bacterial reverse gene mutation tests. First, the current use of the assays is discussed which is mainly limited to screening of test chemicals and testing of drug impurities. Afterwards, several scenarios are presented on how the miniaturised versions can be applied in the future. The selection of the most appropriate scenario will be driven by the outcome of the retrospective validation study and may vary for the different versions of the bacterial reverse gene mutation test.

2.1. Current applications of the miniaturised bacterial reverse gene mutation tests

56. In early stages of product research and development in industry, and in specific regulatory domains where only small amounts can be tested, a full standard bacterial reverse gene mutation test is often not necessary or even not feasible due to the limited availability of test chemical and resources (Hamel et al., 2016). Under these conditions, a miniaturised bacterial reverse gene mutation test is considered an acceptable alternative (**Table 6**). Note that other alternatives exist (e.g., a standard protocol with fewer than 5 strains, Williams et al. 2019). It is beyond the scope of this DRP to compare the miniaturised assays to any alternative approaches or to evaluate concordance of any of the miniaturised assay with the standard assay within any specific regulatory context.

2.1.1. Testing of drug impurities

57. For drugs, the ICH Guideline M7 (ICH, 2017) suggests a miniaturised assay format as one of two alternatives to evaluate an impurity when use of a computational toxicology method identifies a structural alert for mutagenicity in situations where it is impossible to achieve the highest test concentrations in the standard assay e.g. because it is not feasible to isolate or synthesize the impurity. The guideline does not identify a miniaturised format, highlights the need for a proven high concordance between the standard assay and the miniaturised format and also notes that justification for use of the miniaturised format would be needed.

2.1.2. Screening of chemicals

58. Some of the miniaturised bacterial reverse gene mutation tests are already used for screening by (large) industries to predict the outcome of a putative regulatory GLP-compliant study before a report is submitted to the appropriate authorities as part of a regulatory package or dossier (Escobar et al., 2013a in connection with its *corrigendum* Escobar et al., 2013b; Proudlock and Evans, 2016). Within a screening context, advantages of the miniaturised bacterial reverse gene mutation tests are the reduction in the costs, the waste (including biologically and chemically contaminated waste), and the cost of waste disposal (Burke et al., 1996; Kamber et al., 2009; Pant et al., 2016; Proudlock and Evans, 2016) and particularly, the

smaller amount of test chemical needed. The miniaturised bacterial reverse gene mutation tests are mid-throughput tests compared to a slow throughput standard bacterial reverse gene mutation test. The industrial screening is often performed in a very early stage of development, where for a specific aim many new substances are under development and where only a low amount of test chemical is available. In a later stage, when fewer candidate substances remain and more test chemical is available for testing, industries may prefer the standard bacterial reverse gene mutation test.

59. To date, the miniaturised bacterial reverse gene mutation tests are routinely used at several pharmaceutical laboratories to assess mutagenic potential of **drug candidates** prior to running the standard bacterial reverse gene mutation test under GLP conditions (Escobar et al., 2013a,b). Although many (industrial) laboratories throughout the world, including most large contract research laboratories, use these miniaturised bacterial reverse gene mutation tests, very little has been published on their performance or utility with a few notable exceptions (Flückiger-Isler et al., 2004; Kamber et al., 2009; Flückiger-Isler and Kamber, 2012; Escobar et al., 2013; Pant et al., 2016; Spiliotopoulos and Koelbert, 2020). At most industries, a positive result in a screening test leads to the termination of the candidate substance. The miniaturised bacterial reverse gene mutation tests are also used for the evaluation of drug impurities (see 2.1.1) as well as intermediates to ensure worker safety during larger scale production (see 2.1.3). Thus, a highly reliable screening method is crucial. Escobar et al. (2013a,b) reported that the miniaturised bacterial reverse gene mutation tests, used in a tiered approach with *in silico* analysis, are powerful tools for the prediction of mutagenicity in the early stage of development and most helpful to decide which further tests should be selected for impurities and intermediates with the advantage of higher throughput and lower consumption of test chemical.

60. In addition to drugs (including veterinary drugs), the miniaturised bacterial reverse gene mutation tests could also be applied for screening in other chemical domains. For chemicals, biocides, food (additives), pesticides, and cosmetics (ingredients) there is generally enough test chemical available for a standard bacterial reverse gene mutation test. However, as for drugs, the miniaturised tests could be useful as a first step to assess compounds of interest or for testing of impurities or metabolites. For example, the use of the 24-well plate assay was mentioned as a routine pre-screen for novel agrochemicals (Brooks, 1995). For food and food ingredients, the miniaturised bacterial reverse gene mutation tests may have a role when contaminants or impurities need to be tested.

2.1.3. Other applications

61. There are other domains of interest for applying the miniaturised bacterial reverse gene mutation tests. One example includes occupational safety assessment as evaluation of chemicals handled by workers (e.g., intermediates) is needed to ensure that adequate precaution handling will be implemented to guarantee their safety during production. This has to be done long before large scale production. Other domains where only a limited amount of testing material is available include medical devices, aerosols from e-cigarettes and non-intentionally added substances (NIAS) migrating from food contact materials.

62. There may also be applications in environmental monitoring in which the assays are useful. For example, mutagenic potency can be a relevant endpoint when screening soils and

other complex environmental samples. The Ames II test was applied in the drinking water industry where an increased use of surface waters was observed, which carry micropollutants like pesticides, pharmaceuticals and organic solvents, in the preparation of drinking water. The Ames II test was used to determine whether in the cleaning process with UV/H_2O_2 oxidation and granular activated carbon (GAC) filtration could lead to formation of genotoxic compounds (Heringa et al., 2011).

63. Finally, the assays can also be used to identify the fraction containing the bioactivity when a complex mixture is fractionated.

Table 6. Overview of the application of miniaturised bacterial reverse gene mutation test in different regulatory domains.

Domain	Application of miniaturised bacterial reverse gene mutation test
Drugs (including animal drugs)	Testing of impurities Screening of candidate drugs Testing of substances available in low amounts (e.g., intermediates)
New and existing chemicals	Very early screening
Food, food contact materials, food additives Agrochemicals, pesticides and biocides	Very early screening Testing of impurities/metabolites Testing of substances available in low amounts
Consumer products including cosmetics and cosmetic ingredients	Very early screening
Medical devices	Very early screening Testing of impurities Testing of substances available in low amounts
Environmental monitoring	Screening of soils and other complex environmental samples like water
Complex mixtures	Identifying genotoxic fraction

2.2. Possible scenarios for future use of the miniaturised bacterial reverse gene mutation tests

64. For the future use of the miniaturised bacterial reverse gene mutation tests, there are (at least) 3 possible scenarios:

i) as 100% accepted alternative for the standard bacterial reverse gene mutation test;
ii) as alternative test when the standard bacterial reverse gene mutation test is not an option, or
iii) exclusively for screening.

65. The main question for application scenario (i) is if the miniaturised bacterial reverse gene mutation tests are acceptable as standalone tests. One of the first things that should be considered is in how far the miniaturised bacterial reverse gene mutation tests comply with and what distinguishes them from the standard assay. These specific characteristics may strongly determine the final use of the miniaturised bacterial reverse gene mutation tests, its place in notifications and strategies as compared to the standard bacterial reverse gene mutation test, and a putative prospective validation.

66. When using these miniaturised bacterial reverse gene mutation tests, it is important to ensure that the sensitivity for the detection of mutagens is not compromised. There have been some indications and/or concerns that the sensitivity of the miniaturised assays may be lower than for the standard bacterial reverse gene mutation test (Escobar et al., 2013a,b) leading to false negative results. In the standard test, mutation is a quite rare event; depending on the bacterial strain, the background revertant frequency may vary from 5-500 per 10^8 (Levy et al., 2019). Since in the miniaturised bacterial reverse gene mutation tests substantially fewer than 10^8 bacteria are exposed to the test chemical, the probability for a revertant colony to be detected may be lower. In order to ensure the detection of mutagenic test chemicals, in particular the weak responders, it is therefore important to assess whether the format of the miniaturised tests and the lower number of bacteria used allow the adequate detection of rare reversion events. The impact of the other methodological modifications, such as the use of bacterial strains other than those used in the standard bacterial reverse gene mutation test, also needs to be considered. Next to this concern of a potential reduction in test sensitivity, the specificity must not be impacted either.

67. Even if sensitivity and specificity of the miniaturised bacterial reverse gene mutation tests are good, the experience of regulators with the standard bacterial reverse gene mutation test, the place it has demanded in the present strategies for genotoxicity testing and the enormous database may make regulators continue to prefer the standard test. It may be different, however, under special conditions (application scenario (ii)), for instance, if only a small amount of test chemical is available due to difficulties with the isolation or synthesis, such as for (pharmaceutical or agrochemical) impurities, metabolites, environmental substances or extracts derived from medical devices. A role for the miniaturised bacterial reverse gene mutation tests is then certainly acceptable and should be permitted.

68. In application scenario (iii), validation must provide some metrics on sensitivity and specificity to allow a science-based choice on how to use the results and on whether they can be considered final or whether further testing is mandatory. However, screening is not regulated with OECD guidelines and thus out of the scope of this DRP; every industry can decide how and when miniaturised bacterial reverse mutation tests are used.

69. The possible role for the miniaturised bacterial tests may have an effect on a putative prospective validation study. For use as alternative for the standard bacterial reverse gene mutation test, the miniaturised bacterial tests may have to be fully validated. If miniaturised bacterial reverse gene mutation tests are exclusively used for screening or only to generate data that will not ultimately be used in a human safety context, neither GLP nor a validated TG are needed since the assay would be repeated in the full format for later submission.

3 Retrospective data analysis

70. Following the establishment of an EG on the Miniaturised bacterial reverse gene mutation test at the end of 2016, a kick-off meeting was organised on the 28th of February and the 1st of March 2017 to launch the project. The main objectives of the meeting were to identify and characterize the various miniaturised versions of the bacterial reverse gene mutation test, identify data needs and develop plans for the next steps of the project including the collection of data. The different steps of the project are illustrated in **Figure 3**.

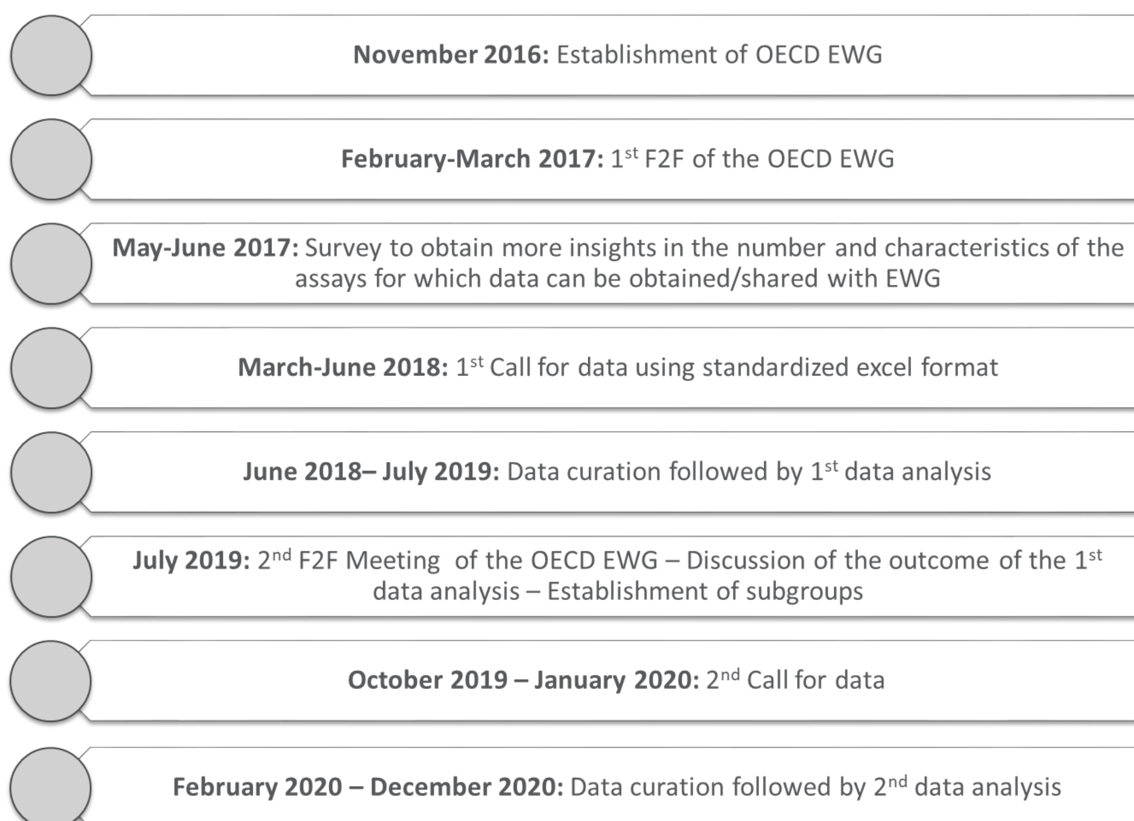

Figure 3. Timeline summarising the different steps of the data collection and data analysis for the retrospective validation study.

3.1. Data collection

3.1.1. Exploratory survey

71. One of the follow-up steps identified during the kick off meeting was the development and circulation of a survey in order to obtain better insights in (i) the amount of data available for the different miniaturised bacterial gene mutation tests and (ii) the extent to which these data could be shared. The information collected with this initial survey was intended to support the decisions of the EG on the further actions required for the assessment of the miniaturised tests including the need for more validation/qualification studies. The complete survey is provided as Annex I of the DRP.

72. The survey was sent to the OECD Working group of the National Coordinators of the Test Guidelines Programme (WNT) with a request to contact relevant experts in their networks, to the EG and to another OECD expert group on genotoxicity. It was directed to experts likely to provide data, i.e., mainly people in laboratories who conduct one or more versions of the miniaturised bacterial reverse gene mutation test. In addition, the survey was also circulated to other networks including, amongst others, the HESI Genetic Toxicology Technical Committee (GTTC) and the European Union Network of Laboratories for the Validation of Alternative Methods (EU-NETVAL). Furthermore, all addressed experts were invited to forward the survey to their contacts, targeting people they knew who might be able to contribute, and asking them to respond to the survey.

73. Overall, the feedback received in the survey was very positive. Sixteen respondents indicated that they have experience with one or more miniaturised versions of the bacterial reverse gene mutation test, among which the majority agreed to share their data. Furthermore, the European Chemicals Agency (ECHA) completed a simple search of the data on Miniaturised bacterial reverse gene mutation tests in their IUCLID database (July 2016). When using the search terms '*miniaturised Ames', or 'miniaturised Ames', or 'bioluminescent Ames', or 'Ames II', or 'Ames MPF'*, anywhere in section 7.6.1. of IUCLID (i.e. Genetic Toxicity in vitro), they found 267 references in 64 dossiers, corresponding to 57 unique substances. It should be noted that the search was not limited to the disseminated information but to all information as submitted in the registration dossiers. However, ECHA was not in a position to fully verify the correctness of all results, nor to confirm whether the assay was conducted on the registered substance or an analogue. Another 7 respondents of the survey indicated that they do not perform any miniaturised versions of the bacterial reverse gene mutation test in their lab.

74. In general, responses to the survey indicated that a substantial number of laboratories were ready to share data that could be used in a retrospective validation study for the various assays, and that a call for data would be likely to generate sufficient data for the analysis of several miniaturised bacterial reverse gene mutation tests.

3.1.2. First Call for data

75. In March 2018, a first call for data was made publicly available on the OECD public website. This was accompanied by letters directed to the WNT (with a request to contact relevant experts in their networks), to relevant OECD expert groups, and to a large number of laboratories, including the respondents to the survey (see above). In case of confidential data, experts were requested to mention this specifically, or to directly code the test chemicals in the submission form.

76. In order to ensure uniformity of submissions across the laboratories, a standardised Excel spreadsheet was created for the data collection. The design (1) enabled inclusion of all variables deemed to be useful for comprehensive data analyses, and (2) confined user data entries via creation of a restricted series of selection options. Alongside the test chemical assessment results, data for all solvent controls were also requested. However, concurrent positive control data did not have to be provided, unless this positive control was being used as the test chemical. Nevertheless, positive control data were submitted by most of the laboratories.

Table 7. Overview of the data fields included in the standardised Excel data collection spreadsheet.

Column title	Data entry limitations
Data Source	None (free text field)
Compound Name	None (free text field)
Compound abbreviation	None (free text field)
CAS Number	None (free text field)
Assay Name	Data entry restricted to the cell drop down menu (6-well plate, 24-well plate, Ames II, Ames MPF, Kado/Microsuspension, Fluctuation, BioLum, Standard (plate incorporation), Standard (preincubation), Other).
Treatment Type	Data entry restricted to the cell drop down menu (Plate incorporation, Preincubation, Fluctuation (liquid) or Other).
Bacterial Exposure vessel	Data entry restricted to the cell drop down menu (6-well plate, 24-well plate, 96-well plate, 384-well plate, test tubes, ~90 mm agar plate or other).
Mutation Scoring Plates	Data entry restricted to the cell drop down menu (6-well plate, 24-well plate, 96-well plate, 384-well plate, Petri-dish (full sized plate) or other).
Date of Assay	Date entry in the format DD MMM YY
Bacterial Strain	List of the most commonly used strains or strain mixes provided in the cell drop down list. If the actual strain used is not on this list, free text entry of the strain name is permissible to enter the specific strain used.
S-9	Data entry limited to the cell drop down options of + S9-mix (in the presence of S-9) or -S9-mix (in the absence of S-9).
S-9 Source	Data entry prompt for the most commonly used S9 sources presented in the cell drop down list (Rat liver, Hamster liver, Human liver or other).
S-9 Inducing Agent	Data entry prompt for the most commonly used chemical S9-mix inducing agents presented in the cell drop down list (Aroclor 1254, Phenobarbital/β-naphthoflavone, uninduced, Other).

S-9 concentration (%) in the S-9 mix	Data entry restricted to integers up to 100.
Volume of S-9 mix or buffer solution added (µL per well/plate)	Data entry restricted to numerical values up to 2 decimal places. Either enter data in this column, or in next column (final S9-mix % in well/plate), whichever is preferred. There is NO requirement to enter data into both these columns.
Final S-9 % in well/plate	Data entry restricted to numerical values up to 100 and to 2 decimal places. Either enter data in this column, or in previous column (volume of S9-mix or buffer solution added (µl per well/plate), whichever is preferred. There is NO requirement to enter data into both these columns.
Total Assay Volume (µL per well or plate)	Data entry restricted to numerical values up to 2 decimal places.
Vehicle/solvent	Water or dimethyl sulphoxide (DMSO) can be selected from the drop-down list, or any other vehicle/solvent can be entered as free text
Test Compound Treatment Concentration	Numerical values only
Units for compound concentration (preferably µg per mL, per well or per plate)	Data entry prompt is for the preferred units to be used for the data as provided in the cell drop down list (µg/mL, µg/plate, µg/well). Free text entry is permissible to enter the appropriate units for the compound concentration in the assay.
Toxicity observed?	Data entry limited to the cell drop down option of Yes (evidence of toxicity observed), No (no evidence of toxicity observed) or N/D (toxicity not determined or data not recorded)
Precipitation observed?	Data entry limited to the cell drop down option of Yes (precipitation observed), No (no precipitation observed) or N/D (test article precipitation not determined or data not recorded)
Number of replicates	Data entry limited to integer values greater than zero
Mean_mutation count or value	Data entry limited to numerical values and to 2 decimal places
Units for Mean mutation count or values	Data entry prompt is for the commonly used units for the data as provided in the cell drop down list (revertants/well, revertants/plate, positive wells/48, positive wells/96, mutation frequency, other). Free text entry is permissible.
(For fluctuation assays only) Number of scoreable wells if other than 48 or 96	Numerical values only
Calculated variability of mutation value	Enter only numerical values to a maximum of 1 decimal place
Measure used for mutation variability	Data entry prompt in the drop-down list is for the more commonly used measures of variability used (Standard Deviation (SD), Coefficient of Variance (%CV), Relative Standard Deviation (RSD), Standard Error of the Mean (SEM), Value at Risk (VAR)). Free text entry is permissible.
Positive, Negative or Equivocal call	Data entry is restricted to the list in the cell drop down menu (Positive, Negative, Equivocal)
Comments	Free text

77. Data was required to be entered in a separate row for each treatment concentration, but not for each replicate (i.e., only mean values for each concentration were entered). In total, there were 29 columns for data entry, plus a further column for optional comments (**Table 7**). For 15 of the 29 columns, drop down lists were used for data entry, which confined user entries to a restricted series of selection options. This simplified and streamlined data compilation, curation and analysis. Respondents were encouraged to use drop down options as far as possible, and to avoid using the 'other' or free text options unless none of the drop-down options were considered appropriate. For the remaining 14 columns, data entries were either only limited by format (e.g., dates in the format DD MM YY, numbers to a set number of

decimal places, etc.), or were free text. The data requested from each submitter were intended to provide sufficient information on the assay design and results considered to be required for the data analyses. Importantly, the assessment call (i.e., positive, negative or equivocal) had to be entered at the appropriate location. Positive calls were entered adjacent to the lowest concentration that elicited a positive response. Negative calls were entered adjacent to the highest tested concentration. This assessment call was made by the submitting laboratory itself based on their own standardised assessment criteria.

78. Data submission guidance was provided in the first tab (i.e., worksheet) of the data collection workbook. Furthermore, a tab with a description of the various assay methodologies identified was provided, as well as one containing a data entry example. Respondents were recommended to provide, wherever possible, corresponding standard bacterial reverse gene mutation test results (i.e., standard plate incorporation and/or preincubation assay) data for test chemicals for which data from a miniaturised version were submitted. This permitted, to the extent possible, direct comparisons of results in the standard version of the assay with the miniaturised version of the assay.

79. The first call for data was closed in June 2018. Data were received from more than 20 laboratories (21 laboratories including one joint submission), covering 5 versions of the miniaturised bacterial reverse gene mutation tests. The data collected served as a basis for an initial retrospective analysis of the various miniaturised versions of the bacterial reverse gene mutation test. However, the initial data analysis showed clear differences in the amount of data submitted for these 5 assays. Most data were received for the 6-well plate assay and the Ames MPF™ test. Although some datasets were submitted for the 24-well plate assay and the Ames II test, these data were considered insufficient to allow a retrospective data analysis. For the miniaturised version of the microsuspension method of the standard assay, data were only received from one lab. Furthermore, literature review showed that no other laboratories have performed this assay. For this reason, the assay was not further considered for the retrospective data analysis.

3.1.3. Second Call for data

80. In order to improve the coverage of the dataset, and to place the 4 miniaturised assays on a more equal footing, the EG, which met in Paris in July 2019 for a second face-to-face meeting, agreed to launch a complementary call for miniaturised bacterial reverse gene mutation test data. This second call was open to any miniaturised bacterial reverse gene mutation test, but with a specific focus on assays for which gaps had been identified, i.e., the 24-well plate and Ames II tests. Seven laboratories submitted data in response to the second call (four of them had already responded to the first call and submitted additional data).

81. The data collection spreadsheet that served for the first call was slightly updated, thereby taking into account small issues encountered during the first data analysis. Furthermore, the following additional recommendations were made for the data submission:

- Whenever possible, paired data had to be provided, i.e., for each test chemical submitted, data were requested both for a miniaturised bacterial reverse gene mutation test and a corresponding standard plate incorporation or preincubation test;

- In case data on a miniaturised bacterial reverse gene mutation test were available without corresponding standard bacterial reverse gene mutation test data, data could still be submitted if the test chemicals belonged to the list of compounds for which standard data had already been collected in the first call. This 'standard Ames compound list' was provided to the submitters.
- If possible, data submitted had to be balanced in terms of positive and negative test chemicals;
- A Tab called 'list of test chemicals and abbreviations' was also included in the Workbook. To facilitate the management of the data, data submitters who submitted data for test chemicals already in this list had to enter test chemical names and abbreviation in the data collection spreadsheet exactly as indicated in the list.

3.2. Comparison of protocol variations among the laboratories

82. One requirement for generating comparable data among laboratories is that the labs use comparable laboratory protocols. Some basic information about laboratory procedures were tabulated in the data collection tables. Out of the 23 laboratories submitting data, 18 responded to a follow-up questionnaire asking for further details, including a description of their criteria for identifying a positive result. About half the laboratories responded to requests for historical control data. Some historical control datasets were based on very few experiments, and laboratories noted that they were still refining criteria for assay acceptance and/or interpretation. Data submitted for the retrospective validation study were used to determine the relative sizes of some parameters of the miniaturised bacterial reverse gene mutation tests and the solvent control counts (**Table 8**). Strains used and strain-specific outcome criteria provided by the laboratories are summarised in **Table 9.**

Table 8. Overview of relative sizes of some parameters and solvent control counts as reported by laboratories participating in the retrospective validation.

Relative sizes of some parameters as reported by laboratories participating in the retrospective validation.										
	Standard			6-Well		24-well		Ames MPF™		Ames II
Relative # viable bacteria plated	100% (10^8 CFU)			20%		5%		10-20%		10-20%
Relative plate area	100% (75 cm^2)			22%		2.5%		Not applicable		Not applicable
Relative volume (Preincubation)	100% (0.65[a] ml)			NA		27[b] or 9%		38[b]%		
Relative volume (long incubation)	100% (2.65[a] ml)			20-25%		5-6.5%		2%		
Solvent controls: Highest and lowest laboratory averages for each strain.										
	Range[c]	Range[d]		Range		Range		Range		Range

E. coli uvrA	21-62	17-82		4.6-17		1.4-2.8		2.1-4.8		
E. coli uvrA pKM101	96-207	28		12-54		6.0-8.0		3.8- 4.8		
S. typhimurium TA100	59-177	78-128		13-37		4.9-7.2		4.5-5.2		
S. typhimurium TA102	247-370	241-360		71				7.1		
S. typhimurium TA1535	15-39	15-58		3.3-12		1.2-3.2		1.2-2.2		
S. typhimurium TA1537	6.7-71	4.3-14		2.0-10.4		4.8		1.2-2.4		
S. typhimurium TA97	96-191	30-162		17-44		6.3-7.9		2.9, 5.8		
S. typhimurium TA98	14-44	8.3-94		3.2-14.		1.7-2.7		1.1-4.5		1.7-2.4
S. typhimurium TAMix										0.9-2.0

Top: Percentage of standard assay amounts based on descriptions provided by laboratories participating in the retrospective validation. [a]Standard assay preincubation mixtures of 0.6 or 0.7 mL, and final volumes of 2.6 or 2.7 mL were averaged for this table. [b]A single "pre-mix" for each concentration is used. Either immediately, or following preincubation, the pre-mix is diluted with soft agar or medium and then aliquoted into replicate wells.

Bottom: Each number in this table reports the average solvent control count per well or plate for all experiments with that strain described by a single laboratory. Data from all solvents and both with and without S9 were consolidated into a single average value. The smaller value represents the laboratory with the smallest overall mean. The larger value represents laboratory with the largest overall mean. A single value indicates only one laboratory reported data for that strain in the respective assay. Values separated by commas are individual results for the only two laboratories submitting data. Means for [c]preincubation and [d]plate incorporation experiments were tabulated separately only for the standard bacterial reverse gene mutation test.

Table 9. Strains used and strain-specific outcome criteria provided by the laboratories.

Assay	*Salmonella strains*						*E coli* WP2	
	TA100	TA98	TA97(a)	TA1537	TA1535	TA102	*uvrA*	*uvrA* pKM101
24-well plate	2 fold	3 fold	2 fold		3 fold	1.5 fold		2 fold
	2 fold & >6	3 fold & >6	2 fold & >6		3 fold & >6			2 fold & >6
	2 fold & >9	3 fold & >6						
6-well plate	2 fold	2 fold		3 fold		2 fold		
	2 fold	2 fold		3 fold	3 fold			2 fold
	2 fold	2 fold	2 fold		3 fold	1.5 fold		
	2 fold	2 fold	2 fold		3 fold			2 fold
	2 fold	2 fold	2 fold	3 fold	3 fold			2 fold
	2 fold	2 fold		3 fold	3 fold			2 fold
	2 fold	2 fold						
	2 fold	2 fold		3 fold	3 fold			

Ames MPF™ Ames II	2-fold over "baseline" (the mean solvent control + 1 standard deviation or 1, whichever is larger)+.

Each row includes the criteria reported by a single lab. Many but not all laboratories provided information about the criteria used to identify a positive result for each strain. Those that did not provide information are not included in the table. Note that for the 24-well plate assay laboratories combined two criteria: a fold increase and a minimum number of revertant colonies.
+Statistical methods are also presented as alternatives in the protocol description submitted by one laboratory.

83. The solvent control counts reported in **Table 8** for the laboratory with the highest average reported were generally 3 to 4 fold higher than the average counts for the laboratory reporting the lowest average for each strain used in the standard assay. A similar level of variation was seen in solvent control count data for the individual experiments used to calculated the averages shown in table 8 (data not shown). This degree of inter- and intra-laboratory variation is consistent with literature reports (Kim and Margolin, 1999; Kato et al., 2018; Levy et al., 2019). Variability in the solvent control counts among laboratories reporting data for miniaturised assays (table 8) was not formally analysed but appears to be similar to the variability seen in standard assay data.

3.2.1. 6-well plate assay

84. A summary of the protocol variations for the 6-well plate assay is provided in **Table 10.** All the submitted data used the plate incorporation version of the assay. Twelve laboratories submitted data, and 8 provided historical control data. The latter has yet to be compared to the concurrent solvent control information in the submitted data. One laboratory reported having a protocol for using the preincubation version of the assay, but no details were provided. Notably, one laboratory reported using 5% S9-mix, which is lower than the typical 10%, but within the range suggested in OECD TG471 (OECD, 1997). Most laboratories scaled all reagents down to 20% of the volumes used in the standard bacterial reverse gene mutation test; three of them used slightly more reagents, with the final volume corresponding to 25% of the amounts used in the standard assay. Note that although this is slightly larger volume variation than in the standard assay (i.e., usually 2.7 or 2.6 mL), one laboratory reported using 35% more volume (i.e., 3.7 mL).

85. Strains used: All but 3 laboratories used at least 5 strains for at least some of the submitted data. The remainder all used *S. typhimurium* TA98 and TA100 alone, or those strains combined with two other strains, i.e., *S. typhimurium* TA1535/1537 or TA1535/TA102, respectively. Among laboratories using 5 strains, all used *S. typhimurium* TA98, TA100 and TA 1535. The laboratories were roughly evenly split between those using *S. typhimurium* TA1537 and those using *S. typhimurium* TA97, and similarly between *S. typhimurium* TA102 and *E. coli* WP2 *uvrA*, the latter with or without pKM101.

86. About two thirds of the laboratories responded to a survey asking about assay interpretation criteria. The respondents all used the same fold rules usually associated with the standard assay, i.e., 3-fold for *S. typhimurium* TA1535 and TA 1537, and 2-fold for the other strains.

87. Number of replicates: The laboratories used a variety of choices for numbers of replicate plates. Replicate plates used for most experiments among the 8 laboratories submitting 6-well plate data:

Solvent	6	6	6	6	6	3	2	2
Test Chemical	6	3	3	3	3	3	2	2

88. The background for each strain was remarkably similar across laboratories, and thus does not explain decisions to use different numbers of replicate plates (**Table 8**). Even though backgrounds were roughly 5% of those for the corresponding strain in the standard bacterial reverse gene mutation test, none of the submitted experiments reported a mean of 0 (i.e., there were no experiments in which there were no colonies on any of the 2, 3 or 6 solvent control plates). The smallest laboratory mean was 1.2 colonies per plate in *S. typhimurium* TA1535 (**Table 8**), and approximately 8% of the individual experiments from all laboratories averaged ≤2 colonies per plate. The mean, standard deviation, maximum, minimum and number of experiments for each laboratory has been tabulated (not shown). Larger historical control datasets provided by the laboratories look similar to the submitted data for the solvent controls. For example, many reported means between 2 and 5 colonies per plate for *S. typhimurium* TA1535 and 1537. Quantitative assay sensitivity, the ability of the assay to detect a small increase above background, declines rapidly as the number of plates with 0 colonies increases. This effect is described in detail for other genetic toxicology assays (OECD 2014).

Table 10. Summary of the protocol variations for the 6-well plate assay.

Consistent among the laboratories		Further investigation may be needed
Information from many laboratories facilitates comparisons		Fold rule criteria applied based on experience with standard assay, not background counts
Protocols are similar among laboratories, facilitating development of a protocol if needed for prospective validation		Plates with 0 colonies after exposure to test chemical might reduce sensitivity to detect weak responses
Most laboratories reported 5 strain experiments.		No experience with preincubation protocols
All strains described in TG471 can be used		
Mean solvent controls were all >0 in all submitted data		
Test acceptance and outcome criteria similar among all laboratories		

3.2.2. 24-well plate assay

89. A summary of the protocol variations for the 24-well plate assay is provided in Table 11. Most of the submitted data came from 3 laboratories using plate incorporation protocols, two others reported exclusively using preincubation protocols. Most laboratories made a “premix” of bacteria and S9-mix (or buffer). One laboratory used the entire premix for preincubation before aliquoting it into the replicate wells, the other preincubated each replicate separately in a smaller volume. Both preincubation protocols resulted in a preincubation mixture with 14% v/v solvent concentration. That is considered acceptable for the standard bacterial reverse gene mutation test, but for common organic solvents, interference with S9-mix metabolism of some test chemicals has been reported under these conditions (Hakura et al., 2010; Mori et al.,1985; Chauret et al., 1998). For the plate incorporation version, both laboratories reported adding the test chemical to the well on top of the bottom agar, and then adding the premix.

90. Most laboratories used *S. typhimurium* TA97a rather than *S. typhimurium* TA1537. One laboratory reported being unable to grow *S. typhimurium* TA97a without an 8-fold reduction in the amount of glucose in the bottom agar. Most laboratories applied the standard fold rules to determine the test outcome, but increased the criterion for TA98 to 3-fold. In addition, some laboratories introduced, as an additional requirement, a minimum increase in revertants over the concurrent solvent control to a minimum value (see **Table 9**). Another laboratory evaluated the test outcome based on a 99% Poisson control interval using the concurrent solvent control. All laboratories used 12 replicate solvent control plates for most experiments. In case experiments were performed with and without metabolic activation, 12 replicates with S9-mix and 12 replicates without S9-mix were included. The number of replicates used for the test chemical was quite variable, with laboratories variously reporting using 2, 3 or 4 per sample under a specific metabolic condition (i.e. either with or without S9-mix). The mean solvent control values calculated from the reported data were considerably higher than might be predicted from starting with only 5% of the bacteria used in the standard assay (**Table 8**). The laboratory mean solvent controls were all >1 colony per well, but means for individual experiments were <1 in about 9% of the reported experiments (not shown). Historical control data were provided by 2 laboratories, plus a third that did not submit experimental data for this assay. Some laboratories compiled the data per plates, others based per experiment (i.e., mean of 12 replicates). Two laboratories compiling data per experiment reported minimum values of 0 for multiple strains, i.e., none of the 12 replicate solvent control plates had any colonies.

Table 11. Summary of the protocol variations for the 24-well plate assay.

Consistent among the laboratories		Further investigation may be needed
Laboratories increased the number of solvent control replicates and all used the 12 replicates		Sensitivity may be limited by the frequency of plates without revertant colonies, both in solvent controls and with test chemical
Protocols are available for both plate incorporation and preincubation		Experimental work may be needed to develop a protocol for a prospective validation study
Most laboratories reported 5-strain experiments		High solvent concentrations in the preincubation assay may compromise metabolism of some test chemicals
All strains described in TG471 (OECD, 1997) can be used		Possible problem with use of standard bottom agar with *S. typhimurium* TA97
		With data from only two companies for each protocol, inter-laboratory reproducibility may be harder to evaluate

3.2.3. Ames II and Ames MPF^{TM}

91. A summary of the protocol variations for the Ames II and Ames MPF^{TM} assays is provided in **Table 12**. Methods and criteria appeared to be uniform across laboratories, consistent with use of a standardised protocol, plus reagents and manuals provided by the vendor. However, one laboratory expressed concerns about lack of information about the contents of some reagents, such as the actual level of histidine, which must be present in limiting amounts. Laboratory solvent control means were all >1 for all bacterial strains (**Table 8**). For the individual Ames MPF^{TM} experiments, 3.5% of the mean solvent control counts were reported to be 0, and <1 for an additional 22% of experiments (not shown). Similarly, there were Ames II experiments without growth in any of the replicate solvent control wells. The ranges of laboratory mean values in the Ames MPF^{TM} assay were quite similar for *S. typhimurium* TA98, TA1535 and 1537, suggesting they are equally suitable for use with the assay.

Table 12. Summary of the protocol variations for the Ames II and Ames MPF™ assays.

Consistent among the laboratories		Further investigation may be needed
Protocols and decision criteria appear to be uniform across laboratories		Sensitivity may be limited by the frequency of wells without revertants, both in solvent controls and with test chemical
All strains described in TG471 (OECD, 1997) can be used in the Ames MPF™ assay		For Ames II™ somewhat fewer test chemicals were submitted; this may reduce coverage of portions of the applicability domain
Most laboratories reported 5 strain experiments for the Ames MPF™ assay		

3.3. Data curation and preparation of the master dataset for analysis

92. Although all data submitted as part of the retrospective review employed a customised data submission spreadsheet, some formatting and data entry issues were encountered. These needed to be addressed before the supplied data could be concatenated and analysed using SAS/STAT v14.1 in SAS v9.4 for Windows (SAS Institute, 2015) and the epiR package in R v.4.1.0 (Steven and Sergeant, 2021; R Core Team, 2020). Before the start of the data analysis, submitted data were reviewed and curated to ensure dataset uniformity. For example, all submitted datasets were screened to identify missing values (e.g., assessment calls), deviations from the prescribed data entry format, and/or problematic formatting (e.g., test date format, use of special characters such as μ and β). In some cases, the submitters were contacted and asked to suitably address data entry issues, e.g., entry duplication. Overall, all datasets were curated (i) to ensure compatibility with the data analysis software (e.g., removal/replacement of special characters and adjustment of test date format), (ii) to insert unique identifiers for coded test chemicals, (iii) to insert abbreviated variable names, (iv) to screen and correct full test chemical names and CAS numbers, (v) to assign an abbreviated chemical name for test chemicals judged to be biologically equivalent (e.g., 9AA was employed as the abbreviation for both 9-Aminoacridine HCl and 9-Aminoacridine hydrochloride monohydrate), (vi) to screen assessment calls (see below), and (vi) to combine all submitted data into a single worksheet (i.e., where submitted data for each test chemical were provided in separate worksheets). With respect to unique identifiers for coded test chemicals, test chemical abbreviations were inserted using the Submitter ID (e.g., a single letter) and a compound number/letter, or the Submitter's unique substance identifier, e.g., Compound G_1, Test9.

93. Following curation, all datasets were imported into SAS v.4 for Windows, concatenated, and the final master dataset screened to identify and correct any remaining inconsistencies and formatting problems. The most common problems related to data entry spelling errors and inconsistencies in (i) test chemical name spelling and format, (ii) test chemical abbreviations, (iii) the spelling and format of solvents, (iv) in the formatting and/or spelling of bacterial strain terms, and (v) in CAS numbers. The final master dataset, which included all recorded responses for all tested concentrations, comprised 61,317 observations and 32 variables. The qualitative-call dataset comprised 8727 assessments; including 429 test chemicals, of which 188 are coded. The latter includes a single call entry for each assay-test chemical-strain-S9-mix combination. For the purposes of the data analyses described below, an assessment refers to the results obtained for a combination of assay, test chemical, strain and S9-mix, e.g., sodium azide tested on the 6-well plate assay using strain *S. typhimurium* TA100 without S9-mix.

3.3.1. Uniformity of assessment calls

94. Submitters were asked to provide a single assessment call (i.e., negative, positive, equivocal) for each experiment for which, in one assay, various concentrations of the same test chemical were applied to a single strain-S9-mix combination. Where submitters inserted multiple call entries, the entries were screened, and a single call inserted in line with the instructions provided (see 3.1). For example, if one concentration was indicated as positive and all other concentrations were indicated as negative, then a single positive call was inserted for that assay-test chemical-strain-S9-mix combination. Where no call was provided by the submitter, the EG inserted a call according to the following assessment criteria:

For miniaturised agar plating tests:

- For all strains except *S. typhimurium* TA1535 and TA1537, a data set is positive if it provides a concentration-related increase in mean revertants with an increase in at least one concentration ≥2-times the mean solvent control value
- For strains *S. typhimurium* TA1535 and TA1537, a data set is positive if it provides a concentration-related increase in mean revertants with an increase in at least one concentration ≥3-times the mean solvent control value
- For 24-well plate assays, a further criterion is included, that any increase is required to be greater than 6 revertants above the solvent level to be considered as positive

For microfluctuation assays:

- A positive result is indicated by a clear concentration-response and fold increase ≥2-times the solvent control value or
- A positive result is indicated by an increase of ≥2-times the baseline (the baseline is obtained by adding one standard deviation to the mean number of positive wells of the solvent control. If the baseline is less than 1.0, the value should be set to 1.0)

3.3.2. Handling of 'multiple assessments' and 'mixed calls'

95. The evaluation of the correspondence between the miniaturised versions of the bacterial reverse gene mutation assay and the standard assay (i.e., the point of reference) requires a single entry for each individual assessment, where an individual assessment is defined as an assay-test chemical-strain-S9-mix combination. However, review of the master dataset revealed multiple entries for selected assessments. Some test chemicals were tested multiple times in the same assay by the same laboratory. In addition, some test chemicals were tested by more than one laboratory. For example, the master dataset included 13 entries for 2-aminoanthracene tested using the standard assay on TA98 with S9-mix. Indeed, for the standard assay, only 64% of the master dataset included single individual assessment entries, i.e., a single entry for a test chemical-strain-S9-mix combination. The remainder of the standard assay portion of dataset included between 2 and 17 entries for selected test chemical-strain-S9-mix combinations (i.e., individual assessments).

96. All assessments with multiple entries were isolated, and subsequently screened to determine a single consensus call. In cases where all calls were in agreement across the replicated entries, a single consensus call was inserted (e.g., positive, negative, equivocal). In cases where there was disagreement amongst the multiple calls for a given individual

assessment, the term ‘mixed’ was inserted. Almost all assessments with disagreements amongst the multiple calls were permanently labeled as mixed, and the assessments subsequently excluded from the correspondence analyses described below. Exceptions were made for some mixed calls for positive controls listed in OECD TG471 (OECD, 1997). After screening 54 instances of mixed positive control calls, 8 were changed to a consensus call of positive or negative based on consistent results across numerous assessments (i.e., >75% of calls in agreement). For example, of the 15 entries for sodium azide tested on the 6-well plate assay with *S. typhimurium* TA100 without S9-mix, 14 were positive; review by a EG subgroup of five people resulted in a consensus call designation of positive. Criteria employed in these EG reviews included the number of calls that supported a particular consensus call, and the widely accepted mutagenicity classification for the test chemical in question in that strain and activation condition. Where there was no clear majority of calls supporting a given consensus (i.e., <75% of calls in agreement), and/or there was no unanimous agreement amongst the five EG members on an overall consensus call, the overall call for that assay-positive control-strain-S9-mix combination remained mixed. In all other cases (i.e., test chemicals that are not positive controls specified in OECD TG471 (OECD, 1997)), the term mixed was retained. As noted, assessments denoted as mixed were excluded from the correspondence analyses described below.

3.3.3. Master dataset description

97. For the purposes of the retrospective performance analyses described herein, only the qualitative call dataset was analysed, i.e., the analyses did not scrutinize the concentration-response relationships for any individual assessments. The tables below (**Table 13** and **Table 14**) provide a brief descriptive overview of the qualitative call dataset employed for the performance evaluation of four miniaturised versions of the bacterial reverse gene mutation assays (i.e., 6-well plate, 24-well plate, Ames MPF™ and Ames II). It is important to note that the data for each assay included a preponderance of negative assessments, i.e., negative calls for a given test chemical-strain-S9-mix combination (see below). A listing of test chemicals included in the master dataset is provided in Annex 2 Supplementary tables Table 48.

Table 13. Descriptive Overview of Collected Data By Assay Type

Variable	Assay Type				
	Standard[1]	6-well plate	Ames MPF™	Ames II	24-well plate
# of Assessments[2]	3599	1503	2540	539	326
# of test chemicals (coded test chemicals)	352 (170)	140 (96)	179 (30)	144 (30)	57 (22)
Without S9-mix (%)	48.8	50.1	50.0	49.9	49.7
With S9-mix (%)	51.2	49.9	50.0	50.1	50.3
Positive Frequency (%)	18.4	24.0	20.6	21.3	38.3
Negative Frequency (%)	78.5	75.1	73.3	75.3	59.8
Equivocal Frequency (%)	3.1	0.9	6.1	3.3	1.8

[1]Plate incorporation and preincubation assays. 47% of assessments preincubation.

[2]Number of individual assessments = counts of test chemical-strain-S9-mix combinations.

Table 14. Descriptive Overview of Collected Data – Strain Use Comparison by Assay Type

Variable	Assay type				
	Standard[1]	6-well plate	Ames MPF™	Ames II	24-well plate
E. coli WP strains frequency (%)[1]	11.0	17.9	15.0		16.6
S. typhimurium TA100 frequency (%)	22.9	22.4	22.4	56.4 (TAMix)	28.8
S. typhimurium TA98 frequency (%)	26.5	22.9	25.4	43.6	27.9
S. typhimurium TA102 frequency (%)	1.7	1.7	0.2		0.3
S. typhimurium TA1535 frequency (%)	19.2	16.7	19.8		12.0
S. typhimurium TA1537 frequency (%)	14.0	10.7	15.8		1.2
S. typhimurium TA97 frequency (%)	4.8	7.6	1.6		13.2

[1]Frequency values are proportions of total individual assessments

98. The master dataset was also described by applying the chemical space concept. The latter is broadly used in drug discovery because of its multiple potential applications; for instance, for exploration of structure-activity relationships (SAR) and navigation through structure-property relationships (SPR), in library design, for compound or dataset description and classification and compound selection (Naveja and Medina-Franco, 2019). Although it is difficult in general to judge chemical space adequacy, its application in the context of the retrospective validation study is relevant to describe the test chemicals and to make sure that known main structural alerts for mutagens (e.g., alkylating agents, etc.) are represented among the test chemicals. For this reason, the chemical space of the test chemicals under investigation was characterised and compared also with the chemical space of known substances i.e., approved drugs in the DrugBank database and industrial chemicals registered under REACH.

99. The analysis and visualization of the chemical space were carried using the method of the principal component analysis (PCA) which aims to capture and report the maximal variance of a dataset for a user-defined set of features, such as fingerprints or molecular descriptors (based on chemical structure and/or physicochemical properties). Basically, the dataset is projected onto a novel reference system defined by principal components (PC), each being a linear combination of the features orthogonal to the other PCs, meaning that it provides unique information. The PCs are associated to a value, generally expressed as a percentage, quantifying how much of the dataset variance is captured by the PC, and ranked in decreasing order.

100. This exercise was initially performed on the test chemicals with known and disclosed chemical structures for each individual assay. Coded test chemicals were excluded. First, a list comprising all test chemicals with known structure was generated. For each test chemical, it was noted in which assay it was tested (i.e., standard, 6-well plate, 24-well plate, Ames MPF™, Ames II). The workflow which is briefly described below, was implemented in the KNIME Analytics Platform (version 4.1.2) [Cebron et al., 2007] using only publicly available nodes. The full list contains 255 SMILES (Simplified Molecular Input Line Entry System) strings for the test chemicals. The SMILES format can be readily used as an input for the calculation of molecular descriptors as well as molecular fingerprints. Using the SMILES, the physicochemical properties available in the RDKit (open-source cheminformatics tool) [Landrum, 2015]; descriptor calculation node (RDKit KNIME integration version 4.0.1.v202002121352) were calculated. Out of the 119 available descriptors, 116 descriptors were calculated. The descriptor average molecular weight (AMW) was excluded as it is a duplicate of the exact molecular weight descriptor as well as the descriptors “Number of specified/unspecified stereo-centers” (NumStereocenters and NumUnspecifiedStereocenters) as they are mostly used to check the quality of the chemical structure rather than describing molecular properties related to activity. The included descriptors define for example type and number of atoms, bonds and rings present in the molecule or polarity and solubility.

101. As the bacterial reverse gene mutation test outcome is not only dependent on physicochemical properties, but also on structural features, the topology of molecules was described by 1024 bit-vector Morgan Fingerprints (ECFP) (diameter=2) [Rogers and Hahn, 2010], calculated with the corresponding KNIME RDKit fingerprint calculation node. The 2D

fingerprints encode circular atom neighborhood in a hashed bit string. Hence, each bit of the fingerprint represents the presence or absence of a structural feature/ substructure. Dimensionality reduction was performed on the full list using PCA. As calculation of a PCA requires standardised numerical values, the 1D descriptors were subjected to Z-score normalization using the "Normalizer" node in KNIME. Thus, the descriptors were standardised to a mean of 0 and a standard deviation of 1. For the fingerprints (2D) this was not necessary, because they contain only binary values.

102. The PCA was calculated using the respective KNIME nodes (PCA compute and PCA apply). The initial PCA used all available SMILES to span the chemical space. Subsequently, the first and second components were used to plot the chemical space and the test chemicals were colored by assay. The plots in **Figure 4** were created using the R library ggplot2 in an R View node within KNIME [R Core Team, 2020; Wickham, 2016].

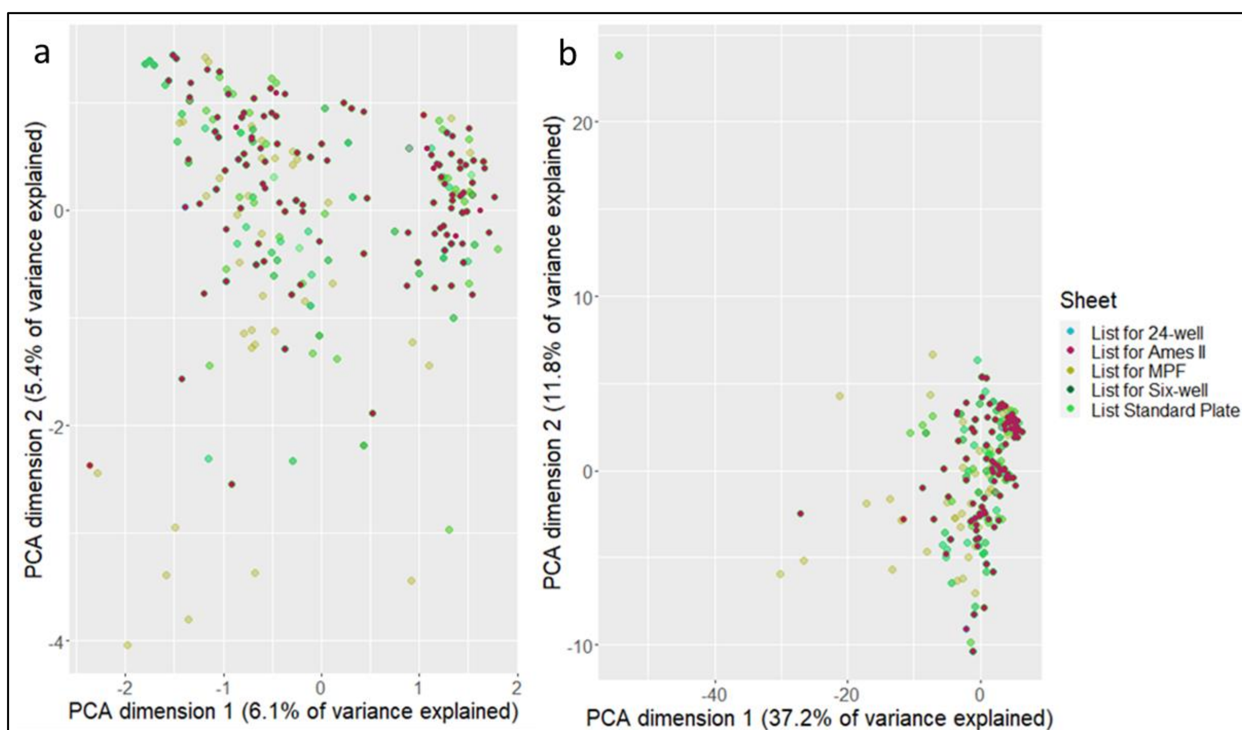

Figure 4. Visualization of the chemical space of the test chemicals used in the different bacterial reverse gene mutation tests as represented by the first and second principal component.

The PCA of chemicals tested in the different assay types (standard plate, 6-well, 24-well, Ames MPF^{TM}, Ames II, was calculated using a) Morgan fingerprints with a diameter of 2, and b) using RDKit physicochemical chemical descriptors. The colors represent the different Assay types.

103. The PCA analysis can help the description of the test chemicals used in the study by evaluating their structural or physicochemical properties ('descriptors'). Figure 4a reports the analysis using Morgan fingerprints, which describe structural features, the topology of molecules; Figure 4b instead, describes the test chemicals by defining for example type and number of atoms, bonds and rings present in the molecule or polarity and solubility. The

descriptor space (figure 3b) is able to explain 50% of the variance and it is better to describe the high dimensionality of the space compared to the fingerprints that explained only 12% of the variance.

104. In the PCA derived from the fingerprints, most of the test chemicals have a second principal component above (–2) with few outliers which comprise of structurally distinct chemicals such as vinblastine, erythromycin or zafirlukast. In the descriptor space for the first principal component this is represented by a high importance of descriptors that are related to the molecular size such as the molecular weight or the number of heavy atoms. Most of those test chemicals, however, were only tested in the Ames MPF™ assay showing a slight bias towards structurally complex test chemicals. The descriptor space has a higher density than the fingerprint space. This is mainly due to streptomycin, which presents itself as an outlier in the upper left corner being significantly larger than other molecules in the dataset. The slight bias of molecules tested in the Ames MPF™ assay is also showing as more of such chemicals are having a first principal component smaller than (–10). Nevertheless, in both plots it can be observed, that the test chemicals are equally distributed over the different assay types. It is important to note that describing molecular similarity in a toxicological context remains a challenging task. A key aspect is the selection of the most appropriate molecular descriptor(s) [Mellor et al., 2019]. The described plots are currently only informative for the chemical space that was tested within the study and limited to the comparison between the different assay types.

105. Thus, as mentioned above, the analysis was extended to better understand biases in the chemical space of the study with respect to the whole chemical space 'Universe' of known approved drugs and registered industrial chemicals, roughly 10000 substances. As such, insights can be obtained whether or not the 255 test chemicals are biased toward certain structural- or physicochemical properties. Thus, the analysis was extended to include a comparison with known registered substances included in REACH (downloaded from the OECD Toolbox, n=7404) and DrugBank approved drugs (n = 2509) [Wishart et al., 2017]. PCA plots were created using the Morgan Fingerprints and RDKit descriptors, as above. The PCA was calculated on each database separately, and then the calculated PCA was applied to the complete list from all assays (**Figure 5**). This approach allows to identify whether an important parameter in the PCA differs in the study dataset as in such a case the chemical spaces would be shifted or even fully separated. As shown in **Figure 5**, it should be noted, that in all PCA plots the test chemicals for the bacterial reverse gene mutation test dataset are in densely populated areas for the descriptor-based PCA (physicochemical) and spread over the chemical space for the fingerprint-based PCA (structure similarity).

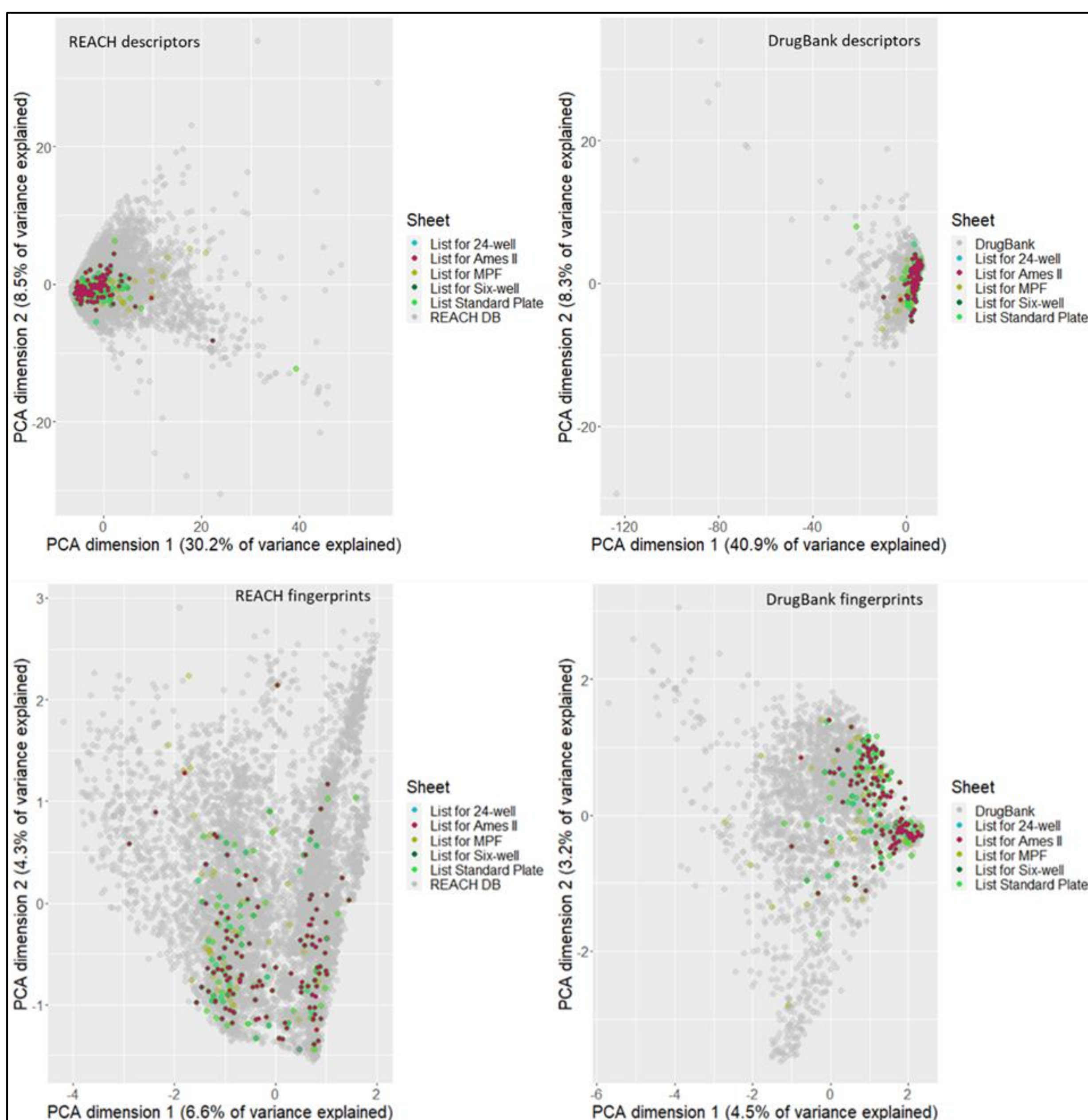

Figure 5. Projection of the chemical space of the bacterial reverse gene mutation tests onto the REACH and DrugBank databases: Physicochemical descriptor-based PCA (upper panels) and fingerprint-based PCA (lower panels).

The PCA of chemicals tested in the different assay types (standard plate, 6-well, 24-well, Ames MPF™, Ames II, was calculated using a) Morgan fingerprints with a diameter of 2, and b) using RDKit physicochemical chemical descriptors. The colours represent the different Assay types.

106. The distribution of the chemicals included in the retrospective validation study over the whole chemical space 'Universe' is also appreciated in a 3D representation (**Figure 6-Figure 10**) which is based on PCA but focusing on the structural similarity of the molecules. In this case, the PCA was performed on the similarity matrix obtained from the atomic pairs

fingerprints of the molecules. The PCs of this analysis correspond to measures of "similarity" and cannot be translated to single properties. PC1 and PC2 contain the "similarity" measure that explains the largest variance of the molecules. Thus, in this type of representation, the distribution of chemicals is purely based on the chemical similarity, preventing clusters of molecules due to a high influence of single properties which may not be related to reactivity in the bacterial reverse gene mutation test as can be molecular weight. The dots that are found further apart correspond to substances that are structurally dissimilar. The analysis was carried out with the same software described above. The application was recently used to describe the chemical space of EURL ECVAM database of test chemicals negative in the standard bacterial reverse gene mutation test [Madia et al., 2020].

107. In the following figures (**Figure 6-Figure 10**), the 3D distribution of chemicals from Drugbank or REACH database and those tested by each single assay type is shown singularly and combined.

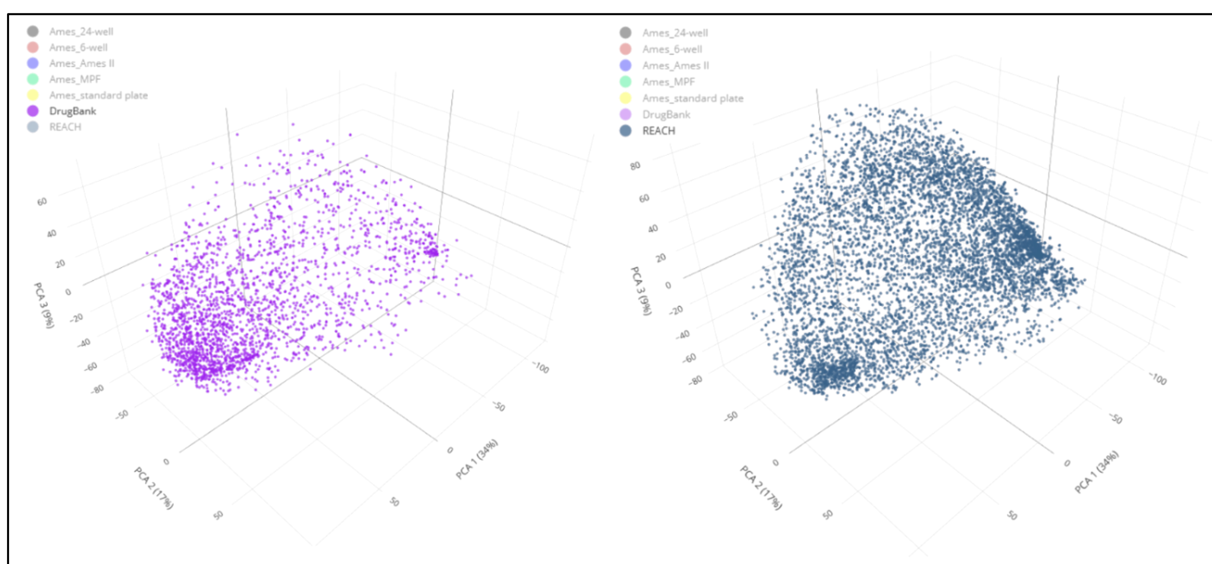

Figure 6. 3D representation of the chemical space (distribution based on structure similarity) of (left) DrugBank approved drugs (n = 2509) and (right) REACH registered industrial chemicals with high quality SMILES-structure-Name relationship (downloaded from the OECD Toolbox n=7404).

Axes and positions of the substances correspond to the first two principal components of the similarity matrix obtained from atom pairs fingerprints. The variance explained by each PC is indicated in parenthesis. It is worth noting that approved drugs and registered REACH chemicals distribute unevenly around the whole space, thus indicating that certain chemical structures are of less use and/or commercial interest (area in the middle).

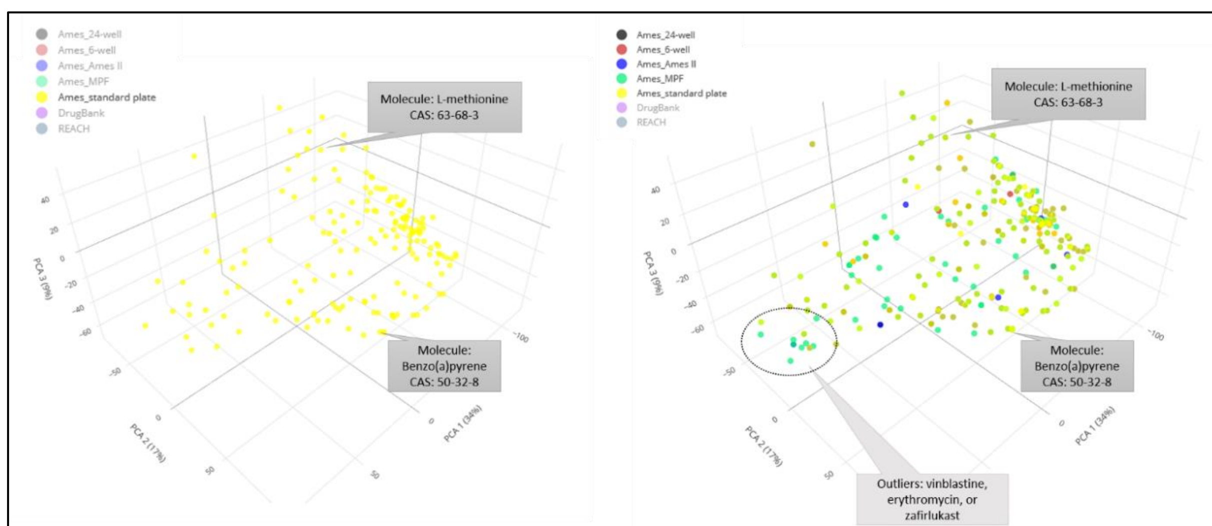

Figure 7. Details of the chemical space covered by chemicals tested either in the standard (left) or in all the miniaturised versions (right) of the bacterial reverse gene mutation assay.

Chemicals have been subdivided by assays types: standard (n=171), 6-well (n=45), 24-well (n=38), Ames MPF^{TM} (n=148), Ames II (n=111) for a total unique chemicals number, n = 227. The position of two representative chemicals, L-methionine and Benzo(a)pyrene is highlighted. Few outliers which comprise of structurally distinct chemicals such as vinblastine, erythromycin or zafirlukast were tested with Ames MPF^{TM} (circle). Axes and positions of the substances correspond to the first two principal components of the similarity matrix obtained from atom pairs fingerprints. The variance explained by each PC is indicated in parenthesis.

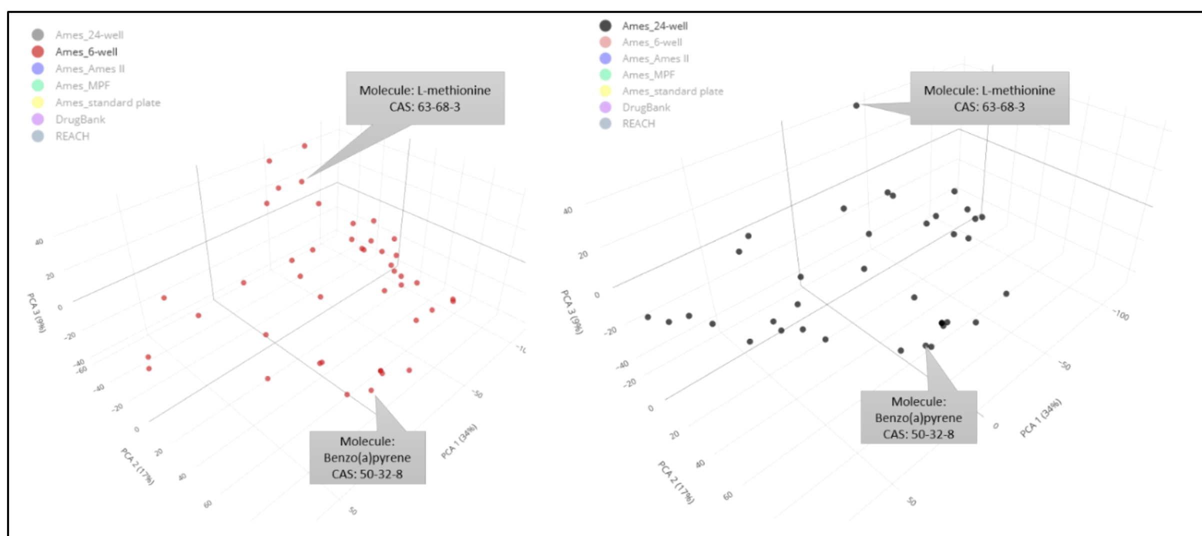

Figure 8. Details of the chemical space covered by chemicals tested either in 6-well (left) or 24-well (right) plate versions of the bacterial reverse gene mutation assay.

Chemicals have been subdivided by assays types: 6-well (n=45) and 24-well (n=38). The position of two representative chemicals, L-methionine and Benzo(a)pyrene tested in both miniaturised bacterial reverse gene mutation assays is reported. Axes and positions of the substances correspond to the first

two principal components of the similarity matrix obtained from atom pairs fingerprints. The variance explained by each PC is indicated in parenthesis.

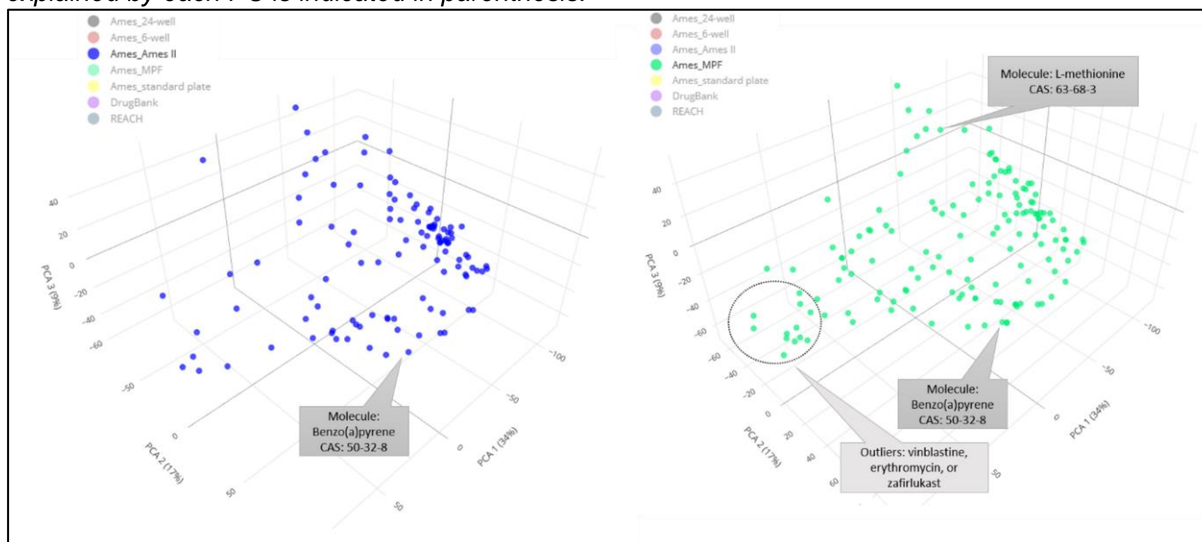

Figure 9. Details of the chemical space covered by chemicals tested either in the Ames II (left) or Ames MPF™ (right) versions of the bacterial reverse gene mutation assay.

Ames substances have been subdivided by assays types: Ames MPF™ (n=148) and Ames II (n=111). The position of two representative chemicals L-Methionine (tested in Ames MPF™ only) and Benzo(a)pyrene is reported. Few outliers which comprise of structurally distinct chemicals such as vinblastine, erythromycin or zafirlukast were tested with Ames MPF™ (circle). Axes and positions of the substances correspond to the first two principal components of the similarity matrix obtained from atom pairs fingerprints. The variance explained by each PC is indicated in parenthesis.

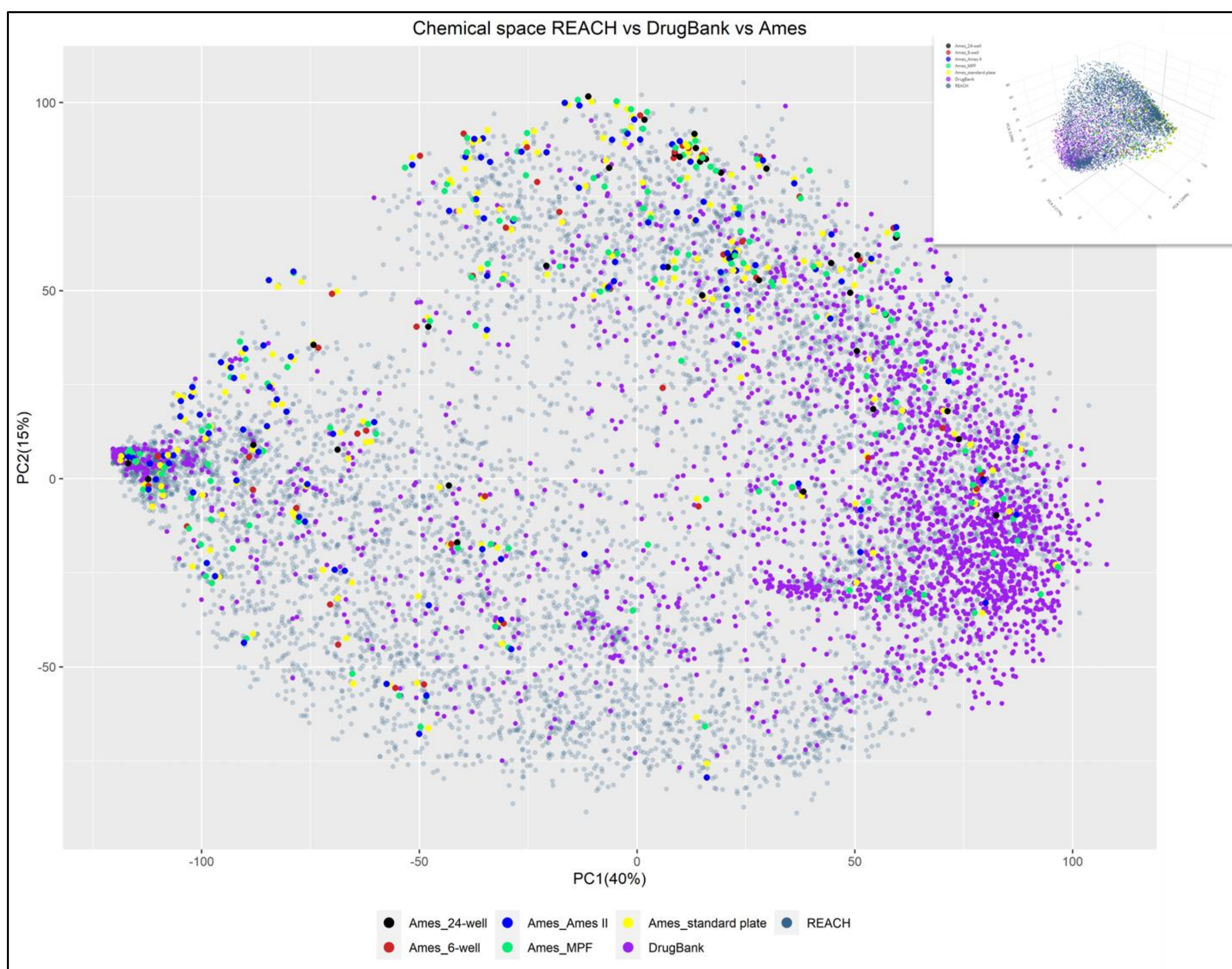

Figure 10. Chemical space covered by the chemicals tested in the various versions of bacterial reverse gene mutation assays (n=227) with REACH registered chemicals with high quality SMILES-structure-Name relationship (downloaded from the OECD Toolbox n=7404) and DrugBank approved drugs (n = 2509).

The plot is a 2D representation of the distribution of test chemicals (with known structure), collected in the retrospective study as compared to the chemical space for registered industrial chemicals and drugs. Chemicals are listed by assays types: standard (n=171), 6-well (n=45), 24-well (n=38), Ames MPF™ (n=148) and Ames II (n=111) for a total unique chemicals number, n = 227. Axes and positions of the substances correspond to the first two principal components of the similarity matrix obtained from atom pairs fingerprints. The variance explained by each PC is indicated in parenthesis. The reference 3D plot is depicted in the upper right corner.

108. The PCA analysis aimed to provide a first qualitative description of the space of test chemicals collected for the retrospective study. Specifically the aim was to identify overlaps or differences among the chemicals tested in the different assays and in comparison with the whole 'universe' of known chemicals (registered under REACH) and drugs (DrugBank database).

109. It was possible to perform the analysis on those test chemicals for which the chemical identity was disclosed: in total >220 test chemicals. This represented only a fraction of all the tested chemicals across the assays. Nevertheless, it was observed that the majority of test chemicals were well distributed across the space and with comparable physicochemical properties. This indicated that the collection contained chemicals with typical physicochemical properties (i.e., rather small molecules with average mass <500 g/mol as most of chemicals in the market are below 500 g/mol; logP in the range of -5 and 5) and that it was not biased towards certain types of chemicals with very specific properties, except for few outliers. Those presented a very high average mass, compared to the rest of chemicals. Streptomycin sulfate, C.I. Direct Blue, zafirlukast, erythromycin, Paclitaxel presented almost 2-fold higher mass thus, driving the higher density observed in the physicochemical descriptors PCA analysis (Figure 4, upper panels). Interestingly, several of the outliers have been tested in the Ames MPF™ and Ames II; streptomycin sulphate was tested also in the standard plate. Nevertheless, only for this chemical, a mismatch was observed in the correspondence analysis (Table 60). The PCA analysis cannot explain the exact relationship between molecular similarities and toxicological effects but confirms that the test chemicals used for this analysis were not allocated to a specific coordinate of the chemical space of the bacterial reverse gene mutation test and represent a good portion of the space. Further analysis, i.e., including functional group characterization, may add to the explanation on the mode of action of each single test chemical, even if this should apply to chemicals tested in the standard bacterial reverse gene mutation test as well. Eventually, the analysis could also be helpful for the selection of reference test chemicals.

3.4. Retrospective data analysis

3.4.1. Data analysis approach

110. The curated and appropriately-formatted master dataset was used to examine binary correspondences between the results obtained using selected miniaturised versions of the bacterial reverse gene mutation assay and the results obtained using the standard assay. Questions addressed relate to the performance of each miniaturised assay, with performance defined as the ability to reliably assess mutagenic activity relative to the standard assay. In all cases, the standard bacterial reserve gene mutation test was used as the point of reference. In other words, for the purposes of these analyses, the standard assay was defined as the ‘gold standard’. Standard assay results included both plate incorporation assay data and the preincubation assay data. All performance and agreement analyses were conducted using SAS/STAT v14.1 in SAS v9.4 for Windows (SAS Institute, 2015) and the epiR package v2.0.19 in R v4.1.0 (Steven and Sergeant, 2021; R Core Team, 2020).

111. The analyses conducted herein followed the approach recently employed by Kirkland et al (2019) for their comparisons of the *in vivo* transgenic rodent (TGR) gene mutation assays and the *in vivo* comet assay. More specifically, 2x2 contingency tables (i.e., cross-tabulation tables) were constructed for each desired analysis, and a host of performance metrics and agreement statistics employed to evaluate the strength of the correspondence between the miniaturised assays and the point of reference (i.e., the standard assay). Following Byrt et al. (1993), **Table 15** illustrates the format of the contingency tables used for the analyses. The table values *a*, *b*, *c* and *d* indicate the incidence of each assay outcome combination, as well

as the row and column totals (i.e., g_1, g_2, f_1 and f_2). More specifically, *a* indicates the 'correct negative' frequency, *d* indicates the 'correct positive' frequency, *b* and *c* indicate the 'false positive' and 'false negative' frequencies, respectively. Values g_1 and g_2 are the total frequencies of standard assay negatives and positives, respectively; f_1 and f_2 are the total frequencies of negatives and positives for the miniaturised assay under investigation. Total dataset negatives and positives were calculated as (g_1+f_1) and (g_2+f_2), respectively; *N* is the total sample size.

Table 15. Format of the 2x2 contingency tables used for correspondence analyses.

		Miniaturised Assay Call		
		Negative	Positive	Row Totals
Standard Assay Call	Negative	*a*	*b*	g_1
	Positive	*c*	*d*	g_2
	Column Totals	f_1	f_2	***Total N***

112. The retrospective data analyses employed six metrics to evaluate the performance of the miniaturised assays under investigation. These included *Sensitivity*, the proportion of standard assay positives (i.e., *true* positives) that are correctly detected by the miniaturised assay, *Specificity*, the proportion of standard assay negatives (i.e., *true* negatives) that are correctly detected by the miniaturised assay, *Positive Predictive Value* (PPV), the proportion of miniaturised assay positives that are correct relative to the standard assay, and the *Negative Predictive Value* (NPV), the proportion of miniaturised assay negatives that are correct relative to the standard assay. With respect to the PPV and NPV values more specifically, they indicate the proportions of miniaturised assay positive and negative calls that are *true* positives and *true* negatives, respectively (Altman and Bland, 1994). Additionally, the complements of sensitivity and specificity were calculated; specifically, the *False Negative* rate (i.e., the proportion of standard assay positives that are incorrectly designated negative by the miniaturised assay), and the *False Positive* rate (i.e., the proportion of standard assay negatives that are incorrectly designated positive by the miniaturised assay). With respect to **Table 15** above, the performance metrics were calculated as:

Sensitivity (%) = correct positives/total standard assay positives = $d/(g_2)*100$
Specificity (%) = correct negatives/total standard assay negatives = $a/(g_1)*100$
Positive Predictive Value (PPV) (%) = correct miniaturised assay positives/total miniaturised assay positives = d/f_2*100
Negative Predictive Value (NPV) (%) = correct miniaturised assay negatives/total miniaturised assay negatives = a/f_1*100
False Positive Rate (%) = miniaturised assay positives where standard is negative/total standard assay negatives = $b/(g_1)*100$
False Negative Rate (%) = miniaturised assay negatives where standard is positive/total standard assay positives = $c/(g_2)*100$

113. The strength of the agreements between the two assessment methods (i.e., miniaturised assay and the standard assay) were evaluated using the Cohen's Kappa statistic

(κ) (Cohen, 1960; Landis and Koch, 1977). Kappa values were calculated according to Equation 1, where *Po* is the observed agreement between assay pairs, and *Pe* is the hypothetical probability of chance agreement. *Po* and *Pe* values were calculated according to equations 2 and 3; see **Table 15** for variable definitions (Nurjannah and Siwi, 2017).

$$\kappa = \frac{(Po-Pe)}{1-Pe} \quad (1)$$

$$Po = \frac{(a+d)}{N} \quad (2)$$

$$Pe = Ppositive + Pnegative = \left(\frac{g2}{N} \times \frac{f2}{N}\right) + \left(\frac{g1}{N} \times \frac{f1}{N}\right) \quad (3)$$

However, as noted by Byrt et al. (1993) and Kirkland et al. (2019), the Kappa statistic is not reliable when response prevalence is unbalanced and/or response proportions are biased. Following Byrt et al. (1993) and Flight and Julious (2015), Kirkland et al. (2019) recommended the use of the Prevalence Index (PI) and Bias Index (PI) to assess response prevalence asymmetry and bias. With respect to **Table 15,** the PI is calculated as $(a-d)/N$; the BI is calculated as $(b-c)/N$. The BI reflects the differences in the proportions of positive and negative calls for the assay formats being compared (i.e., miniaturised assay and standard plate incorporation assay). The PI reflects the balance between the prevalence of negative and positive results for the two assessment methods being compared, i.e., total negatives ($f_1 + g_1$) relative to total positives ($f_2 + g_2$). Larger absolute values of PI and BI are indicative of increasing response asymmetry and bias, respectively. A PI of zero indicates balanced prevalence of negatives and positives; a BI of zero indicates lack of bias. Calculation and scrutiny of PI and BI values were used to determine the suitability of agreement statistics such as Cohen's Kappa.

114. Flight and Julious (2015) and Byrt et al. (1993) noted that contingency table values that yield a high PI, which reflects lack of balance between the prevalence of negatives and positives, can deflate Kappa values. Thus, as noted by Kirkland et al. (2019), even for high concordance, Kappa values can be unexpectedly low. Although there is no threshold beyond which a PI value can be uniformly designated as excessive, problematic deflation of high Kappa values (i.e., >0.7) generally requires PI values greater than about 0.5 (Burn and Weir, 2011). PI is generally not considered problematic if Kappa is >0.9. High bias (BI), which reflects differences between the relative probabilities of negatives and positives for the two assessment methods, can, in contrast, yield unexpectedly high Kappa values. This so-called Kappa paradox (Cichetti and Feinstein, 1990; Feinstein and Cichetti, 1990), whereby Kappa values are deflated or inflated, can be addressed via the use of more robust statistics that adjust the agreement rating for unbalanced response prevalence and response bias. Following Byrt et al. (1993), Kirkland et al. (2019) noted that the Prevalence-and-Bias-Adjusted-Kappa (PABAK) statistic is a suitable alternative to the standard Kappa value; more specifically, the PABAK is less influenced by contingency table asymmetry and bias. Consequently, in accordance with the recommendations of Byrt et al. (1993) and Kirkland et al. (2019), the agreement analyses conducted herein employed two statistics, i.e., Kappa and PABAK. PABAK values were calculated according to equation 4. Although there is controversy regarding appropriate interpretation of agreement statistics such as Kappa and PABAK, values presented herein were interpreted following the recommendations of Landis and Koch (1977), as cited in Cunningham (2009) (**Table 16**).

$$PABAK = 2Po - 1 \quad (4)$$

Table 16. Agreement categorization based on Kappa/PABAK statistic values.

Kappa/PABAK Value	Strength of Agreement
<0.20	Poor
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 1.00	Almost perfect

It should be noted that although these categories might be considered useful for assessing the performance of the miniaturised assays examined herein, they should be interpreted with caution. Moreover, they should not be over-interpreted. Indeed, Kirkland et al. (2019) noted that confidence limits on performance statistics such as Kappa and PABAK will be wide when sample sizes are ≤95. In such cases, Kirkland et al. (2019) suggest that the lower confidence limits of the performance statistics be used to evaluate performance according to the categorization scheme presented in **Table 16**.

115. Analyses of 2x2 contingency tables proceeded incrementally, beginning with an evaluation of the correspondences based on all individual assessments, where, as noted earlier, each individual assessment is defined as the call corresponding to a test chemical-strain-S9-mix combination. The correspondence analysis did not require that standard and miniaturised assay data for an assessment come from the same lab(s). This was followed by evaluations of correspondences based on two types of 'overall calls'. Since overall call data based on the results for a test chemical collected in multiple strains are generally used to decide whether or not a test chemical is a mutagen, it could be argued that the overall call analyses are most important. Within this retrospective validation study, firstly, overall calls were determined using the responses in all 5 bacterial strains recommended in OECD TG471 (OECD, 1997), i.e., (i) *S. typhimurium* TA98, (ii) *S. typhimurium* TA100, (iii) *S. typhimurium* TA1535, (iv) *S. typhimurium* TA97, 97a or 1537, and (v) *E. coli* WP2 or *S. typhimurium* TA102. The overall call designation did not differentiate between the different genotypes of *E. coli* WP2. The second type of overall call correspondence analyses employed overall calls based on (i) *S. typhimurium* TA98 and (ii) TA100 responses only. With respect to the former, a 5-strain overall call was only possible if 10 assessments were available to fullfill the required conditions, i.e., a call in each of 5 strains with and without S9-mix. For this analysis it was not necessary that all 10 assessments came from the same laboratory. Similarly, a 2-strain overall call was only possible if calls were available for all required conditions, i.e., 4 assessments arising from tests using the *S. typhimurium* TA98 and TA100 strain with and without S9-mix. A positive response on any of the test conditions was deemed sufficient for an overall positive call, i.e., positive on *any* of 10 conditions for the 5-strain overall call or positive on *any* of 4 conditions for the 2-strain overall call. Designation of a negative overall call required a negative response call across all required test conditions, i.e., negative on *all* of the 10 conditions for the 5-strain overall call, or negative on *all* of the 4 conditions for the 2-strain overall call. For the Ames II assay, both types of overall calls were based on results for TA98 and TAMix only. For the purposes of the performance analyses conducted herein, the data were deemed acceptable if the row and column totals are all >10.

116. For each analysis, the aforementioned performance metrics and agreement statistics are provided, along with their respective 95% confidence intervals. Confidence intervals on performance metrics were determined using the aforementioned epiR package, following the method of Simel et al. (1991). The Simel et al. method assumes that proportional performance

metrics (e.g., sensitivity and specificity) are drawn from a normal distribution. Confidence intervals on agreement statistics (i.e., Kappa and PABAK) were determined using epiR following the method of Donner and Eliasziw (1992). The Donner and Eliasziw approach is based on a chi-squared goodness-of-fit test that is consistent with models typically used for analyses of binary incidence data.

117. For all initial analyses, equivocal calls were excluded. Follow-up analyses investigated changes in the results if equivocal calls were alternatively denoted positive or negative. Additional analyses investigated strain-specific correspondences, correspondences based only on preincubation assessments, and correspondences based only on standard plate incorporation assessments (i.e., without preincubation).

118. QC (Quality Control) measures were routinely employed to evaluate the accuracy of the datasets used for any particular analysis; moreover, the accuracy of the presented results. For example, for the 5- and 2-strain overall call analyses, the master dataset was inspected to determine the validity of randomly-selected overall calls, i.e., to determine if overall calls correspond with the required spectrum of calls across the relevant strain-S9-mix combinations. Additionally, correspondence call mismatches were routinely screened and enumerated to verify the accuracy of the analyses and the linked performance metrics.

3.4.2. Results of Assay correspondence analyses

119. As noted, performance metrics determined via analyses of 2x2 contingency table data reflect the ability of the miniaturised assays to correctly assign a dichotomous response relative to that observed using the standard assay (i.e., the point of reference). Additionally, the aforementioned agreement statistics reflect the strength of the agreement.

120. As noted in Table 13 above, small fractions of the assessment calls were denoted as equivocal by the data submitters. A list of test chemicals that yielded equivocal responses on each assay are provided in supplementary Table 49-Table 53. As noted by Kirkland et al. (2019), equivocal calls can be attributed to situations where the results obtained are ambiguous or inconsistent within a study (i.e., positive and negative across replicate tests), and/or the response did not achieve statistical significance, and/or did not exceed a recommended threshold (e.g., 2-fold above concurrent control), and/or did not exceed the range of historical controls. It could be argued that equivocal calls are not negative; moreover, that the analyses described below should be repeated with the equivocal calls designated as positive (Kirkland et al., 2019). For completeness, the analyses described below could also be repeated with equivocal calls designated as negative. Both analyses were conducted (i.e., equivocals as positive or equivocals as negative); however, the results obtained failed to show any appreciable changes in performance metrics and agreement statistics (data not shown).

121. The first correspondence analysis investigated agreement based on **individual assessments**, e.g., the calls associated with each test chemical-strain-S9-mix combination. **Table 17** shows the contingency table data for each of the miniaturised assays investigated, i.e., Ames MPF™ assay, 6-well plate assay, 24-well plate assay, and Ames II assay. The performance statistics for each assay are included in **Table 18**. **Table 19** shows the number of test chemicals and assessments included in the analyses, as well as the PI and BI. The

results of all following correspondence analyses are reported using the same format. The results for test chemicals that created correspondence mismatches for each miniaturised assay (i.e., false positives and false negatives) are described in Supplementary Table 54-Table 57. As indicated before, test chemicals for which there were equivocal calls in a single assessment, or mixed results among multiple assessments, were excluded from these analyses.

122. The results obtained show substantial to almost perfect agreement (See **Table 16**) between the miniaturised and standard assays. The high absolute value of the PI for the 6-well, Ames MPF™ and Ames II (TA98 only) assay data suggests evaluating agreement using PABAK rather than Kappa. In all cases, the PABAK value indicates almost perfect agreement (i.e., >0.81), with lower confidence limits at least indicating substantial agreement (See **Table 18**). Here and throughout the described analyses (see below), the absolute values of the BI are all low (i.e., close to zero), indicating lack of bias (See Table 19).

123. High specificity across the assays examined reflects excellent performance; however, there are marked differences in the sensitivity of the microfluctuation assays (i.e., Ames MPF™ and Ames II). The results for the Ames II (TA98 only) and Ames MPF™ assays show lower sensitivity, and a concomitant high frequency of false negatives (i.e., >20%), relative to the 6-well and 24-well plate assays. With respect to the NPV values, the results indicate a high probability that a negative call on any of the miniaturised assays will correctly correspond to a standard assay negative. In contrast, and in congruence with the sensitivity and false negative rate, the PPV is lower for the two microfluctuation tests. This is particularly noteworthy for the Ames II assay; it suggests more frequent misclassification of standard assay mutagens as non-mutagens.

Table 17. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values are based on comparisons of individual assessments (e.g., combinations of test chemical, strain and S9-mix). Equivocal responses excluded from the analyses.

		6-well plate assay			Ames MPF™ assay			Ames II assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	751	12	763	515	13	528	141	8	149	69	5	74
	Positive	11	139	150	39	107	146	5	19	24	6	72	78
	Column Totals	762	151	913	554	120	674	146	27	173	75	77	152

Table 18. Performance of miniaturised bacterial mutations assays based on comparisons of individual assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	92.7 (87.3-96.3)	98.4 (97.3-99.2)	92.1 (86.5-95.8)	98.6 (97.4-99.3)	1.6 (0.8-2.7)	7.3 (3.7-12.7)	0.91 (0.84-0.97)	0.95 (0.92-0.97)
Ames MPF™ assay	73.3 (65.3-80.3)	97.5 (95.8-98.7)	89.2 (82.2-94.1)	93.0 (90.5-94.9)	2.5 (1.3-4.2)	26.7 (19.7-34.7)	0.76 (0.68-0.83)	0.85 (0.80-0.88)
Ames II assay[4]	79.2 (57.8-92.9)	94.6 (89.7-97.6)	70.4 (49.8-86.2)	96.6 (92.2-98.9)	5.4 (2.4-10.3)	20.8 (7.1-42.2)	0.70 (0.55-0.85)	0.85 (0.75-0.92)
24-well plate assay	92.3 (84.0-97.1)	93.2 (84.9-97.8)	93.5 (85.5-97.9)	92.0 (83.4-97.0)	6.8 (2.2-15.1)	7.7 (2.9-16.0)	0.86 (0.70-1.01)	0.86 (0.75-0.93)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$
[4]TA98 Only

Table 19. Number of assessments, number of substances, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on individual assessments; equivocal responses excluded.

Mini Assay	Number of Test chemicals	Number of Assessments	PI	BI
6-well plate assay	121	913	0.67	0.001
Ames MPF™ assay	125	674	0.61	-0.004
24-well plate assay	42	152	-0.02	-0.007
Ames II assay[1]	97	173	0.71	0.017

[1]TA98 Only

124. The second analysis investigated assay correspondences based on **overall calls**, i.e., response calls determined using the results obtained across a series of test conditions. More specifically, correspondences based on (i) 5-strain overall calls, and (ii) 2-strain overall calls. As indicated earlier, overall calls were only possible if all strain-S9-mix conditions for a particular test chemical were available in the dataset, e.g., 10 conditions for a 5-strain overall call and 4 conditions for a 2-strain overall call. A positive response for any of the relevant strain-S9-mix combinations was deemed sufficient for an overall positive call; designation of a negative overall call required a negative response for all of the relevant strain-S9-mix combinations. Overall calls for the Ames II assay were based on results for TA98 and TAMix only.

125. The 6-well plate assay was the only miniaturised assay for which there were sufficient 5-strain overall call data for a correspondence analysis. The results presented below in **Table 20**-**Table 22** show excellent performance of the 6-well plate assay. Indeed, all listed performance metrics indicate improved performance relative to the **individual assessment** results presented in **Table 17**-**Table 19**; the confidence intervals of the performance metrics and agreement statistics reflect their reliability. Given the moderate absolute value of the PI (i.e., > 0.3), the slightly-elevated PABAK is the more reliable indicator of agreement. Analogous to what was noted earlier, the absolute value of the BI was negligible. The data employed for the 5-strain overall call analyses, with assay correspondence mismatches highlighted in bold, are summarised in Supplementary **Table 58**-**Table 61**. For the 6-well plate assay, the single mismatched test chemical was a proprietary compound, as were almost all of the other test chemicals with negative results (i.e., 35 of 45), and most of the remaining positive test chemicals (i.e., 13 of 21).

126. In some instances, insufficient data prohibited meaningful evaluations of test performance. This occurred when the calculated performance metrics were undefined or considered too uncertain by the EG, i.e., extreme confidence intervals. In such cases, Altman and Bland (1994) described the performance parameters as “ineffectual”. They also noted that when positive calls on the reference assay (i.e., standard assay) are rare, the ability to confidently interpret a positive on the alternate assay declines. This is particularly true for the 24-well plate data, as well as 5-strain overall call data for the Ames MPF™ and Ames II assays. For example, the 5-strain overall call contingency table data for the Ames II assay (i.e., no positive call and an incorrect call for a single standard assay positive) results in an undefined PPV. Moreover, the false negative and false positive rates are 100 and 0%, respectively.

Similarly, the 5-strain overall call false negative rate for the Ames MPF^{TM} assay is 16.7%, with a confidence interval from 0.4 to 64%; the confidence limit for the false positive rate is 0-49%. Thus, in accordance with Altman and Bland (1994), the 5-strain overall call performance statistics for the 24-well plate, Ames MPF^{TM} and Ames II assays were considered by the EG to be "ineffectual". Confidence intervals for the several other performance metrics were similarly broad (data not shown); they were also deemed problematic and ineffectual.

127. With respect to the correspondence analyses based on 2-strain overall calls, **Table 23** below shows the contingency table data for each of the miniaturised assays, i.e., Ames MPF^{TM} assay, 6-well plate assay, Ames II assay, and 24-well plate assay. **Table 24** shows the 2-strain overall call performance statistics for each miniaturised assay investigated. With respect to the 24-well assay, performance could not be meaningfully evaluated due to, as noted above, data limitations and the problematic distribution of the result calls. More specifically, the limited row and column totals resulted in imprecise performance statistics that, in accordance with Altman and Bland (1994), were deemed to be uninformative. **Table 25** shows the number of test chemicals included in the 2-strain overall call analysis, as well as the PI and BI values. As noted above, the data employed for these analyses, with assay correspondence mismatches highlighted in bold, are summarised in **Table 58**-**Table 61**.

128. Data were available for enough test chemicals to evaluate performance of all assays except the 24-well plate assay. For the 6-well, Ames MPF^{TM} and Ames II assays, the results show improved performance relative to the **individual assessment** performance data presented in **Table 17**-**Table 19**. Again, given the elevated absolute values of the PI, the PABAK is the more reliable agreement indicator. In all cases, the absolute values of the BI were negligible. Importantly, the results show perfect performance for the 6-well plate assay, with confidence intervals for the performance metrics and agreement statistics similar to those seen in **Table 18** for the **individual assessment** analyses. With respect to the Ames MPF^{TM} and Ames II assays, the results show improved performance relative to the **individual assessment** analyses; however, the improvements should be interpreted with caution due to the wide confidence limits of the performance metrics and agreement statistics. Overall, the results for the Ames MPF^{TM} assay results show marginal improvements in sensitivity and PPV, with a concomitant decline in false negatives. However, the results further show small declines in specificity and NPV, with a concomitant increase in false positives. Although the PPV for the Ames II assay increased to 80%, the confidence limit is 28.4 to 99.5%; this is likely attributable to the rarity of positive calls on the standard assay (i.e., the point of reference). Similarly, although the false negative rate declined slightly to 20%, it is only based on 5 standard assay positives, and the confidence limit is 6.8 to 49.9%. Importantly, 4 of those 5 standard assay positives are listed in TG471 (OECD, 1997) as positive controls, including sodium azide, a positive control for strains *S. typhimurium* TA100 and TA1535 without metabolic activation. Interestingly, sodium azide was the single false negative test chemical in this analysis.

129. In some instances, positive 5-strain overall calls are associated with negative 2-strain overall calls, i.e., circumstances for which a positive call was elicited only on a strain other than *S. typhimurium* TA98 and TA100. In two instances, this resulted in changes in correspondence between the miniaturised assay and the standard assay. CmpdQ8 was positive in the standard assay using 5 strains, but negative when only 2 strains were used, and also negative in Ames MPF^{TM} assays with 5 strains or 2 strains. There was a similar

finding for N2_10 in the 6-well plate assay comparison. Interestingly, there were positive 5-strain calls for 21 test chemicals in both Standard and 6-well plate assays, but 10 of the 21 were negative in both assays when based on 2-strain overall calls.

Table 20. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values based on comparisons of 5-strain overall calls. 5 strains included S. typhimurium TA100, TA98, TA1535, TA1537 or 97 or 97a, and E. coli WP2 strain or S. typhimurium TA102. Strains for Ames II assay included only S. typhimurium TA98 and TAMix. Equivocal responses excluded from the analyses.

		6-well plate assay			Ames MPF™ Assay			Ames II Assay[1]			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	40	0	40	7	0	7	24	0	24	1	0	1
	Positive	1	21	22	1	5	6	1	0	1	0	1	1
	Column Totals	41	21	62	8	5	13	25	0	25	1	1	2

[1]Overall call based on TA98 and TAMix

Table 21. Performance of miniaturised bacterial mutations assays based on comparisons of 5-strain overall call assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	95.4 (77.2-99.9)	100 (91.2-100)	100 (83.9-100)	97.6 (87.1-99.9)	0.0 (0.0-8.8)	4.5 (0.1-23.0)	0.96 (0.72-1.21)	0.97 (0.83-1.00)
Ames MPF™ assay	Insufficient data							
Ames II assay[4]	Insufficient data							
24-well plate assay	Insufficient data							

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$
[4]Overall call based on TA98 and TAMix

Table 22. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on 5-strain overall call assessments; equivocal responses excluded.

Mini Assay	Number of Test Chemicals	PI	BI
6-well plate assay	62	0.31	-0.02
Ames MPF™ assay	13	Insufficient data	
Ames II assay[1]	25	Insufficient data	
24-well plate assay	2	Insufficient data	

[1]Overall call based on TA98 and TAMix

Table 23. 2 x 2 contingency tables showing call frequency correspondence values for each miniaturised assay versus the standard assay. Frequency values based on comparisons of 2-strain overall calls. With the exception of the Ames II assay, calls based on responses for S. typhimurium TA98 and TA100. For the Ames II assay, calls based on S. typhimurium TA98 and TAMix. Equivocal responses excluded from the analyses.

		6-well plate assay			Ames MPF™ Assay			Ames II Assay[1]			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	77	0	77	29	1	30	26	1	27	1	0	1
	Positive	0	20	20	4	13	17	1	4	5	0	6	6
	Column Totals	77	20	97	33	14	47	27	5	32	1	6	7

[1]Overall call based on TA98 and TAMix

Table 24. Performance of miniaturised bacterial mutations assays based on comparisons of 2-strain overall call assessments; equivocal responses excluded. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	100 (83.2-100)	100 (95.3-100)	100 (83.2-100)	100 (95.3-100)	0 (0.0-4.7)	0 (0.0-16.8)	1.0 (0.80-1.20)	1.0 (0.92-1.00)
Ames MPF™ assay	76.5	96.7	92.9	87.9	3.3	23.5	0.76	0.79

	(50.1-93.2)	(82.8-99.9)	(66.1-99.8)	(71.8-96.6)	(0.08-17.2)	(6.8-49.9)	(0.48-1.04)	(0.54-0.93)
Ames II assay[4]	80.0 (28.4-99.5)	96.3 (81.0-99.9)	80.0 (28.4-99.5)	96.3 (81.0-99.9)	3.7 (0.09-19.0)	20.0 (0.5-71.6)	0.76 (0.42-1.11)	0.88 (0.58-98)
24-well plate assay	Insufficient data							

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$
[4]Overall call based on TA98 and TAMix

Table 25. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses based on 2-strain overall call assessments; equivocal responses excluded.

Mini Assay	Number of Test Chemicals	PI	BI
6-well plate assay	97	0.59	0
Ames MPF™ assay	47	0.34	-0.06
Ames II assay[1]	32	0.69	0
24-well plate assay	7	Insufficient data	

[1]Overall call based on TA98 and TAMix

130. Since the Ames MPF™ and Ames II assays are microfluctuation assays, it could be argued that they may be more comparable with the preincubation version of the standard assay, i.e., rather than the plate incorporation version without preincubation. Those analyses, which are based on **individual assessments** (i.e., each test chemical-strain-S9-mix combination), were conducted, and the results presented below in **Table 26**-**Table 28**. For comparative interpretation, these results can be compared to the results presented above in **Table 17**-**Table 19**. Assay correspondence mismatches associated with these analyses are not shown. Removing standard assay plate incorporation data reduced the numbers of test chemicals and assessments available for comparison with the Ames MPF™ assay. There was a small decrease in test chemicals, and an increase in assessments for comparison with the Ames II assay. The elevated absolute values of the PI were similar to those in the previous analyses (e.g., Table 19), necessitating interpretation of the PABAK agreement statistic. The absolute values of the BI were consistently low. The PABAK agreements for these smaller sets of test chemicals were similar to those for the original analyses (i.e., <0.85). The results obtained show decreases in the frequency of false negatives for the Ames MPF™ and Ames II assays, and a small decrease in the frequency of false positives for the Ames MPF™ assay. Importantly, the results show a decline in the PPV for the Ames II assay, indicating a lower probability that this miniaturised assay will generate a positive call that correctly corresponds to a positive call on the preincubation version standard assay.

131. Similar comparisons between the preincubation version of the standard bacterial reverse gene mutation assay and preincubation versions of the 6-well and 24-well plate assays were not possible due to data limitations. Only 22 of the 1503 assessments based on the 6-well plate assay employed the preincubation method; the dataset does not contain any corresponding 6-well and standard assay preincubation data. Similarly, only 90 of the 326 assessments based on the 24-well plate assay employed the preincubation method. The dataset only includes 8 assessments of coded test chemicals that employed both the 24-well plate and the standard preincubation assays; all assessments yielded negative responses on both assays.

132. Similarly, the miniaturised agar plating tests (i.e., without preincubation) were compared to the standard plate incorporation assay (i.e., without preincubation). Additionally, the Ames II and Ames MPF™ microfluctuation assays were compared to the standard plate incorporation test (i.e. without preincubation). The results of these comparisons are shown below in **Table 29**-**Table 31**. Again, for interpretation, these results can be compared to those presented in **Table 17**-**Table 19**. Again, assay correspondence mismatches associated with these analyses are not shown. Since the Ames II and Ames MPF™ assays involve a liquid suspension incubation, they might be expected to be more sensitive than agar plate testing methods (Gee et al., 1998; Kamber et al., 2009; Flückiger-Isler & Kamber, 2012). Thus, it might not be surprising that the results obtained show an increase in the false positive rate for the Ames MPF™ assay. However, there was no similar increase in the false positive rate for the Ames II assay. In contrast, the results also show a marked increase in the frequency of false negatives for both the Ames MPF™ and Ames II assays. This increase is reflected by the concomitant marked decrease in sensitivity. It might also be noted that the false negative rate for the 24-well plate assay almost doubled relative to that obtained when the assessment was not restricted to the standard plate incorporation assay format (i.e., without preincubation). Importantly, the results show a marked decline in PPV values for the Ames II and Ames MPF™ assays, reflecting low probabilities that these miniaturised assays will generate a positive call

that correctly corresponds to a positive call on the standard assay. This result is supported by the related increases in false negatives. The results also reveal declines in the NPV for the Ames MPF™ and 24-well plate assays, reflecting declines in the probability that a negative response correctly corresponds to a negative call on the standard plate incorporation assay (i.e., without preincubation). This result is supported by the related increases in false positives. Interestingly, the reduction in the PABAK agreement statistic for both the 24-well plate assay (without preincubation) and the Ames MPF™ assay (i.e., PABAK<0.8) reflects marked reductions in the agreement between these assays and the standard plate incorporation assay.

Table 26. 2 x 2 contingency tables showing call frequency correspondence values for each the Ames MPF™ and Ames II assays versus the preincubation version of the standard assay. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		Ames MPF™ assay			Ames II assay[1]		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	424	7	431	159	10	169
	Positive	20	101	121	3	17	20
	Column Totals	444	108	552	162	27	189

[1]TA98 only

Table 27. Performance of the Ames MPF™ and Ames II assays based on comparisons of individual assessments with the preincubation version of the standard assay. Equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[1]	PABAK
MPF assay	83.5 (75.6-89.6)	98.4 (96.7-99.3)	93.5 (87.1-97.4)	95.5 (93.1-97.2)	1.6 (0.7-3.3)	16.5 (10.4-24.4)	0.85 (0.77-0.93)	0.90 (0.86-0.94)
Ames II assay[2]	85.0 (62.1-96.8)	94.1 (89.4-97.1)	63.0 (42.4-80.6)	98.1 (94.7-99.6)	5.9 (2.9-10.6)	15.0 (3.2-37.9)	0.68 (0.54-0.83)	0.86 (0.77-0.93)

[1]$p<0.0001$
[2]TA98 only

Table 28. Number of assessments, number of substances, PI and BI for correspondence analyses between the Ames MPF™ and Ames II assays and the preincubation version of the standard bacterial reverse gene mutation test based on individual assessments. Equivocal responses not included in the analyses.

Mini Assay	Number of Substances	Number of Assessments	PI (Prevalence Index)	BI (Bias Index)
MPF assay	97	552	0.58	-0.024

Ames II assay[1]	85	189	0.75	0.037

[1]TA98 only

Table 29. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the plate incorporation version of the standard assay (i.e., without preincubation). The results for the 6-well and 24-well plate assays are also plate incorporation versions of the assay (i.e., without preincubation). Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		6-well plate incorporation			Ames MPF™ assay			Ames II assay[1]			24-well plate incorporation		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	704	14	718	116	12	128	48	3	51	57	5	62
	Positive	9	111	120	23	37	60	2	4	6	9	53	62
	Column Totals	713	125	838	139	49	188	50	7	57	66	58	124

[1]TA98 only

Table 30. Performance of the miniaturised assays based on comparisons of individual assessments with the plate incorporation version of the standard assay (i.e., without preincubation). Equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	92.5 (86.2-96.5)	98.1 (96.8-98.9)	88.8 (81.9-93.7)	98.7 (97.6-99.4)	1.9 (1.1-3.2)	7.5 (3.5-13.8)	0.89 (0.82-0.96)	0.94 (0.92-0.96)
Ames MPF™ assay	61.7 (48.2-73.9)	90.6 (84.2-95.1)	75.5 (61.1-86.7)	83.5 (76.2-89.2)	9.4 (4.9-15.8)	38.3 (26.1-51.8)	0.55 (0.41-0.69)	0.63 (0.50-0.73)
Ames II assay[4]	66.7 (22.3-95.7)	94.1 (83.8-98.8)	57.1 (18.4-90.1)	96.0 (86.3-99.5)	5.9 (1.2-16.2)	33.3 (4.3-77.7)	0.57 (0.31-0.82)	0.82 (0.61-0.94)
24-well plate assay	85.5 (74.2-93.1)	91.9 (82.2-97.3)	91.4 (81.0-97.1)	86.4 (75.7-93.6)	8.1 (2.7-17.8)	14.5 (6.9-25.8)	0.77 (0.60-0.95)	0.77 (0.64-0.87)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]p<0.0001
[4]TA98 Only

Table 31. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the plate incorporation version of the standard assay (i.e., without preincubation) based on individual assessments. Equivocal responses not included in the analyses.

Mini Assay	Number of Test Chemicals	Number of Assessments	Prevalence Index	Bias Index
6-well plate incorp.	109	838	0.71	0.006
Ames MPF™ assay	41	188	0.42	-0.059
Ames II assay[1]	31	57	0.77	0.018
24-well plate incorp.	34	124	0.03	-0.032

[1]TA98 only

133. The final analyses involved strain-specific comparisons of the miniaturised assays with the standard bacterial reverse gene mutation assay. The analyses, which are based on **individual assessments**, include performance assessments based on response data for (i) *S. typhimurium* TA98 only, (ii) *S. typhimurium* TA100 only, (iii) *S. typhimurium* TA1535 only, (iv) *S. typhimurium* TA97 or 97a only, and (v) *E. coli* WP2 or *S. typhimurium* TA102 only. No differentiation was made between the different genotypes of *E. coli* WP2. The results obtained are provided below in **Table 32**-**Table 46**. Again, for comparative interpretation, these results can be compared to those presented in **Table 17**-**Table 19**. Again, assay correspondence mismatches associated with these analyses are not shown.

134. Similar to what was observed for the correspondence analyses described earlier (e.g., **Table 19**), the overall strain-specific evaluations of the Ames MPF^{TM}, Ames II and 6-well plate assays were associated with elevated absolute values of the PI and low values for BI. Agreement was almost perfect for each assay, with PABAK values varying little from strain to strain, or compared with the values associated with the previously described assessments for all strains combined. For the 6-well plate assay there were increases in the false negative rates for strains *S. typhimurium* TA1537 and TA1535, with corresponding decreases in *S. typhimurium* TA98, TA100 and the TA102-*E. coli* WP2 combination. There was also an increase in false positive calls for the *S. typhimurium* TA102-*E. coli* WP2 combination. The most prominent change for the 24-well plate assay was, as in the 6-well plate assay, an increase in false positives for the *S. typhimurium* TA102-*E. coli* WP2 combination. However, all of those changes were small compared to the width of the confidence intervals. There was no discernable pattern in the individual strain analyses for the Ames MPF^{TM} assay.

135. With respect to the *S. typhimurium* TA98 correspondence analyses, the results do not show any appreciable differences in performance relative to what was observed for the analyses based on all strains (see **Table 17**-**Table 19**).

136. With respect to the *S. typhimurium* TA100 correspondence analyses, which only examined the 6-well plate, Ames MPF^{TM} and 24-well plate assays, the results show a marked decline in false negative rate for the 6-well plate assay, and a concomitant increase in sensitivity. The opposite was observed for the Ames MPF^{TM} assay, i.e., a marked increase in false negatives, and concomitant decrease in sensitivity. The PPV and NPV metrics show no appreciable decline in the predictive capacity of the miniaturised assays examined.

137. With respect to the *S. typhimurium* TA1535 correspondence analyses, which only examined the 6-well plate, Ames MPF^{TM} and 24-well plate assays, the results show marginal increases in false negative and false positive frequencies for the 24-well plate assay, and concomitant decreases in sensitivity and specificity. The results also show a small increase in false negatives for the Ames MPF^{TM} assay. Interestingly, the false positive rate for the 6-well plate assay was zero, with concomitant specificity of 100%. Although the PPV and NPV metrics are similar to those presented in **Table 17**, the results show marginal declines in PPV for the Ames MPF^{TM} and 24-well plate assays. This indicates a marginal decline in the ability to correctly predict a standard assay positive call; it is commonly associated with an increase in the frequency of false negatives.

138. With respect to the *S. typhimurium* TA1537-TA97 correspondence analyses, which only examined the 6-well plate, Ames MPF™ and 24-well plate assays, the results show an increase in the false negative rate for the 6-well plate assay, and a concomitant decrease in sensitivity. Conversely, the performance of both the 24-well and Ames MPF™ assays was improved with respect to both sensitivity and specificity, i.e., reductions in the frequencies of false positives and false negatives. The marked reduction in the frequency of false positives for the Ames MPF™ assay is particularly noteworthy; it is associated with a specificity of 99.1%. Importantly, the PPV and NPV metrics show no appreciable decline in the predictive capacity of the miniaturised assays examined. In fact, the PPV for the Ames MPF™ assay was markedly increased; this is expectedly associated with a decline in the frequency of false negatives.

139. With respect to the *E. coli* WP2- *S. typhimurium* TA102 correspondence analyses, which only examined the 6-well plate, Ames MPF™ and 24-well plate assays, the results show a decrease in the false negative rate for the 6-well plate assay, and a concomitant increase in sensitivity. Conversely, the performance of the 24-well plate assay was reduced with respect to both sensitivity and specificity, i.e., increases in the frequencies of false positives and false negatives. Similar to what was observed for *S. typhimurium* TA1537-TA97, the marked reduction in the frequency of false positives for the Ames MPF™ assay is particularly noteworthy. Indeed, it is associated with a specificity of 100%, i.e., no false positives. Importantly, the results show a marked decrease in PPV for 6- and 24-well plate assays, and a decline in NPV for the Ames MPF™ assay.

Table 32. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA98 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		6-well plate assay			Ames MPF™ assay			Ames II assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	191	3	194	134	6	140	141	8	149	10	0	10
	Positive	1	24	25	11	28	39	5	19	24	2	16	18
	Column Totals	192	27	219	145	34	179	146	27	173	12	16	28

Table 33. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA98 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	96.0 (79.6-99.9)	98.4 (95.5-99.7)	88.9 (70.8-97.6)	99.5 (97.1-100)	1.6 (0.3-4.5)	4.0 (0.1-20.4)	0.91 (0.78-1.04)	0.96 (0.91-0.99)
MPF assay	71.8 (55.1-85.0)	95.7 (90.9-98.4)	82.4 (65.5-93.2)	92.4 (86.8-96.2)	4.3 (1.6-9.1)	28.2 (15.0-44.9)	0.71 (0.56-0.85)	0.81 (0.71-0.89)
Ames II assay	79.2 (57.8-92.9)	94.6 (89.7-97.7)	70.4 (49.8-86.2)	96.6 (92.2-98.9)	5.4 (2.3-10.3)	20.8 (7.1-42.2)	0.70 (0.55-0.85)	0.85 (0.75-0.92)
24-well plate assay	Insufficient data							

[1]Positive Predictive value
[2]Negative Predictive Value
[3]p<0.0001

Table 34. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA98 only; equivocal responses not included in the analyses.

Mini Assay	Number of Test Chemicals	Number of Assessments	PI	BI
6-well plate assay	117	219	0.76	0.009
Ames MPF™ assay	107	179	0.59	-0.028
Ames II assay	97	173	0.71	0.017
24-well plate assay	19	28	Insufficient data	

Table 35. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA100 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		6-well plate assay			Ames MPF™ assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	178	3	181	125	3	128	15	1	16
	Positive	1	34	35	12	25	37	1	19	20
	Column Totals	179	37	216	137	28	165	16	20	36

Table 36. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA100 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	97.1 (85.1-99.9)	98.3 (95.2-99.7)	91.9 (78.1-98.3)	99.4 (96.9-100)	1.7 (0.3-4.8)	2.9 (0.1-14.9)	0.93 (0.80-1.07)	0.96 (0.91-0.99)
Ames MPF™ assay	67.6 (50.2-82.0)	97.7 (93.3-99.5)	89.3 (71.8-97.7)	91.2 (85.2-95.4)	2.3 (0.5-6.7)	32.4 (18.0-49.8)	0.71 (0.56-0.86)	0.82 (0.71-0.90)

24-well plate assay	95.0 (75.1-99.9)	93.8 (69.8-99.8)	95.0 (75.1-99.9)	93.8 (69.8-99.8)	6.2 (0.2-30.2)	5.0 (0.1-24.9)	0.89 (0.56-1.21)	0.89 (0.63-0.99)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$

Table 37. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments . Strain S. typhimurium TA100 only; Equivocal responses not included in the analyses.

Mini Assay	Number of Test Chemicals	Number of Assessments	PI	BI
6-well plate assay	113	216	0.67	0.009
Ames MPF™ assay	100	165	0.61	-0.055
24-well plate assay	28	36	-0.11	0.000

Table 38. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA1535 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

Standard Assay		6-well plate assay			Ames MPF™ assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
	Negative	125	0	125	131	3	134	10	1	11
	Positive	5	29	34	6	14	20	1	8	9
	Column Totals	130	29	159	137	17	154	11	9	20

Table 39. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA1535 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	85.3 (68.9-95.0)	100.0 (97.1-100)	100.0 (0.88-100)	96.2 (91.3-98.7)	0.0 (0.0-2.9)	14.7 (5.0-31.1)	0.90 (0.75-1.06)	0.94 (0.86-0.98)
Ames MPF™ assay	70.0 (45.7-88.1)	97.8 (93.6-99.5)	82.4 (56.6-96.2)	95.6 (90.7-98.4)	2.2 (0.5-6.4)	30.0 (11.9-54.3)	0.72 (0.57-0.88)	0.88 (0.78-0.95)
24-well plate assay	88.9 (51.8-99.7)	90.9 (58.7-99.8)	88.9 (51.8-99.7)	90.9 (58.7-99.8)	9.1 (0.2-41.3)	11.1 (0.3-48.2)	0.80 (0.35-1.23)	0.80 (0.37-0.98)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$

Table 40. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA1535 only; equivocal responses not included in the analyses.

Mini Assay	Number of Test	Number of Assessments	PI (Prevalence Index)	BI (Bias Index)

	Chemicals			
6-well plate assay	86	159	0.60	-0.031
Ames MPF™ assay	89	154	0.76	-0.019
24-well plate assay	13	20	-0.10	0.000

Table 41. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA97 or TA1537 only. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		6-well plate assay			Ames MPF™ assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	152	2	154	115	1	116	21	1	22
	Positive	4	31	35	9	28	37	1	21	22
	Column Totals	156	33	189	124	29	153	22	22	44

Table 42. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. Strain S. typhimurium TA97 or TA1537 only; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	88.6 (73.3-96.8)	98.7 (95.4-99.8)	93.9 (79.8-99.3)	97.4 (93.6-99.3)	1.3 (0.2-4.6)	11.4 (3.2-26.7)	0.89 (0.75-1.03)	0.94 (0.86-0.98)
Ames MPF™ assay	75.7 (58.8-88.2)	99.1 (95.3-100)	96.6 (82.2-99.9)	92.7 (86.7-96.6)	0.9 (0.0-4.7)	24.3 (11.8-41.2)	0.81 (0.65-0.96)	0.87 (0.77-0.94)
24-well plate assay	95.4 (77.2-99.9)	95.4 (77.2-99.9)	95.4 (77.2-99.9)	95.4 (77.2-99.9)	4.5 (0.1-22.8)	4.5 (0.1-22.8)	0.91 (0.61-1.20)	0.91 (0.69-0.99)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$

Table 43. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. Strain S. typhimurium TA97 or TA1537 only; equivocal responses not included in the analyses.

Mini Assay	Number of Test Chemicals	Number of Assessments	PI	BI
6-well plate assay	85	189	0.64	-0.011
Ames MPF™ assay	63	153	0.57	-0.052
24-well plate assay	15	44	0.00	0.000

Table 44. 2 x 2 contingency tables showing call frequency correspondence values for miniaturised assay results versus results for the standard bacterial reverse gene mutation assay; S. typhimurium TA102 - E. coli WP2 combination. Frequency values are based on individual assessments. Equivocal responses not included in the analyses.

		6-well plate assay			Ames MPF™ assay			24-well plate assay		
		Negative	Positive	Row Totals	Negative	Positive	Row Totals	Negative	Positive	Row Totals
Standard Assay	Negative	157	6	163	50	0	50	27	4	31
	Positive	1	28	29	6	19	25	2	17	19
	Column Totals	158	34	192	56	19	75	29	21	50

Table 45. Performance of the miniaturised assays based on comparisons of individual assessments with the standard bacterial reverse gene mutation assay. S. typhimurium TA102 - E. coli WP2 combination; equivocal responses not included in the analyses. Values in parentheses are 95% confidence limits.

Mini Assay	Sensitivity (%)	Specificity (%)	PPV (%)[1]	NPV (%)[2]	False + (%)	False – (%)	Kappa[3]	PABAK
6-well plate assay	96.6 (82.2-99.9)	96.3 (92.2-98.6)	82.4 (65.5-93.2)	99.4 (96.5-100)	3.7 (1.4-7.8)	3.4 (0.1-17.8)	0.87 (0.73-1.01)	0.93 (0.85-0.97)
Ames MPF™ assay	76.0 (54.9-90.6)	100.0 (92.9-100)	100.0 (82.4-100)	89.3 (78.1-96.0)	0.0 (0.0-7.1)	24.0 (9.4-45.1)	0.81 (0.59-1.03)	0.84 (0.67-0.94)
24-well plate assay	89.5 (66.9-98.7)	87.1 (70.2-96.4)	81.0 (58.1-94.6)	93.1 (77.2-99.2)	12.9 (3.6-29.8)	10.5 (1.3-33.1)	0.75 (0.47-1.02)	0.76 (0.51-0.91)

[1]Positive Predictive value
[2]Negative Predictive Value
[3]$p<0.0001$

Table 46. Number of assessments, number of test chemicals, Prevalence Index (PI) and Bias Index (BI) for correspondence analyses between the miniaturised assays and the standard bacterial reverse gene mutation assay based on comparisons of individual assessments. S. typhimurium TA102 - E. coli WP2 combination; equivocal responses not included in the analyses.

Mini Assay	Number of Test Chemicals	Number of Assessments	PI	BI

6-well plate assay	85	192	0.67	0.026
Ames MPF™ assay	33	75	0.41	-0.080
24-well plate assay	16	50	0.20	0.040

3.4.3. Summary of Assay Correspondence Analyses

- Overall, the agreement statistic values (e.g., PABAK – see 3.4.1) reflect excellent performance of all the miniaturised assays examined. More specifically, there was a high correspondence between the miniaturised assay calls and the standard assay calls; indeed, the agreement statistic values are predominately above 0.81. Agreement statistic values (i.e., Kappa and PABAK) above 0.81 are indicative of "almost perfect" agreement (**Table 16**). The few exceptions remained in the "substantial" agreement range (i.e., 0.61-0.80). The performance metrics (e.g. false negatives, false positives, sensitivity, specificity, positive predictive value, negative predictive value) generally reflect good correspondence between the miniaturised assays and the standard assay. **Table 62** contains an overall summary of performance metrics and agreements statistics for the analyses conducted herein.
- For **all analyses** investigating performance of the 6-well, Ames MPF™ and Ames II assays, the high absolute value of the PI reflects a high level of response asymmetry, i.e., high relative prevalence of negatives. As noted, high absolute value of PI can deflate the Kappa agreement statistic, making it necessary to use the PABAK value to assess the strength of the agreement.
- The uniformly low absolute value of the BI indicates no appreciable data bias, i.e., differences in the relative proportions of positive and negative calls for the 2 assay formats being compared (i.e., miniaturised assay vs. standard bacterial reverse gene mutation test).
- Results based on comparisons of **individual assessments** (i.e., each combination of test chemical, strain and S9-mix) revealed excellent performance of the two miniaturised agar plating assays, i.e., the 6- and 24-well plate assays. Moreover, based on the available data and considering the caveats mentioned in the following paragraphs, performance of these assays is superior to that observed for the two microfluctuation assays (i.e., Ames MPF™ and Ames II). More specifically, analysis of cross-tabulations across 121 test chemicals and 913 assessments reveals both high sensitivity and specificity for the 6-well plate assay, and concomitantly low false positive and false negative rates of 1.6 and 7.3%, respectively. With respect to the 24-well plate assay, the results reveal lower specificity compared to the 6-well plate assay, and a concomitantly higher false positive rate. These latter results are based on a different and much smaller number of test chemicals and assessments, i.e., 42 and 152, respectively. The PPV and NPV values also indicate excellent ability of the two miniaturised plate incorporation assays (i.e., 6- and 24-well) to correctly predict a standard assay call, i.e., PPVs between 92 and 94%, and NPVs between 92 and 99%. For both assays, the 95% confidence intervals for each of these performance measures (i.e., PPV and NPV) were relatively narrow.
- With respect to the two microfluctuation assays, overall performance based on comparison of **individual assessments** is lower relative to that of the agar-based tests. Although the specificity of the assays, and the corresponding false positive rates, are similar to those observed for the miniaturised agar plating tests, the observed sensitivity is decidedly lower (i.e., below 80%). The corresponding false negative rates

are consequently high relative to the miniaturised agar plating assays. According to the results of the analyses presented herein, which are based on the collected data, the Ames MPF™ and Ames II were unable to correctly classify 26.7% and 20.8% of the assessments that elicit a positive response on the standard bacterial reverse gene mutation test, respectively. Furthermore, with respect to the PPV value, the results confirm a lower ability to correctly predict a standard assay positive. The result is particularly noteworthy for the Ames II assay (i.e., PPV = 70%). Importantly, small samples sizes contributed to wide confidence limits of some performance statistics, e.g., false negative rate for Ames II. As noted below (Paragraph 152), this indicates that the results of some retrospective performance analyses must be interpreted with caution.

- Unfortunately, data limitations restricted the ability to conduct correspondence analyses based on **overall calls**. The 6-well plate assay was the only data set for which the 5-strain overall call analysis could be conducted; analysis of the 5-strain overall call dataset revealed marked improvement in the performance of the assay compared to the analysis based on individual assessments. The 2-strain overall call analyses showed improvements in the performance of the 6-well, Ames MPF™ and Ames II assays compared to analyses based on **individual assessments**. Importantly, the results obtained showed perfect performance of the 6-well plate assay, with relatively modest confidence intervals for the performance metrics and agreement statistics. Although the 2-strain overall call analyses show improvements in the performance of the Ames MPF™ and Ames II assays, the wide confidence limits of the performance metrics and agreement statistics are noteworthy. These wide confidence limits are consistent with aforementioned data analyses limitations, which are addressed in more detail below.
- Restricting the **individual assessment** analyses to the preincubation version of the standard assay, or alternatively, the standard plate incorporation assay without preincubation, resulted in some unexpected changes in performance. More specifically, since the Ames II and Ames MPF™ assay are microfluctuation assays, the expectation was a decline in false positives when comparing to only the preincubation version of the standard plate incorporation assay. However, aside from a small decline in false positives for the Ames MPF™ assay, this was not seen. In contrast, the results showed an unexpected decline in the frequency of false negatives.
- Restricting the **individual assessment** analyses to the standard plate incorporation assay without preincubation resulted in unexpected declines in the performance of the Ames MPF™ and Ames II assay, particularly a decline in sensitivity, and concomitant increase in false negatives. Similarly, restriction of the analyses to the standard assay without preincubation was associated with an unexpected decline in the performance of the 24-well plate assay, i.e., reduced sensitivity, and increased false negative rate.
- The **strain-specific analyses** based on **individual assessments** revealed few marked changes in the performance of the miniaturised assays. Nevertheless, the analyses did reveal some differences between performance analyses based on **individual assessments** and the strain-specific performance analyses. Exceptions include some declines in the performance of the assays examined when analyses were based only on strains *S. typhimurium* TA100, TA1535, TA1537-TA97 and *E. coli* WP2-

S. typhimurium TA102 (i.e., Ames MPF^{TM}, 24-well and 6-well), and some improvements in performance of the 6-well plate assay when analyses were based only on strains *S. typhimurium* TA98, TA100 and *E. coli* WP2-TA102.

- For **all analyses** conducted, alternately designating equivocals as positive or negative only resulted in marginal changes in performance of the miniaturised assays investigated. Nevertheless, with respect to the correspondence based on comparisons of **individual assessments**, there were some declines in the performance of the 24-well, Ames MPF^{TM} and Ames II assays (data not shown).

Data Analyses Limitations

140. It is important to highlight limitations of the retrospective performance evaluations described herein. First and foremost, it is important to emphasize that the evaluations of each miniaturised assay were necessarily based on different sets of test chemicals. More specifically, the indicated differences in the numbers of test chemicals and assessments used to evaluate each of the miniaturised assays reflects the fact that the performance analyses were based on different chemicals. Consequently, performance metrics such as sensitivity, specificity, PPV and NPV, and their associated confidence limits, must be interpreted with caution.

141. In the strictest sense, the performance metrics for each miniaturised assay, and the attendant predictive value of that assay, cannot be credibly compared with those associated with another miniaturised assay (Altman and Bland, 1994). Indeed, rigorous cross-assay performance evaluations would need to be based on results for a common set of test chemicals. Although it may be possible to restrict the retrospective analyses conducted herein to a common set of test chemicals, those analyses were considered outside the scope of the current DRP. Importantly, even if the retrospective analyses were to be restricted to a common set of test chemicals, dataset bias relative to the population of available chemicals (e.g., a preponderance of mutagens) could still have a profound effect on the ability to use the results to predict future performance of the miniaturised assays examined. As noted by Altman and Bland (1994), the calculated predictive values of an alternative test (i.e., PPV and NPV) depend on the prevalence of positive and negative responses in the test that is denoted the point of reference. For example, with respect to these analyses, if the dataset examined has a low frequency of positive responses on the standard assay, the results obtained will provide better relative assurance about negative predictions. Indeed, in the strictest sense, ideal evaluation of the performance of any alternative assay (i.e., *true* miniaturised assay performance) requires a dataset that adequately represents the population of available chemicals and its associated balance of positive and negative responses; moreover, the applicability domain of the endpoint under consideration.

142. As noted earlier, instances where responses to a given test chemical were *mixed* (i.e., different results across multiple assessments), resulted in exclusion from the current analyses. This approach was justified by an inability to objectively assign a single consensus call when the calls in the collected data were not consistent across assessments (i.e., intra- and/or inter-laboratory variability). However, this exclusion can differentially restrict the ability to examine correspondence between miniaturised assay calls and the standard assay calls; moreover, differentially bias the correspondence analyses and associated results. Consequently, going

forward, efforts should be made to incorporate these *mixed* calls into the correspondence analyses. More specifically, mixed calls could be more rigorously scrutinised to determine if a consensus might be possible. For example, instances with multiple positives and a single equivocal might be designated as positive. Similarly, instances with multiple negatives and a single equivocal might be designated as negative. Such a change in call designation would require careful review of the original data submissions to scrutinize, for example, call criteria, tested concentration ranges, etc. Additionally, although more challenging, mixed calls might be resolved in instances where an inconsistent pattern of responses might not be surprising. For example, responses without S9 for test chemicals known to require metabolic activation (e.g., BaP). Similarly, for instances where a test chemical is known to yield variable responses for a particular strain (e.g., 4NQO and *S. typhimurium* TA1535), test chemicals are known to be difficult to detect due to requirement(s) for specialised metabolic activation (e.g., 1NA), or test chemicals known to be non-mutagenic on the standard assay that have been shown to yield erratic responses on some miniaturised assays (e.g., TBA). Alternatively, mixed calls might also be resolved in cases where the response expectation for a specific test chemical is well recognised. For example, OECD TG 471 recommends sodium azide as a positive control for *S. typhimurium* base-pair mutation strains TA100 and TA1535 without S9-mix (OECD, 1997). However, the collected data indicated mixed calls for *S. typhimurium* TA100 and TA1535 without S9-mix on the Ames MPF™ assay. More specifically, the submitted data included the following for *S. typhimurium* TA100 without S9-mix: three positive calls, one negative call and one equivocal call. Similarly, the submitted data included the following for *S. typhimurium* TA1535 without S9-mix: two positive calls and two negative calls. Consequently, neither a 5- overall call nor a 2-strain overall call were possible for sodium azide on the Ames MPF™ assay, despite the fact that the chemical is a well-known positive control. Implementation of appropriate data scrutiny steps could permit resolution of such dataset problems, thus maximising the availability of data for assay correspondence assessment. It is interesting to note that a weighted data analysis approach might be employed to handle mixed calls that are differentially biased in favour of a specific consensus call, e.g., higher weights for a positive call when the majority of the submitted assessment calls are positive. Such an approach might enable incorporation of mixed call occurrences into the retrospective performance analyses

143. Supplementary **Table 58**-**Table 61** document assay mismatches, i.e., miniaturised assay calls that differ from those recorded for the standard assay. Since the mismatch frequency directly impacts performance metrics (e.g., sensitivity, specificity, PPV and NPV), mismatch instances should also be carefully reviewed and scrutinised. In some instances, it may be possible to determine the likely determinant of the mismatch, with implementation of judicious remedial measures that improve the reliability and accuracy of the performance assessments. For example, inter-laboratory differences in call criteria may account for some mismatches, with the situation ameliorated by implementation of a uniform set of call criteria across the entire retrospective dataset. In addition, variability in test chemical and solutions purity might also contribute to assay mismatches, and it may be possible to determine if different laboratories used test chemicals with differing purity levels. Lastly, scrutiny of the mismatches could consider known problems related to the use of improperly stored and/or expired Ames MPF™ assay reagents. The test method developer has noted that such issues could have contributed to a high frequency of false negatives, and this possibility should be more thoroughly considered. More specifically, it may be necessary to remove some Ames MPF™ assay results if they are determined to be associated with an unreliable negative. Thus, similar to what was noted above for mixed calls, further dataset scrutiny might also permit resolution of dataset problems associated with unexpected mismatches.

144. Altman and Bland (1994) noted that evaluations of test performance require knowledge about the likelihood that a test outcome is correct, i.e., the PPV and NPV. Indeed, they state that the entire rationale underscoring the development of an alternative test is accurate "diagnosis"; thus, it is necessary to determine the likelihood that the alternative test will yield a correct "diagnosis". However, the authors also note that NPV and PPV values obtained in an individual study cannot be universally applied; particularly in instances where the distribution of responses for the point of reference in the analysed dataset is differentially skewed relative to the population of existing responses (i.e., responses for the population of tested chemicals). The analyses presented herein noted high PI values, and the noteworthy contingency table asymmetry, i.e., preponderance of negative calls on the standard assay. Consequently, with respect to these analyses, if the prevalence of a particular call on the standard assay is particularly high, then it is more likely that the miniaturised assay evaluated will generate a matching call. More specifically, the data analyses presented herein indicate that the retrospective performance assessments are based on a dataset with a high prevalence of standard assay negative calls. Thus, the NPV might be perceived as circumstantially elevated relative to true negative predictivity, i.e., the rarer the positive outcome on the standard assay, the better the confidence that a negative on a miniaturised assay is correct. Indeed, as noted by Altman and Bland (1994), a very high prevalence of negatives for the reference assay will necessarily result in an elevated frequency of miniaturised assay false positives. Additionally, the low prevalence of standard assay positive calls might be perceived as circumstantially contributing to a low PPV relative to true positive predictivity. That being said, it is important to emphasize that in most cases the PPV and NPV values described herein were both quite high, i.e., >85%. Thus, as noted earlier, reliable performance assessments, including assessments of positive and negative predictivity, ideally require response distributions for the point of reference assay that appropriately correspond to the entire population of responses.

145. It might also be noted that the utility of the retrospective data analyses is expectedly affected by sample size. Indeed, 2- and 5-strain overall call analysis sample sizes are small (<100) for all miniaturised assays, and some assay performance statistics consequently have wide confidence intervals. This means that care is needed in drawing conclusions from the retrospective performance analyses; indeed, the revealed patterns should not be over-interpreted. Nevertheless, the results obtained are useful for exploring the degree of correspondence between the standard assay and the miniaturised assays examined, with discordant results providing important insights. Relatedly, as noted earlier (i.e., Paragraph 114 of Section 3.4.1), and in accordance with Kirkland et al (2019), categorization of performance (**Table 16**) based on Kappa and PAPAK might employ the lower confidence limits of the performance statistic values. Determination of the sample sizes required to realistically evaluate the performance of alternative test procedures depends on the required statistical power, and the expected level of agreement. Since, as noted by Sim and Wright (2005), agreement between two alternative test methods will be better than zero by chance alone, specifying a null hypothesis that Kappa=0 has little value. Alternatively, specification of a minimum acceptable Kappa value is context dependent. Indeed, since the miniaturised assays evaluated herein are all versions of the same bacterial reverse gene mutation assay, the expected Kappa should be quite high, i.e., high minimum performance requirements. Importantly, prior to implementing a prospective evaluation regarding the performance of miniaturised versions of the bacterial mutation assays, a sample size calculation should be conducted such that the study "has a stated probability of detecting a statistically significant

Kappa coefficient" (Altman and Bland, 1994). As an example, if the proportion of standard assay positives is 50%, the expected Kappa is 0.8, the statistical power is 0.8, and the minimum acceptable level of agreement is 0.6, then 126 samples (i.e., test chemicals) will be required (Sim and Wright, 2005).

Follow-up Analyses

146. Although the retrospective performance analyses presented herein provide a great deal of information regarding the ability of the investigated miniaturised assays to reliably assess bacterial mutagenicity, follow-up analyses are certainly warranted. More specifically, since the analyses of each assay was based on a different set of test chemicals, and the standard assay dataset comprises a preponderance of standard assay negative calls, it will likely be necessary to conduct follow-up, prospective performance evaluations. Such evaluations would necessarily employ a common set of carefully-selected test chemicals, and proceed in a manner that is consistent with the provisions outlined in OECD Guidance Document (GD) 34. Ideally, to the extent possible, the selected set of test chemicals should appropriately represent the population of available chemicals; moreover, the attendant distribution of mutagenic potency.

147. Additional analyses of data collected during this retrospective validation study, including quantitative analyses of the concentration response data, will likely aid in the design of any prospective validation studies. These analyses are unlikely to change but may refine the overall conclusions. Additionally, follow-up analyses of the retrospective and/or prospective datasets will need to include quantitative concentration-response analyses. Such analyses will permit comparative evaluation regarding the ability of the miniaturised assays to effectively detect weak responders. More specifically, quantitative analyses could compare the minimum amount of test chemical required to elicit a significant response.

148. A substantial effort will be needed to develop methods quantitative examinations of concentration-response data. The challenges arise from previously described properties of Ames test data, applicable to both standard and miniaturised assays. These include distributions of count data which vary among strains in ways which make it difficult to use the same methods for all strains or to choose methods that can adequately correct for deviations from the distributions used to construct the statistical tests (see Paragraph 24). Variability in background counts among the strains, between laboratories, and on different days within the same laboratory (See Table 8 and Paragraph 83) will also complicate design of these methods. Moreover, the variability observed in solvent control count data increases. This, together with the increasing numbers of revertants induced by mutagens (Margolin et al., 1994; Kato et al., 2018), further complicate development of methods for determining dose response. Note that these methods may be necessary to make the mutagenic potency estimates needed to identify weak responders. Techniques not previously described in published analyses of Ames test data, such as Bayesian analyses or benchmark dose modelling may circumvent some of these challenges. The EG decided not to undertake these analyses due to the time which would be needed for method development and the likelihood that the analyses would not likely change the overall conclusions and recommendations of the DRP. Since it is likely that none of the miniaturised assays will be validated without at least

some prospective study data, it would be more appropriate to conduct further analyses of this retrospective dataset during the design of those prospective studies.

4 Discussion

4.1. Interpretation of results

149. The retrospective performance analyses presented herein were based on data submitted by over 20 laboratories comprising 8727 assessments; including 429 test chemicals, of which 188 were coded. Although the PCA analysis cannot explain the exact relationship between molecular similarities and toxicological effects, it confirmed that the test chemicals used for the analysis were not allocated to a specific coordinate of the chemical space of the bacterial reverse gene mutation test and represent a good portion of the space. For the retrospective analysis, the qualitative designations (i.e., mutagen or non-mutagen) based on miniaturised bacterial reverse gene mutation tests were compared to those associated with the preincubation and/or plate incorporation version of the standard assay (i.e., the point of reference). The term qualitative is used here to differentiate these analyses from any future analyses regarding the magnitude of the mutagenic responses, and/or the minimum amount of test article required to yield a response that would be designated as positive.

150. Although the **individual assessment** analyses provide useful insights into the performance of the miniaturised bacterial reverse gene mutation assays examined, results based on the 5-strain overall call comparisons might be deemed most relevant with respect to the use of the assays for mutagenicity screening. Each test chemical included in the 5-strain overall call comparison was necessarily analysed in a miniaturised assay with the complete set of bacterial strains described in OECD TG471 (OECD, 1997), both with and without S9-mix metabolic activation. Unfortunately, there were too few 5-strain overall data to conduct meaningful analyses for the 24-well plate, Ames MPF™ and Ames II assays (see paragraph 128). Additional analyses examined correspondences associated with 2-strain overall calls; there were sufficient data to examine the 6-well plate, Ames MPF™ and Ames II assays (see paragraph 132). Lastly, the **individual assessment** analyses were parsed into strain-specific analyses; moreover, analyses based only on standard assay pre-incubation results, or alternatively, only on standard assay plate-incorporation results (i.e., without preincubation).

151. The results of the correspondence analyses are described in detail in section 3.4.2, and summarised in section 3.4.3. This section provides a brief comparative overview of the results obtained; **Table 62** provides a comparative summary of the results for all correspondence analyses.

- **6-well plate assay:** As compared with the other miniaturised assays examined, the dataset included 6-well assay data from more laboratories and for more test chemicals. It was the only assay for which enough data were available to conduct a 5-strain overall call analysis. Overall agreement for the correspondence analysis, as measured by PABAK, was almost perfect for the 62 test chemicals for which data were available.

Performance measures such as sensitivity, specificity, PPV, and NPV were high, with concomitantly low values levels of false positives and false negatives. Many laboratories have historically used miniaturised assays for screening based only on *S. typhimurium* strains TA98 and TA100, which was reflected by the higher number of test chemicals (97) included in the 2-strain overall call analyses. Compared with the 5-strain overall call analysis, agreement and performance were as good or better; the narrower confidence intervals are not surprising since the 2-strain overall call dataset is larger. However, it should be noted that the additional test chemicals included in the 2-strain overall call analyses were not mutagenic in either the 6-well plate or standard assays. Thus, some performance measures (e.g., false negatives, sensitivity) were necessarily based on essentially the same overall call data. With respect to agreement and performance, the overall call analyses were similar to the individual assessment results; although, for the overall call assessments, there was a small but consistent improvement in each measure. Repeating the correspondence analysis using only standard assay plate incorporation results had no appreciable effect on any of the measures of agreement or performance. Strain specific analyses resulted in increased sensitivity, and fewer false negatives, for *S. typhimurium* strains TA98, TA100, and the combined *E. coli* WP2 and *S. typhimurium* TA102 strains, with a corresponding decrease in sensitivity for *S. typhimurium* TA1537 and TA1535. However, PABAK and the other performance measures did not vary among the strains except for an increase in false positive rate for the combined *E. coli* WP2 and *S. typhimurium* TA102 strains.

- **Ames MPF™ assay:** The second largest portion of the dataset is associated with the Ames MPF™ assay. For the 2-strain overall call analysis, most performance measures were similar to that obtained for the 2-strain overall call analysis with the 6-well plate assay, except for a notable decrease in sensitivity, and corresponding increase in false negatives. Although PABAK was in the “almost perfect” range, it was at the lower end of that range. Marginal overlap with the PABAK 95% confidence interval for the 6-well assay evaluation suggests a significantly lower level of agreement. For the **individual assessment** analysis, there was a similar pattern of agreement and performance, except for sensitivity and false negatives. When standard assay data were segregated into preincubation and plate incorporation results, there were small but consistent improvements in all measures of agreement and performance. The strain-specific analyses did not reveal any consistent patterns of differences relative to the individual assessment results.

- **Ames II assay:** Overall agreement for the Ames II assay was “almost perfect” for the overall call analyses. However, as for Ames MPF™ assay, the sensitivity was somewhat lower, and associated false negatives higher, relative to the miniaturised agar-plating assays. Patterns such as improvement in the overall call analysis relative to the **individual assessments** analysis, and better correspondence with standard assay preincubation data, were similar to that observed for the Ames MPF™ assay. However, this comparison should be interpreted with caution in light of the wider confidence intervals around the measures of agreement and performance; the wider intervals are consistent with the smaller number of test chemicals in the dataset. For

example, there were 32 test chemicals for the Ames II assay 2-strain overall call analysis, compared with 47 test chemicals for the Ames MPF^{TM} assay analyses.

- **24-well plate assay:** Overall call analyses could not be conducted for the 24-well plate assay, the miniaturised assay with the smallest dataset. In the analysis of **individual assessments**, overall agreement was at the lower end of the "almost perfect" range, and lower than agreement for the 6-well plate assay. Most measures suggested performance similar to that observed for the 6-well plate assay, with the exception that specificity and NPV were slightly lower, and false positives higher. There was a consistent decline in agreement and performance when the 24-well plate assay assessments were compared to standard assay plate incorporation data only. However, it should be noted that the differences were greatest for measures with the widest confidence intervals.

4.2. Limitations of the Retrospective Performance Analyses

152. The main limitations of the correspondence analyses can be summarised as follows:

- Each correspondence analysis is based on data for a **different set of test chemicals**. The test chemicals in the dataset are those that were chosen by each of the laboratories participating in the retrospective validation; thus, the agreement statistics and performance metrics necessarily represent the outcome for the set of test chemicals investigated. As noted earlier (see 3.4.3), this is a significant source of bias, with repercussions for comparative interpretation of the results across the miniaturised assays investigated. Cross-assay comparisons of the test chemical set used for the correspondence analyses are complicated by inclusion of a substantial number of coded compounds (i.e., unidentified proprietary compounds); comparisons are also complicated by exclusion of mixed calls and lack of assessment of mutagenic potency. The latter are required for identification of weak responders.

- For most of the analyses, the datasets included a **prevalence of standard assay non-mutagens**, with concomitant impacts on performance metrics such as PPV and NPV. As noted, an asymmetric standard assay dataset, which biasedly includes a predominance of non-mutagens, will result in inflated NPV values, and associated increases in false positives. Conversely, the PPV will be concurrently low, with an associated decrease in false negatives.

- In several instances, the submitted data for some miniaturised assays included a **low incidence of revertants in the solvent controls**. Indeed, for some experiments there were no revertant colonies in any of the solvent control wells. Examples of such low solvent control counts were noted for *S. typhimurium* strains TA1535, TA1537 and TA98 in the Ames MPF^{TM} assay and the 24-well plate assay. Solvent control counts this low may compromise the ability of the miniaturised assay to detect small increases in response. The variations in numbers of replicates used by some laboratories may also alter the ability to detect weak responders.

- As noted, ***mixed* call information was not incorporated into the analyses**. Some of these mixed calls may inevitably be related to poor reproducibility across replicate experiments (i.e., intra- and inter-laboratory reproducibility). If reproducibility differs between the standard assay and a miniaturised assay, the absence of mixed call data may result in removal of important data regarding assay performance. Moreover, differential removal of mixed calls across the miniaturised assays could result in differential suppression of the ability to examine correspondences with the standard assay. As noted, mixed calls might be expected based on the nature of the test chemical, and the mechanisms underlying induction of a positive response. Indeed, response heterogeneity justifies the necessity of using 5 strains to appropriately detect a broad range of mutagenic substances.

- **Variations in the criteria used to interpret test results** may impact inter-laboratory reproducibility, and, by extension, the correspondence between miniaturised assay calls and standard assay calls. Using lab-specific criteria (**Table 9**), data submitters were asked to provide a single assessment call (i.e., negative, positive, equivocal) for each experiment. For instances where no call was provided, the EG inserted a call using the criteria outlined in Section 3.3.1.

- **Variations in protocols** employed by the laboratories that submitted data may impact inter-laboratory reproducibility for the 6- and 24-well plate assays. Examples include varying the number of replicates used for solvent controls and/or test chemical in both assays, and use of pre-mixes for the 24-well plate assay. None of the laboratories used a pre-incubation protocol for the 6-well plate assay, and very little of the 24-well plate assay data were generated using a preincubation protocol. OECD TG471 suggests that certain chemical classes require use of pre-incubation protocols (OECD, 1997); the ability of the miniaturised assays to detect such chemicals remains to be examined.

- Although the protocols for each of the miniaturised assays necessarily require addition of a smaller amount of test chemical relative to that needed for the standard bacterial reverse gene mutation test, there is as yet **no quantitative** analysis comparing the limits of detection.

- Moreover, it should be noted that evaluating the performance of miniaturised bacterial reverse gene mutation tests is challenged by the fact that comparison is done against the standard bacterial reverse gene mutation test, by far the most used genotoxicity test ever. It is present in the first tier of every strategy for genotoxicity known and thousands of compounds have been studied in the test. Given the limitations listed above and, the outcome of the current retrospective validation study regulators may accept a positive result in a miniaturised bacterial reverse gene mutation test as an indication for a genotoxic test chemical or, at best, to consider it as a genotoxic test chemical, but, more importantly, they will not consider a negative result in a miniaturised test as sufficient proof that a test chemical is non-genotoxic. As long as there is no conclusive evidence that the miniaturised bacterial reverse gene mutation test gives for most test chemicals identical results as the standard bacterial reverse gene mutation test, it will remain

difficult to convince regulators to accept a miniaturised bacterial reverse gene mutation test as a reliable alternative for the standard assay. When setting up additional work (see 4.3), it is therefore important to clearly define which knowledge gaps need to be addressed for each of the miniaturised bacterial reverse gene mutation tests (e.g. quantitative analysis, more detailed evaluation of the applicability domains,…).

4.3. Next steps

153. The correspondence analyses described in this DRP have contributed to significant advancement in our understanding of the relative performance of the 4 miniaturised bacterial reverse gene mutation assays evaluated. However, there was a consensus within the workgroup that more work would need to be done to create a fully validated test guideline for any one of the miniaturised assays examined; alternatively, to explicitly include any of the miniaturised assays in OECD TG471.

154. Data collected, but not yet analysed as part of the retrospective performance analyses, may provide a significant amount of the information needed for more complete validation. For example, the following prospective opportunities have been identified:

- Characterize the degree to which the test chemicals are representative of the domain of applicability for the standard assay (see **Table 47**). The EG proposed that a high priority would be ensuring inclusion of a diverse set of weak responders. Within this context, a more in depth PCA analysis, i.e., including functional group characterization, may provide insights in the mode of action of each single test chemical.
- For each miniaturised assay, assess and evaluate intra- and inter-laboratory reproducibility. These analyses would include assessments of the impact of reproducibility on mixed calls and call mismatches. A benchmark for reproducibility is the widely cited estimate of 85% (Piegorsch and Zeiger, 1991).
- Quantitative examination of the collected mutagenicity concentration-response data, while challenging, may provide valuable information, including identification of weak responders, and more complete assessments of correspondence and reproducibility. The results may also provide insights regarding the design of prospective validation studies. Examples of potentially useful quantitative comparisons between each miniaturised assay and the standard assay might include:
    - Assess and evaluate relative quantitative sensitivity (e.g., the lowest concentration required to elicit a significant positive response) in the overall chemical set; moreover, for mutagens in selected chemical categories (e.g., weak responders);
    - Determine whether *no revertants* observations for solvent controls adversely affects assay sensitivity, particularly for mutagens expected to induce a weak response;
    - Determine whether the call criteria commonly used to evaluate miniaturised assay responses are effectively aligned with the criteria employed for the standard assay.

- As indicated earlier, effective retrospective performance analysis of any miniaturised version of the bacterial reverse mutation test must be based on a set of chemicals that is appropriately representative of the endpoint's applicability domain. Moreover, a set of chemicals for which the response pattern is reasonably representative of the population of chemicals that have been subjected to bacterial mutagenicity assessment. Table 47 summarizes the characteristics of test chemicals that would appropriately represent the applicability domain of the bacterial reverse mutation test; these characteristics would need to be considered for any effective validation of miniaturised alternatives to the standard assay described in OECD TG 471. A more in depth PCA analysis could also be helpful for the selection of reference test chemicals.
- Comparison of concordance of the miniaturised assays with alternative formats requiring less test material such as a standard format test using fewer than 5 strains.

Table 47. Characteristics of test chemicals that adequately represent the applicability domain of the bacterial reverse gene mutation test

Characteristic	Required Range
Mutagenic potency	Include a diverse set of weak responders
Mutagenic mechanism for strain-specific detection	GC substitution, A:T substitution, base pair insertion/deletion
Solubility	Water soluble, moderately non-polar, and highly non polar
Metabolic requirements	Chemicals or chemical classes well-, moderately-, and poorly-metabolised by induced rat liver S9, e.g., Aroclor-induced or PB/NF[1]-induced. Chemicals or chemical classes that require a specialised type of S9 (i.e., specialised inducer), and/or an elevated level of S9. Chemicals or chemical classes that are preferentially detected using a preincubation protocol.

[1]Phenobarbitol/β-naphthoflavone

155. Going forward, any effective performance evaluation for a miniaturised version of the bacterial reverse mutation test would need to adhere to the modular validation stages outlined in OECD GD (Guidance Document) 34 (OECD, 2005). More specifically, as noted in GD34, the following validation modules are recommended:

(i) Test definition (including purpose, need and scientific basis);
(ii) intra-laboratory repeatability and reproducibility;
(iii) inter-laboratory transferability;
(iv) inter-laboratory reproducibility;
(v) predictive capacity (accuracy);
(vi) applicability domain; and,
(vii) performance standards.

156. The initial chapters of this DRP address the information requirements of module (i). The retrospective performance analysis was primarily designed to collect the data needed to address modules (v) and (vi). However, analysis of the collected data might also be used to address the requirements associated with some other modules (e.g., intra-laboratory repeatability and inter-laboratory reproducibility). Information gaps that remain after complete analysis of the collected data will inform the design of a prospective validation study. Issues that will need to be addressed by a prospective validation include:

1) Coverage of areas in the applicability domain not adequately covered in the retrospective analysis;

2) Ability of low background strains to detect small increases relative to the solvent controls, particularly for the microfluctuation assays;
3) Performance of the microfluctuation assays and the 24-well plate assay in the presence of unusually high concentrations of non-aqueous solvents (i.e., >10% v/v);
4) Optimal S9 concentration, particularly for detection of carcinogens known to be poorly metabolised by rat liver S9-mix;
5) Adequacy of miniaturised agar-plate-based assays conducted with preincubation;
6) Performance standards and demonstration of laboratory proficiency.

5 Recommendations

157. Several miniaturised versions of the bacterial reverse gene mutation assay have been developed and are already in use, particularly for early screening of new products. Despite the current use of the methods for specific reasons, there is currently no scientific consensus on the general regulatory use of these methods. However, under specific conditions (e.g. when limited test chemical is available such as for some impurities or metabolites), data from one or more of the miniaturised versions of the bacterial reverse gene mutation test may be among the alternatives to contribute to an overall safety assessment.

158. The retrospective analysis described in this DRP provides valuable insights into qualitative aspects of the performance of these miniaturised assays compared to the standard Ames assay. Conducting further confirmatory work in order to facilitate the regulatory acceptance of these methods in the future should be considered.

159. The concordance analysis shows a good agreement between the miniaturised assay calls and the standard assay calls. However, it presents several limitations and does not consider quantitative aspects, particularly those related to identification of mutagens which produce weakly positive responses in the test. For these reasons, none of the four evaluated miniaturised bacterial reverse gene mutation tests can at this stage be recommended as an alternative for the standard OECD TG471 bacterial gene mutation test (OECD, 1997), nor can any of the four miniaturised bacterial reverse gene mutation tests be explicitly included in the current OECD TG 471. Because there are areas of application for miniaturised bacterial reverse gene mutation assays and because the results from the retrospective study are promising, it is recommended that, if additional resources allow, further work be conducted in the future, including the quantitative exploitation of the data collected in the context of the retrospective analyses and, as appropriate, a prospective validation study.

160. Follow-up quantitative analyses is essential. Concentration-response data will provide insight regarding sensitivity and specificity, call criteria, and inter- and intra-laboratory variability for each of the assays examined and would provide the strong basis to determine the relevance to conduct prospective validation studies.

6 References

Agurell E, Stensman C. 1992. Salomenella mutagenicity of three complex mixtures assayed with the microsusoension technique. A WHO/IPCS/CSCM study. Mutat Res 276:87-91.

Alves DKM, Kummrow F, Cardoso AA, Morales DA, Umbuzeiro GA. 2016. Mutagenicity profile of atmospheric particulate matter in a small urban center subjected to airborne emission from vehicle traffic and sugar cane burning. Environ Mol Mutagen 57:41–50.

Altman DG, Bland JM. 1994. Diagnostic tests 2: predictive values. Brit Med J 309:102.

Ames, BN. 1971. The Detection of Chemical Mutagens with Enteric Bacteria. In A. Hollaender (Ed.), *Chemical Mutagens: Principles and Methods for Their Detection Volume 1* (pp. 267-282). Boston, MA: Springer USA.

Ames BN, Dursten, WE, Yamasaki E. and Lee, FD. 1973. Carcinogens are Mutagens: A Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection. Proc. Nat. Acad. Sci. USA 70:2281-2285.

Aubrecht J, Osowski JJ, Persaud P, Cheung Ackerman JR, Lopes SH, Ku WW. 2007. Bioluminescent Salmonella reverse mutation assay: A screen for detecting mutagenicity with high throughput attributes. Mutagenesis 22:335–342.

Ayrton AD, Neville S, Ioannides C. 1992. Cytosolic activation of 2-aminoanthracene: implications in its use as diagnostic mutagen in the Ames test. Mutat Res 265:1-8.

Braun K. 2001. Automation of the Ames II Assay: high through-put screening of mutagenic substances; Aventis Pharma Deutschland GmbH, DI & A, Lead Optimization, Drug Safety Evaluation. MipTec ICAR, Basel, Switzerland.

Bridges BA. 1972. Simple bacterial systems for detecting mutagenic agents. Lab Pract 21:413-416.

Brooks TM. 1995. The use of a streamlined bacterial mutagenicity assay, the MINISCREEN. Mutagenesis 10:447-448.

Burn CC, Weir AAS. 2011. Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. Vet J 188:166-170.

Burke DA, Wedd DJ, Burlinson B. 1996. Use of the Miniscreen assay to screen novel compounds for bacterial mutagenicity in the pharmaceutical industry. Mutagenesis 11:201–205.

Byrt T, Bishop J, Carlin JB. 1993. Bias, prevalence and Kappa. J Clin Epidemiol 46:423-449.

Cariello NF, Piegorsch WW. 1996. The Ames test: the two-fold rule revisited. Mutat Res 369:23-31.

Cebron MR, Dill N, Gabriel F, Kötter TR, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B. 2007. KNIME: The Konstanz Information Miner. In Studies in Classification, Data Analysis, and Knowledge Organization (GfKL), Berlin, Germany: Springer.

Cerná M, Pastorková A, Vrbíková V, Smíd J, Rössner P. 1999. Mutagenicity monitoring of airborne particulate matter (PM10) in the Czech Republic. Mutat Res Genet Toxicol Environ Mutagen 444:373–386.

Chauret N, Gauthier A, Nicoll-Griffith DA. 1998. Effect of Common Organic Solvents on in Vitro Cytochrome P450-Mediated Metabolic Activities in Human Liver Microsomes. Drug Metab Dispos 26:1-4.

Cicchetti DV, Feinstein AR. 1990. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 6:551-558.

Claxton LD, Allen J, Auletta A, Mortelmans K, Nestmann E, Zeiger E. 1987. Guide for the Salmonella typhimurium/mammalian microsome tests for bacterial mutagenicity. Mutat Res 189:83-91.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46.

Côté C, Blaise C, Delisle CE, Meighen EA, Hansen PD. 1995. A miniaturized Ames mutagenicity assay employing bioluminescent strains of Salmonella typhimurium. Mutat Res 345:137–146.

Crebellia R, Contia L, Monarca S, Feretti D, Zerbini I, Zani C, Veschetti E, Cutilli D, Ottaviani M. 2005. Genotoxicity of the disinfection by-products resulting from peracetic acid- or hypochlorite disinfected sewage wastewater. Water Res 39:1105–1113.

Cunningham M. 2009. More than Just the Kappa Coefficient: A Program to Fully Characterize Inter-Rater Reliability between Two Raters. SAS Global Forum 2009, paper 242-2009.

DeMarini DM. 2000. Influence of DNA repair on mutation spectra in Salmonella. Mutat Res 450:5-17.

DeMarini DM, Shelton ML, Abu-Shakra A, Szakmary A, Levine JG. 1998. Spectra of spontaneous frameshift mutations at the hisD3052 allele of Salmonella typhimurium in four DNA repair backgrounds. Genetics 149:17-36.

Diehl M, Willaby S, Snyder R. 2000. Comparison of the results of a modified miniscreen and the standard bacterial reverse mutation assays. Environ Mol Mutagen 36:72–77.

Di Giorgio C, Malleret L, Gueydon-Morin C, Rigaud S, De Meo M. 2011. Comparison of two extraction procedures for the assessment of sediment genotoxicity: Implication of polar organic compounds. Mutat Res 725:1–12.

Donner A, Eliasziw M. 1992. A goodness-of-fit procedure to inference procedures for the Kappa statistic: confidence interval construction, significance-testing and sample size estimation. Stat Med 11:1511-1519.

Egorova OV, Ilyushina NA, Rakitskii VN. 2020. Mutagenicity evaluation of pesticide analogs using standard and 6-well miniaturized bacterial reverse mutation tests. Toxicol In Vitro 69:105006.

Escobar PA, Kemper RA, Tarca J, Nicolette J, Kenyon M, Glowienke S, Sawant SG, Christensen J, Johnson TE, McKnight C, Ward G, Galloway SM, Custer L, Gocke E, O'Donovan MR, Braun K, Snyder RD, Mahadevan B. 2013a. Bacterial mutagenicity screening in the pharmaceutical industry. Mutat Res 752:99-118.

Escobar PA, Kemper RA, Tarca J, Nicolette J, Kenyon M, Glowienke S, Sawant SG, Christensen J, Johnson TE, McKnight C, Ward G, Galloway SM, Custer L, Gocke E, O'Donovan MR, Braun K, Snyder RD, Mahadevan B. 2013b. Bacterial mutagenicity screening in the pharmaceutical industry. Mutat Res 753(2):155.

Feinstein AR, Cicchetti DV. 1990. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 6:543-549.

Flamand N, Meunier J, Meunier P, Agapakis-Causs EC. 2001. Mini mutagenicity test: A miniaturized version of the Ames test used in a prescreening assay for point mutagenesis assessment. Toxicol In Vitro 15:105–114.

Flight L, Julious SA. 2015. The disagreeable behaviour of the kappa statistic. Pharm Stat 14:74–78.

Flückiger-Isler S, Baumeister M, Braun K, Gervais V, Hasler-Nguyen N, Reimann R, Van Gompel J, Wunderlich HG, Engelhardt G. 2004. Assessment of the performance of the Ames II assay: a collaborative study with 19 coded compounds. Mutat Res 558:181-197.

Flückiger-Isler S and Kamber M. 2012. Direct comparison of the Ames microplate format (MPF) test in liquid medium with the standard Ames pre-incubation assay on agar plates by use of equivocal to weakly positive test compounds. Mutat Res 747:36-45.

Flückiger-Isler S, Kamber M. 2014. Chapter 2: The Ames II and Ames MPF Penta I Assay: A Liquid Microplate Format Modification of the Classic Ames Test. L. María Sierra and Isabel Gaivão (eds.), Genotoxicity and DNA Repair: A Practical Approach, Methods in Pharmacology and Toxicology, Springer Science and Business Media, New York, USA.

Gee P, Maron DM, Ames BN. 1994. Detection and classification of mutagens: a set of base-specific Salmonella tester strains. Proc Natl Acad Sci USA 91:11606-11610.

Gee P, Sommers CH, Melick AS, Gidrol XM, Todd MD, Burris RB, Nelson ME, Klemm RC, Zeiger E. 1998. Comparison of the responses of base-specific Salmonella tester strains with the traditional strains for identifying mutagens: Results of a validation study. Mutat Res 412:115–130.

Gervais V, Bijot D, Claude N. 2003. Assessment of a screening experience with the Ames II™ test and future prospects. European Environmental Society 33th Annual Meeting: From Hazard to Risk, Aberdeen, p. 120.

Green MH, Muriel WJ, Bridges BA. 1976. Use of a simplified fluctuation test to detect low levels of mutagens. Mutat Res 38:33-42.

Hakura A, Hori Y, Uchida K, Sawada S, Suganuma A, Aoki T, Kazuo T. 2010. Inhibitory Effect of Dimethyl Sulfoxide on the Mutagenicity of Promutagens in the Ames Test. Genes Environ 32:53-60.

Hamada C, Wada T, Sakamoto Y. 1994. Statistical characterization of negative control data in the Ames Salmonella/microsome test. Environ Health Perspect, 102 Suppl 1, 115-119.

Hamel A, Roy M, Proudlock R. 2016. The bacterial reverse mutation test. In: Proudlock R, editor. Genetic Toxicology Testing, A Laboratory Manual. London, UK: Academic Press (Elsevier Inc.). pp 79–138.

Hartman PE, Hartman Z, Stahl RC, Ames BN. 1971. Classification and Mapping of Spontaneous and Induced Mutations in The Histidine Operon of Salmonella. In E. W. Caspari (Ed.), Academic Press. Adv Genet 16:1-34.

Heringa MB, Harmsen DJ, Beerendonk EF, Reus AA, Krul CA, Metz DH, IJpelaar GF. 2011. Formation and removal of genotoxic activity during UV/H(2)O(2)-GAC treatment of drinking water. Water Res. 45:366-74.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). 2017. ICH guideline M7(R1): Assessment and control of DNA reactive

(mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk – Step 5. Geneva, Switzerland. https://database.ich.org/sites/default/files/M7_R1_Guideline.pdf

Kado NY, Langley D, Eisenstadt E. 1983. A simple modification of the Salmonella liquid-incubation assay: Increased sensitivity for detecting mutagens in human urine. Mutat Res 121:25–32.

Kado NY, Guirguis GN, Flessel CP, Chan RC, Chang KI, Wesolowski JJ. 1986. Mutagenicity of Fine (Less Than 2.5 Microns) Airborne Particles: Diurnal Variation in Community Air Determined by a Salmonella Micro Preincubation (Microsuspension) Procedure. Environ Mutagen 8:53-66.

Kamber M, Flückiger-Isler S, Engelhardt G, Jaeckh R, Zeiger E. 2009. Comparison of the Ames II and traditional Ames test responses with respect to mutagenicity, strain specificities, need for metabolism and correlation with rodent carcinogenicity. Mutagenesis 24:359-366.

Kato M, Sugiyama KI, Fukushima T, Miura Y, Awogi T, Hikosaka S, Kawakami K, Nakajima M, Nakamura M, Sui H, Watanabe K, Hakura A. 2018. Negative and positive control ranges in the bacterial reverse mutation test: JEMS/BMS collaborative study, Genes Environ. 4;40:7.

Kirkland D, Kasper P, Martus H-J, Müller L, van Benthem J, Madia F, Corvi R. 2016. Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. Mutat Res 795:7-30.

Kirkland D, Levy DD, LeBaron MJ, Aardema MJ, Beevers C, Bhalli J, Douglas GR, Escobar PA, Farabaugh CS, Guerard M, Johnson GE, Kulkarni R, Le Curieux F, Long AS, Lott J, Lovell DP, Luijten M, Marchetti F, Nicolette JJ, Pfuhler S, Roberts DJ, Stankowski Jr LF, Thybaud V, Weiner SK, Williams A, Witt KL, Young R. 2019. A comparison of transgenic rodent mutation and in vivo comet assay responses for 91 chemicals. Mutat Res 839:21-35.

Kirkland DJ. 1994. Statistical evaluation of mutagenicity test data: recommendations of the U.K. Environmental Mutagen Society. Environ Health Perspect 102 Suppl 1:43-47.

Kupchella E, Koch WH, Cebula TA. 1994. Mutant alleles of tRNA (Thr) genes suppress the hisG46 missense mutation in Salmonella typhimurium. Environ Mol Mutagen 23:81-88.

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159-174.

Landrum G. 2015. RDKit: Open-Source Cheminformatics Software, http://www.rdkit.org.

Levy DD, Zeiger E, Escobar P, Hakura A, van der Leede B, Kato M, Moore MM, Sugiyama K. 2019. Recommended Criteria for the Evaluation of Bacterial Mutagenicity Data (Ames Test). Mutation Research - Genetic Toxicology and Environmental Mutagenesis 848: 403075.

Lorge E, Gervais V, Becourt-Lhote N, Maisonneuve C, Delongeas JL, Claude N. 2007. Genetic toxicity assessment: employing the best science for human safety evaluation part IV: a strategy in genotoxicity testing in drug development: some examples. Toxicol Sci 98:39–42.

Madia F, Kirkland D, Morita T, White P, Asturiol D, Corvi R. 2020. EURL ECVAM Genotoxicity and Carcinogenicity Database of Substances Eliciting Negative Results in the Ames Test: Construction of the Database, Mutat Res 854–855:503199.

McCann J, Horn L, Kaldor J. 1984. An evaluation of Salmonella (Ames) test data in the published literature: application of statistical procedures and analysis of mutagenic potency. Mutat Res 134:1-47.

Mellor CL, Marchese Robinson RL, Benigni R, Ebbrell D, Enoch SJ, Firman JW, Madden JC, Pawar G, Yang C, Cronin MTD. 2019. Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use. Regul Toxicol Pharmacol 101:121-134.

Miller JE, Vlasakova K, Glaab WE, Skopek TR. 2005. A low volume, high-throughput forward mutation assay in Salmonella typhimurium based on fluorouracil resistance. Mut Res 578:210–224.

Mori Y, Yamazaki H, Toyoshi K, Emi Y, Uchida K, Tsutsumi M, Konishi Y. 1985. Inhibitory effect of organic solvents on the mutagenicity of N-Nitrosodialkylamines in Salmonella. Mutat Res 142:153-158.

Mortelmans K, Zeiger E. 2000. The Ames Salmonella/microsome mutagenicity assay. Mutat Res 455:29-60.

Muster W, Albertini S, Chetelat A, Kirchner S, Gocke E. 2000. A Microsuspension Screening Version of the Ames Test to Predict the Outcome of GLP Assays for Gene Mutations. Environ Mol Mutagen 35:44.

Naveja JJ, Medina-Franco JL. 2019. Finding Constellations in Chemical Space Through Core Analysis. Front Chem 7:510.

Nicolette J, Murray J, Sonders P. 2015. 6-Well Bacterial Reverse Mutation Assay: Retrospective Evaluation of its Appropriateness for Assessing Mutagenicity of Impurities. Environ Mol Mutagen 56:S77.

Nicolette J, Dakoulas E, Pant K, Crosby M, Kondratiuk A, Murray J, Sonders P, Kulkarni R, Jayakumar G, Mathur M, Patel A, Vicente R, Datta K, Kolaja K. 2018. A comparison of 24 chemicals in the six-well bacterial reverse mutation assay to the standard 100-mm Petri plate bacterial reverse mutation assay in two laboratories. Regul Toxicol Pharmacol. 100:134-160.

Nurjannah, I, Siwi, SM. 2017. Guidelines for analysis on measuring interrater reliability of nursing outcome classification. Int J Res Med Sci 5:1169-1175.

OECD. 1997. Test Guideline No. 471, Bacterial Reverse Mutation Test. Health effects, section 4. Paris, France. oecd-ilibrary.org

OECD. 2005. Guidance document 34 on the validation and international acceptance of new or updated test methods for hazard assessment. Series on testing and assessment, number 34. Paris, France. https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-gd34.pdf

Ohta T, Tokishita S, Tsunoi R, Ohmae S, Yamagata H. 2002. Characterization of Trp(+) reversions in Escherichia coli strain WP2uvrA. Mutagenesis 17:313-316.

Pant K, Bruce S, Sly J, Klug Laforce M, Springer S, Cecil M, Andrus E, Dakoulas E, Wagner VO 3rd, Hewitt NJ, Kulkarni R. 2016. Bacterial mutagenicity assays: Vehicle and positive control results from the standard Ames assay, the 6- and 24-well miniaturized plate incorporation assays and the Ames II™ assay. Environ Mol Mutagen 57:483-496.

Prival MJ. 2001. Anomalous mutagenicity profile of cyclohexanone oxime in bacteria: cell survival in background lawns. Mutat. Res. 491:1-9.

Proudlock R, Evans K. 2016. The Micro-Ames Test: A direct comparison of the performance and sensitivities of the standard and 24-well plate versions of the bacterial mutation test. Environ Mol Mutagen 57:687-705.

Rainer B, Pinter E, Prielinger L, Coppola C, Marin-Kuan M, Schilter B, Apprich S, Tacker M. 2021. Direct comparison of the lowest effect concentrations of mutagenic reference substances in two Ames test formats. Toxics 9:152.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rogers D, Hahn M. 2010. Extended connectivity-fingerprints. J Chem Inf Model 50:742–754.
SAS Institute 2015. SAS/STAT® v14.1 User's Guide. SAS Institute, Cary, NC, USA.

Sawant S, Smith AB, Xu L. 2017. Validation of 6-and 24-Well Plate Bacterial Mutagenicity Assay. Int J Toxicol 36:59.

Schlosser MJ, Soelter SG, Wells M. 2007. Evaluation of a MicroAmes screen that is fully predictive for the regulatory Salmonella/E. coli mammalian microsome mutagenicity test, presentation delivered at the 46th annual meeting of the Society of Toxicology, Charlotte, NC, March 25-29, 2007, Abstract 1663.

Sim J, Wright CC. 2005. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 85:257-267.

Simel DL, Samsa GP, Matchar DB. 1991. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 44:763-770.

Smith R, Whitwell J, Keig-Shelvin Z, Ballantyne M, Bundy D, Steele A. 2017. Early genetic toxicology screening, mini Ames and micro TK6 with 20 mg of test article. Mutagenesis 32:621.

Spiliotopoulos D, Koelbert C. 2020. Assessment of the miniaturized liquid Ames microplate format (MPF™) for a selection of the test items from the recommended list of genotoxic and non-genotoxic chemicals. Mutat Res 856-857:503218.

Steven M, Sergeant E. 2021. Package epiR, Tools for Analysis of Epidemiological Data v. 2.0.19. cran.r-project.org/web/packages/epiR/epiR.pdf

Thybaud V, Lorge E, Levy DD, van Benthem J, Douglas GR, Marchetti F, Moore MM, Schoeny R. 2017. Main issues addressed in the 2014-2015 revisions to the OECD Genetic Toxicology Test Guidelines. Environ Mol Mutagen 58(5):284-295.

Umbuzeiro GA, Roubicek DA, Rech CM, Sato MIZ, Claxton LD. 2004. Investigating the sources of the mutagenic activity found in a river using the salmonella assay and different water extraction procedures. Chemosphere 54:1589–1597.

Watanabe T, Kohan MJ, Walsh D, Ball LM, DeMarini DM, Lewtas J. 1995. Mutagenicity of nitrodibenzopyranones in the Salmonella plate-incorporation and microsuspension assays. Mutat Res 345:1–9.

Wells et al. 2008. Validation of a MicroAmes Reverse Mutation Mammalian Microsome Screen using ICH-Compliant Salmonella and E. coli Tester Strains (EMS 2008).

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, USA.

Williams RV, DeMarini DM, Stankowski LF Jr, Escobar PA, Zeiger E, Howe J, Elespuru R, Cross KP. 2019. Are all bacterial strains required by OECD mutagenicity test guideline TG471 needed? Mutat Res Genet Toxicol Environ Mutagen. 2019 Dec;848:503081. doi: 10.1016/j.mrgentox.2019.503081.

Wilson JD Jr, Cariello NF. 1997. The Ames miniscreen assay: volatility of sodium azide can cause an increase in the reversion frequencies of adjacent, untreated wells. Environ Mol Mutagen. 29:217-219.

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46(D1): D1074-D1082.

Witkin EM. 1956. Time, temperature, and protein synthesis: a study of ultraviolet-induced mutation in bacteria. Cold Spring Harb Symp Quant Biol 21:123-140.

Zwarg JRRM, Morales DA, Maselli BS, Brack W, Umbuzeiro GA. 2018. Miniaturization of the Microsuspension Salmonella/Microsome Assay in Agar Microplates. Environ Mol Mutagen 59:488-501.

Annex I. OECD Expert group Survey on Miniaturised Ames Tests

The OECD Expert group on the miniaturised bacterial gene mutation tests (Miniaturised Ames Test) will be doing a comprehensive review of the miniaturised versions of the Ames test.

Background

The Bacterial Reverse Gene Mutation Test (Ames test – OECD TG 471) is one of the most widely used tests for early mutagenicity detection due to its relatively simple format and short assay times and its high relevance and reliability for testing genotoxic and carcinogenic compounds.

Several miniaturised versions of the bacterial gene mutation test have been developed and are already commonly used, particularly for early screening in product development. Potential advantages of these miniaturised versions include a significant reduction of test material, a reduction of costs and/or the possibility for simultaneous analyses of large number of samples. However, since they are not described in a Test Guideline, they don't benefit from Mutual Acceptance of Data (MAD), and results from these miniaturised bacterial gene mutations tests are not currently widely accepted by regulatory agencies. Thus, there is a need to evaluate these miniaturised versions.

During the 28th Meeting of the National Coordinators of the Test Guidelines Programme (April 2016), a Standard Project Submission Form (SPSF) for the preparation of a Detailed Review Paper (DRP) on the available miniaturised bacterial gene mutation tests was approved for inclusion in the Test Guidelines work plan. The aim of the DRP is to provide an overview of the existing miniaturised bacterial gene mutation tests and to assess whether these tests have the same pattern of sensitivity, specificity and reliability as the standard OECD TG 471 Ames test. Based on this assessment, recommendations for potential next steps will be made (e.g. no further actions needed, inclusion of the miniaturised bacterial gene mutation tests as an alternative to the standard bacterial gene mutation tests in OECD TGs).

An Expert Group (EG) on the miniaturised Ames tests was established by the OECD including experts from different member states/countries. During its kick-off meeting in February 2017, the EG decided to develop a survey in order to obtain better insights in (i) the amount of data available for the different miniaturised bacterial gene mutation tests and (ii) the extent to which these data can ultimately be shared. The information collected with this initial survey will help to support the decisions of the EG on the further actions required for the assessment of the miniaturised tests. As appropriate, this initial survey will be followed with a request for actual data sharing.

This initial survey is to determine 1) what information is already available and 2) if more validation/qualification studies would be useful to conduct.

<table>
<tr><th colspan="4">Questions</th></tr>
<tr><td colspan="2">Institution
Name</td><td colspan="2"></td></tr>
<tr><td colspan="2">Contact Information
Name
Email</td><td colspan="2"></td></tr>
<tr><td colspan="2">In your institution do you use any form of miniaturised Ames? Yes/No</td><td colspan="2"></td></tr>
<tr><td colspan="2">If yes, please name them
(e.g. 24-well plate, 6-well plate, Ames II, MPF, Bioluminescent Ames, 96-well plate, others)</td><td colspan="2"></td></tr>
<tr><td colspan="3">For each assay listed above please provide the following information (add more rows if needed)</td><td>Comments</td></tr>
<tr><td rowspan="10">Assay 1</td><td>Briefly describe the assay</td><td></td><td></td></tr>
<tr><td>How many chemicals have you tested in this assay (e.g. during the qualification of the assay)?</td><td></td><td></td></tr>
<tr><td>Please provide list of chemicals (with CAS#)</td><td></td><td></td></tr>
<tr><td>Do you have corresponding data in a full plate Ames?
If yes, for which chemicals in the above list?</td><td></td><td></td></tr>
<tr><td>Can you share the data with the OECD?</td><td></td><td></td></tr>
<tr><td>How many Proprietary compounds have you tested in this assay?</td><td></td><td></td></tr>
<tr><td>Do you have corresponding data in a full plate Ames?
If yes, for how many of these proprietary compounds?</td><td></td><td></td></tr>
<tr><td>Can you share the data with the OECD (coded/blinded)?</td><td></td><td></td></tr>
<tr><td>In your experience, please comment on the strengths and weaknesses and of this assay. for example are there chemicals or classes that this methodology is inappropriate for?</td><td></td><td></td></tr>
</table>

	In your company how is this assay used? (e.g., screening, test impurities, test metabolites, etc.)		
	Have the results from the miniaturised version been submitted to regulatory authorities? If so, were the results considered acceptable for regulatory use? If not, why?		
Assay 2	Briefly describe the assay		
	How many **chemicals** have you tested in this assay (e.g. during the qualification of the assay)?		
	Please provide list of chemicals (with CAS#)		
	Do you have corresponding data in a full plate Ames? If yes, for which chemicals in the above list?		
	Can you share the data with the OECD?		
	How many **Proprietary** compounds have you tested in this assay?		
	Do you have corresponding data in a full plate Ames? If yes, for how many of these proprietary compounds?		
	Can you share the data with the OECD (coded/blinded)?		
	In your experience, please comment on the strengths and weaknesses and of this assay, for example are there chemicals or classes that this methodology is inappropriate for?		
	In your company how is this assay used? (e.g., screening, test impurities, test metabolites, etc.)		
	Have the results from the miniaturised version been submitted to regulatory authorities? If so, were the results considered acceptable for regulatory use? If not, why?		
Assay 3… Add as many question blocks as the number of assays you wish to describe.			

Annex 2: Supplementary tables

Table 48. Test chemicals assessed and included in the retrospective analyses. An additional 188 coded test chemicals were assessed and included in the analyses.

#	Full Chemical Name	Abbreviation	CAS Number
1	(2-Chloroethyl)trimethylammonium chloride	CTMAC	999-81-5
2	(S)-1-[(R)-alpha-Methylbenzyl]-2-aziridinemethanol	MBAM	173143-73-2
3	1,2,3-Benzotriazole	BTAZ	95-14-7
4	1,2,3-Trichloropropane	TCP	96-18-4
5	1,2-Dichloroethane	12DIC	107-06-2
6	1,2-Dichloropropane	12DCP	78-87-5
7	1,2-Dimethylhydrazine	12DMH	540-73-8
8	1,3-Diphenylguanidine	13DPG	102-06-7
9	1,6-Dinitropyrene	16DNP	42397-64-8
10	1-Chloro-2-propanol	1CP	127-00-4
11	1-Chlorobutane	CBut	109-69-3
12	1-Ethyl-3-nitro-1-nitrosoguanidine	ENNG	4245-77-6
13	1H-Pyrazole-4-boronic acid	PBA	763120-58-7
14	1-Naphthylamine	1NA	134-32-7

15	1-Nitropyrene	1NP	5522-43-0
16	2,3-Butanedione	23BD	431-03-8
17	2’,3’-Dideoxyadenosine	DDAD	4097-22-7
18	2,4-Diaminotoluene	24DAT	95-80-7
19	2,4-Dichlorophenol	24DCP	120-83-2
20	2,4-Dinitrotoluene	24DNT	121-14-2
21	2,6-Diaminotoluene	26DAT	823-40-5
22	2,6-Dinitrotoluene	26DNT	606-20-2
23	2-(2-Furyl)-3-(5-nitro-2-furyl)acrylamide	AF2	3688-53-7
24	2-(4-Aminophenyl)-6-methylbenzothiazole	2APMB	92-36-4
25	2-Acetylaminofluorene	2AAF	53-96-3
26	2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine	PhIP	105650-23-5
27	2-Amino-3-methylimidazol[4,5-*f*]quinoline	IQ	76180-96-6
28	2-Amino-5-nitrophenol	2AmNP	121-88-0
29	2-Aminoanthracene	2AA	613-13-8
30	2-Aminoanthraquinone	AAQ	117-79-3
31	2-Aminobenzimidazole	2ABI	934-32-7
32	2-Aminofluorene	2AF	153-78-6
33	2-Bromo-4,6-dinitroaniline	2BDNA	1817-73-8
34	2-Chloropyrimidine	2CP	1722-12-9
35	2-Ethyl-1,3-hexanediol	2EHD	94-96-2
36	2-Hydroxybenzoic acid	2HBA	69-72-7
37	2-Naphthylamine	2NA	91-59-8
38	2-Naphthylamine-1-sulfonic acid	2NAS	81-16-3
39	2-Nitrobenzyl chloride	2NBC	612-23-7
40	2-Nitroethanol	2NE	625-48-9

41	2-Nitrofluorene	2NF	607-57-8
42	2-Nitropropane	2NP	79-46-9
43	3,3’,5,5’-Tetramethylbenzidine	3355TMB	54827-17-7
44	3,6-Dinitrobenzo[*a*]pyrene	DNBAP	128714-76-1
45	3-Methylcholanthrene	3MCA	56-49-5
46	3-Nitrobenzanthrone	3NBA	17117-34-9
47	3-Nitrofluoranthene	3NF	892-21-7
48	4,4’-Methylenebis(2-chloroaniline)	MBCA	101-14-4
49	4,4’-Methylenedianiline	4MDA	101-77-9
50	4,4’-Thiodianiline	4TA	139-65-1
51	4-Aminobiphenyl	4ABP	92-67-1
52	4-Methoxycarbonylphenylboronic acid	4MCPB	99768-12-4
53	4-Nitroanisole	4NA	100-17-4
54	4-Nitropyrene	4NP	57835-92-4
55	4-Nitroquinoline-*N*-oxide	4NQO	56-57-5
56	5-Azacytidine	5AC	320-67-2
57	5-Fluorouracil	5FU	51-21-8
58	6-Aminochrysene	6AC	2642-98-0
59	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0
60	6-Mercaptopurine	6MP	50-44-2
61	6-Methylquinoline	6MQ	91-62-3
62	7,12-Dimethylbenz[*a*]anthracene	DMBA	57-97-6
63	8-Hydroxyquinoline	8HQ	148-24-3
64	9,10-Dimethylanthracene	DMAN	781-43-1
65	9-Aminoacridine	9AA	90-45-9
66	9-Aminoacridine hydrochloride	9AA	134-50-9

67	9-Aminoacridine hydrochloride monohydrate	9AA	52417-22-8
68	9-Aminoanthracene	9AANT	779-03-3
69	Acetaldehyde oxime	AcOx	107-29-9
70	Acetamide	ACM	60-35-5
71	Acrolein	ACR	107-02-8
72	Acrylamide	ACY	79-06-1
73	Acrylonitrile	ACRY	107-13-1
74	Aflatoxin B1	AFLB1	1162-65-8
75	Allyl alcohol	AOH	107-18-6
76	Allyl glycidyl ether	AGE	106-92-3
77	Amitrol	AMTr	61-82-5
78	Ampicillin trihydrate	AmpT	7177-48-2
79	Aniline	ANI	62-53-3
80	Anthracene	ANT	120-12-7
81	Azidothymidine	AZT	30516-87-1
82	Azoxybenzene	AZB	495-48-7
83	Benzene	BENZ	71-43-2
84	Benzidine	BZD	92-87-5
85	Benzo[*a*]pyrene	BAP	50-32-8
86	Benzo[*k*]fluoranthene	BkF	207-08-9
87	Benzyl alcohol	BenOH	100-51-6
88	Beta-butyrolactone	BBL	3068-88-0
89	C.I. Acid Red 26	PX	3761-53-3
90	C.I. Direct Blue 1	Blue1	2610-05-1
91	C.I. Disperse Blue 373	Blue373	51868-46-3
92	C.I. Disperse Orange 30	Orange30	5261-31-4[a]

93	C.I. Disperse Orange 37	Orange37	13301-61-6[b]
94	C.I. Disperse Violet 93	Violet93	52697-38-8
95	C.I. Disperse Yellow 3	Yellow3	2832-40-8
96	Cadmium chloride	CdCl	10108-64-2
97	Catechol	CAT	120-80-9
98	Chloramphenicol	CAP	56-75-7
99	Chloroacetaldehyde	CAC	107-20-0
100	Chlorpheniramine maleate	CPAM	113-92-8
101	Cisplatin	Cplat	15663-27-1
102	Colchicine	COLCH	64-86-8
103	Crotonaldehyde	CROT	123-73-9
104	Crystal violet	CV	548-62-9
105	Cumene hydroperoxide	CuHyp	80-15-9
106	Cumene hydroperoxide, 80%	CuHyp	80-15-9
107	Cyclohexanone	CH	108-94-1
108	Cyclophosphamide	CPA	50-18-0
109	Cytosine arabinoside	CYAR	147-94-4
110	DL-Menthol	MEN	1490-04-6[c]
111	Danthron	DThr	117-10-2
112	Daunomycin	DAUN	20830-81-3
113	Di(2-ethylhexyl) phthalate	DEHP	117-81-7
114	Diaminodiphenyl ether	DDPE	101-80-4
115	Diethanolamine	DIET	111-42-2
116	Diethylstilbestrol	DES	56-53-1
117	Dimenhydrinate	DMHyd	523-87-5
118	Dimethoate	DIMET	60-51-5

119	Dimethyl yellow	DMY	60-11-7
120	4-Dimethylaminobenzenediazosulfonic acid sodium salt	DMAB	140-56-7
121	Dimethylcarbamoyl chloride	DMCC	79-44-7
122	Diphenylnitrosamine	DPN	86-30-6
123	Emodin	EMOD	518-82-1
124	Ephedrine hydrochloride	EPHED	50-98-6
125	Epichlorohydrin	ECH	106-89-8
126	Epinephrine	Epin	51-43-4
127	Erythromycin	ERY	114-07-8
128	Ethidium bromide	EtBr	1239-45-8
129	Ethionamide	EtiA	536-33-4
130	Ethyl acrylate	EthAc	140-88-5
131	Ethylene diamine	EDA	107-15-3
132	Ethylene glycol diethyl ether	EGDE	629-14-1
133	Ethyl methanesulfonate	EMS	62-50-0
134	Etoposide	ETOP	33419-42-0
135	Eugenol	Eug	97-53-0
136	Fisetin	FIS	528-48-3
137	Formaldehyde	Form	50-00-0
138	Gamma-butyrolactone	GBL	96-48-0
139	Glutaraldehyde	Glut	111-30-8
140	Glyoxal	GLY	107-22-2
141	Hexachloroethane	HClE	67-72-1
142	Hexamethylphosphoramide	HMP	680-31-9
143	Hydralazine	HYDL	86-54-4
144	Hydrazine dihydrochloride	HYDZ	5341-61-7

145	Hydrazine sulfate	HYDZ	1184-66-3
146	Hydrogen peroxide	H2O2	7722-84-1
147	Hydroquinone	HQ	123-31-9
148	Hydroxylamine	HA	7803-49-8
149	ICR-170	ICR170	146-59-8
150	Isobutyraldehyde	IsoBut	78-84-2
151	Isopropyl N-(3-chlorophenyl)carbamate	ICPC	101-21-3
152	L-Methionine	LMETH	63-68-3
153	D-Limonene	DLIMO	5989-27-5
154	Limonene	LIMO	138-86-3
155	Maltol	Malt	118-71-8
156	D-Mannitol	MAN	69-65-8
157	Melamine	MELA	108-78-1
158	2-Methoxy-5-nitroaniline	MetNA	99-59-2
159	Methyl carbamate	MCarb	598-55-0
160	Methyl vinyl ketone	3BUT	78-94-4
161	Methyl methanesulfonate	MMS	66-27-3
162	Michler's ketone	MK	90-94-8
163	Mitomycin C	MMC	50-07-7
164	Morin	MOR	480-16-0
165	N,N,N',N'-Tetramethyl-3,6-acridinediamine hydrochloride hydrate	TMAD	89722-22-5[d]
166	N,N,N'N-Tetramethyl-4,4'-methylenedianiline	NTM	101-61-1
167	N,N'-Dicyclohexylthiourea	DCTU	1212-29-9
168	N,N'-Ethylenethiourea	ETU	96-45-7
169	N-Ethyl-N-nitrosourea	ENU	759-73-9
170	N-Methylaniline	NMAN	100-61-8

171	N-Methyl-N’-nitro-N-nitrosoguanidine	MNNG	70-25-7
172	N-Nitrosodimethylamine	DMN	62-75-9
173	N4-Aminocytidine	N4AC	57294-74-3
174	Neutral Red, 92%	NRd	553-24-2
175	Nitrobenzene	NB	98-95-3
176	Nitrofurantoin	NFn	67-20-9
177	Nitromethane	NiMET	75-52-5
178	Norfloxacin	NOR	70458-96-7
179	O-Benzylhydroxylamine hydrochloride	BHAHCl	2687-43-6
180	O-Toluidine	TOL	95-53-4
181	Olaquindox	OQ	23696-28-8
182	Orthanilic acid	ORTH	88-21-1
183	Phenanthrene	Phen	85-01-8
184	Phenformin hydrochloride	PhenF	834-28-6
185	Phenol	PHL	108-95-2
186	Phenol Red	PhR	143-74-8
187	Phthalic acid	PTacid	88-99-3
188	Phthalic anhydride	PTA	85-44-9
189	Potassium dichromate	KCrO	7778-50-9
190	Primidone	Prim	125-33-7
191	Proflavine	PF	92-62-6
192	Proflavine dihydrochloride	PF	531-73-7
193	Propanedial	PPD	542-78-9
194	Pyrene	Pyr	129-00-0
195	Pyrene-1,6-quinone	PQ	1785-51-9
196	Pyridine	Pyri	110-86-1

197	Quercetin	QUER	117-39-5
198	Quercetin dihydrate	QUER	6151-25-3
199	Quinoline	Quin	91-22-5
200	Resorcinol	Res	108-46-3
201	Riboflavin 5'-phosphate sodium	RBFPS	130-40-5
202	Safrole	SAF	94-59-7
203	Sodium azide	SA	26628-22-8
204	Sodium arsenite	NaAs	7784-46-5
205	Sodium diclofenac	NaDIC	15307-79-6
206	Sodium lauryl sulphate	SLS	151-21-3
207	Sodium nitrite	SN	7632-00-0
208	Sterigmatocystin	SMC	10048-13-2
209	Streptomycin sulfate	STREP	3810-74-0
210	Streptonigrin	STRPN	3930-19-6
211	Streptonigrin methyl ester	STRPNme	3398-48-9
212	Sucrose	SUCR	57-50-1
213	Sulfisoxazole	SULX	127-69-5
214	Taxol	Tax	33069-62-4
215	Tetraethylthiuram disulfide	TETD	97-77-8
216	Tetrakis(hydroxymethyl)phosphonium chloride	TKPC	124-64-1
217	Thiourea	TU	62-56-6
218	Topiramate	Top	97240-79-4
219	Triethyl phosphate	TEP	78-40-0
220	Trimethyl phosphate	TMP	512-56-1
221	Tris(2,3-dibromopropyl) phosphate	TDBP	126-72-7
222	Tris(2-chloroethyl) phosphate	TCEP	115-96-8

223	Tris(2-ethylhexyl) phosphate	TEHP	78-42-2
224	Urea	UREA	57-13-6
225	Urethane	URET	51-79-6
226	Vinblastine Sulfate	VinS	143-67-9
227	Zafirlukast	ZaFk	107753-78-6
228	Zonisamide sodium salt	ZonNa	68291-98-5
229	m-Phenylenediamine	MPD	108-45-2
230	m-Toluidine	MTOL	108-44-1
231	n-Butyl chloride	NBC	109-69-3
232	o-Anisidine	oAN	90-04-0
233	o-Anthranilic acid	oAA	118-92-3
234	o-Dianisidine	oD	119-90-4
235	p-Chloroaniline	PCAN	106-47-8
236	p-Chloroaniline hydrochloride	PCAN	20265-96-7
237	p-Nitrophenol	pNP	100-02-7
238	p-Toluidine	PTOL	106-49-0
239	tert-Butyl alcohol	TBA	75-65-0
240	tert-Butyl carbazate	TBZ	870-46-2
241	tert-Butylhydroquinone	tBHQ	1948-33-0

[a]CAS number 12223-23-3 indicated on the ECHA website
[b]CAS number 12223-33-5 indicated in ChemicalBook
[c]PubChem also lists CAS number 89-78-1
[d]PubChem lists 65-61-2 for the hydrochloride

Table 49. Test chemicals that elicited equivocal responses on the standard bacterial reverse gene mutation test. All entries reflect the call provided by the data submitter.

Obs #	Chemical name	Abbreviation	Strain	S9
1	1,3-Diphenylguanidine	13DPG	TA100	+S9
2	1,3-Diphenylguanidine	13DPG	TA1535	-S9
3	1,3-Diphenylguanidine	13DPG	TA1537	+S9
4	1,3-Diphenylguanidine	13DPG	TA1537	-S9
5	2-Aminoanthracene	2AA	TA100	+S9
6	2-Amino-5-nitrophenol	2AmNP	TA100	-S9
7	2-Amino-5-nitrophenol	2AmNP	TA1537	+S9
8	2-Amino-5-nitrophenol	2AmNP	TA1537	-S9
9	2-Nitrobenzyl chloride	2NBC	TA98	-S9
10	2-Nitroethanol	2NE	E. coli WP2 uvrA pKM101	-S9
11	2-Nitropropane	2NP	TA98	-S9
12	3-Methylcholanthrene	3MCA	TA98	+S9
13	4,4-Thiodianiline	4TA	TA98	+S9
14	5-Azacytidine	5AC	TA1535	-S9
15	6-Methylquinoline	6MQ	TA98	+S9
16	2-Aminoanthraquinone	AAQ	TA98	-S9
17	Amitrol	AMTr	TA1535	+S9
18	Anthracene	ANT	TA100	+S9
19	Anthracene	ANT	TA1535	+S9
20	Anthracene	ANT	TA98	+S9
21	Azoxybenzene	AZB	TA98	+S9
22	Acetaldehyde oxime	AcOx	TA100	+S9
23	Benzo[a]pyrene	BAP	TA1535	+S9
24	1,2,3-Benzotriazole	BTAZ	TA1535	+S9
25	Benzo[k]fluoranthene	BkF	TA98	+S9
26	Chlorpheniramine maleate	CPAM	TA100	-S9
27	Cadmium chloride	CdCl	TA98	-S9
28	Di-(2-ethylhexyl) phthalate	DEHP	TA100	+S9
29	Diethylstilbestrol	DES	TA98	+S9
30	Diethanolamine	DIET	TA100	+S9

31	Dimethylcarbamoyl chloride	DMCC	TA98	+S9
32	Dimethylcarbamoyl chloride	DMCC	TA98	-S9
33	Dimenhydrinate	DMHyd	TA100	+S9
34	Dimenhydrinate	DMHyd	TA100	-S9
35	Dimenhydrinate	DMHyd	TA1535	+S9
36	Dimenhydrinate	DMHyd	TA1535	-S9
37	N-Nitrosodimethylamine	DMN	TA1535	+S9
38	Dimethyl yellow	DMY	TA100	+S9
39	Dimethyl yellow	DMY	TA98	+S9
40	Ethylene diamine	EDA	TA100	-S9
41	Ethylene diamine	EDA	TA1535	+S9
42	Ethylene diamine	EDA	TA1535	-S9
43	Ethylmethanesulfonate	EMS	TA98	+S9
44	Ethyl acrylate	EthAc	TA100	+S9
45	G_15	G_15	TA100	+S9
46	G_20	G_20	TA100	+S9
47	Glutaraldehyde	Glut	TA100	-S9
48	D-Mannitol	MAN	TA1535	-S9
49	4,4-Methylene bis(2-chloroaniline)	MBCA	TA98	+S9
50	Methylmethanesulfonate	MMS	TA1535	-S9
51	Methylmethanesulfonate	MMS	TA1537	+S9
52	Methylmethanesulfonate	MMS	TA98	+S9
53	Methylmethanesulfonate	MMS	TA98	-S9
54	Maltol	Malt	TA100	+S9
55	Maltol	Malt	TA1535	+S9
56	N,N,N,N-Tetramethyl-4,4-methylenedianiline	NTM	TA98	+S9
57	Orthanilic acid	ORTH	TA98	-S9
58	Phenanthrene	Phen	TA100	+S9
59	Primidone	Prim	TA100	-S9
60	Primidone	Prim	TA1535	-S9
61	Pyrene	Pyr	TA1537	+S9
62	Quinoline	Quin	TA98	+S9
63	Sulfisoxazole	SULX	TA1535	+S9
64	p-Nitrophenol	pNP	TA100	+S9

65	p-Nitrophenol	pNP	TA98	-S9
66	tert-Butylhydroquinone	tBHQ	TA98	-S9

Table 50. Test chemicals that elicited equivocal responses on the 6-well plate assay. All entries reflect the call provided by the data submitter.

Obs #	Chemical name	Abbreviation	Strain	S9
1	2-Aminoanthracene	2AA	TA100	+S9
2	2-Aminoanthracene	2AA	TA100	-S9
3	2-Aminoanthracene	2AA	TA97	-S9
4	4-Methoxycarbonyl phenylboronic acid	4MCPB	TA97	+S9
5	4-Nitroquinoline-N-oxide	4NQO	TA98	+S9
6	6-Mercaptopurine	6MP	E. coli WP2 uvrA	+S9
7	6-Mercaptopurine	6MP	E. coli WP2 uvrA pKM101	+S9
8	9-Aminoanthracene	9AANT	TA1537	-S9
9	Benzo[a]pyrene	BAP	TA1535	+S9
10		CmpdD	TA1535	+S9
11	7,12-Dimethylbenz[a]anthracene	DMBA	TA98	-S9
12	H_2	H_2	TA98	+S9
13	tert-Butyl alcohol	TBA	E. coli WP2 uvrA	+S9
14	tert-Butyl alcohol	TBA	TA98	+S9

Table 51. Test chemicals that elicited equivocal responses on the Ames MPF™ assay. All entries reflect the call provided by the data submitter.

Obs #	Chemical name	Abbreviation	Strain	S9
1	1,3-Diphenylguanidine	13DPG	TA100	+S9
2	1,3-Diphenylguanidine	13DPG	TA1537	+S9
3	1,3-Diphenylguanidine	13DPG	TA1537	-S9
4	2,4-Diaminotoluene	24DAT	E. coli WP2 uvrA pKM101	+S9
5	2,4-Diaminotoluene	24DAT	TA100	+S9
6	2,4-Diaminotoluene	24DAT	TA1537	+S9
7	2-Aminoanthracene	2AA	TA100	+S9
8	2-Acetylaminofluorene	2AAF	E. coli WP2 uvrA pKM101	+S9
9	2-Acetylaminofluorene	2AAF	TA98	-S9
10	2-Bromo-4,6-dinitroaniline	2BDNA	TA100	+S9
11	2-Bromo-4,6-dinitroaniline	2BDNA	TA98	+S9
12	2-Ethyl-1,3-hexanediol	2EHD	TA1535	+S9
13	2-Ethyl-1,3-hexanediol	2EHD	TA98	+S9
14	2-Nitrofluorene	2NF	TA100	+S9
15	2-Nitrofluorene	2NF	TA98	+S9
16	4-Nitroquinoline-N-oxide	4NQO	E. coli WP2 uvrA pKM101	+S9
17	4-Nitroquinoline-N-oxide	4NQO	TA100	+S9
18	4-Nitroquinoline-N-oxide	4NQO	TA1537	+S9
19	4-Nitroquinoline-N-oxide	4NQO	TA1537	-S9
20	4-Nitroquinoline-N-oxide	4NQO	TA98	+S9
21	5-Azacytidine	5AC	TA1535	-S9
22	6-Aminochrysene	6AC	TA98	+S9
23	6-Mercaptopurine	6MP	TA1535	-S9
24	9-Aminoacridine	9AA	TA98	-S9
25	9-Aminoacridine hydrochloride monohydrate	9AA	TA98	-S9
26	Aflatoxin B1	AFLB1	E. coli WP2 uvrA pKM101	+S9
27	Amitrol	AMTr	TA1535	+S9
28	Amitrol	AMTr	TA1535	-S9
29	Azoxybenzene	AZB	TA98	+S9
30	Benzo[a]pyrene	BAP	E. coli WP2 uvrA pKM101	+S9

31	Benzo[a]pyrene	BAP	TA100	+S9
32	Benzo[a]pyrene	BAP	TA1535	+S9
33	1,2,3-Benzotriazole	BTAZ	E. coli WP2 uvrA pKM101	+S9
34	1,2,3-Benzotriazole	BTAZ	TA1535	+S9
35	1,2,3-Benzotriazole	BTAZ	TA1535	-S9
36	1,2,3-Benzotriazole	BTAZ	TA1537	+S9
37	Cyclophosphamide	CPA	TA100	+S9
38	Cyclophosphamide	CPA	TA1535	-S9
39	Cyclophosphamide	CPA	TA1537	+S9
40	Cyclophosphamide	CPA	TA98	+S9
41	Crotonaldehyde	CROT	E. coli WP2 pKM101	-S9
42	Cadmium chloride	CdCl	E. coli WP2 uvrA pKM101	+S9
43	Cisplatin	Cplat	TA100	+S9
44	Cisplatin	Cplat	TA98	+S9
45	Cumene hydroperoxide	CuHyp	E. coli WP2 uvrA pKM101	-S9
46	Cumene hydroperoxide	CuHyp	TA98	+S9
47	Cumene hydroperoxide	CuHyp	TA98	-S9
48	Daunomycin	DAUN	TA100	+S9
49	Daunomycin	DAUN	TA98	+S9
50	Diethanolamine	DIET	E. coli WP2 uvrA pKM101	-S9
51	7,12-Dimethylbenz[a]anthracene	DMBA	E. coli WP2 uvrA pKM101	+S9
52	7,12-Dimethylbenz[a]anthracene	DMBA	E. coli WP2 uvrA pKM101	-S9
53	7,12-Dimethylbenz[a]anthracene	DMBA	TA100	+S9
54	7,12-Dimethylbenz[a]anthracene	DMBA	TA1535	-S9
55	7,12-Dimethylbenz[a]anthracene	DMBA	TA1537	-S9
56	7,12-Dimethylbenz[a]anthracene	DMBA	TA98	+S9
57	Danthron	DThr	TA98	+S9
58	N-Ethyl-N-nitrosourea	ENU	E. coli WP2 uvrA pKM101	+S9
59	N-Ethyl-N-nitrosourea	ENU	E. coli WP2 uvrA pKM101	-S9
60	N-Ethyl-N-nitrosourea	ENU	TA100	-S9
61	N-Ethyl-N-nitrosourea	ENU	TA1535	-S9
62	N-Ethyl-N-nitrosourea	ENU	TA1537	+S9
63	N-Ethyl-N-nitrosourea	ENU	TA1537	-S9
64	N-Ethyl-N-nitrosourea	ENU	TA98	-S9

65	Epinephrine	Epin	TA1537	-S9
66	Ethyl acrylate	EthAc	TA1535	+S9
67	Eugenol	Eug	E. coli WP2 uvrA pKM101	+S9
68	Formaldehyde	Form	E. coli WP2 uvrA pKM101	-S9
69	Hydroquinone	HQ	E. coli WP2 uvrA pKM101	-S9
70	Hydroquinone	HQ	TA98	+S9
71	2-Amino-3-methylimidazol[4,5-f]quinoline	IQ	E. coli WP2 uvrA pKM101	+S9
72	2-Amino-3-methylimidazol[4,5-f]quinoline	IQ	TA1535	+S9
73	2-Amino-3-methylimidazol[4,5-f]quinoline	IQ	TA1537	-S9
74	Limonene	LIMO	TA100	+S9
75	D-Mannitol	MAN	E. coli WP2 uvrA pKM101	-S9
76	Methyl carbamate	MCarb	E. coli WP2 uvrA pKM101	-S9
77	Methyl carbamate	MCarb	TA98	+S9
78	DL-Menthol	MEN	E. coli WP2 uvrA pKM101	-S9
79	Methylmethanesulfonate	MMS	TA100	-S9
80	Methylmethanesulfonate	MMS	TA1535	-S9
81	Methylmethanesulfonate	MMS	TA1537	-S9
82	Methylmethanesulfonate	MMS	TA98	+S9
83	Methylmethanesulfonate	MMS	TA98	-S9
84	Maltol	Malt	E. coli WP2 uvrA pKM101	+S9
85	Maltol	Malt	TA1535	+S9
86	Maltol	Malt	TA1535	-S9
87	Maltol	Malt	TA1537	+S9
88	Maltol	Malt	TA1537	-S9
89	Sodium diclofenac	NaDIC	E. coli WP2 uvrA pKM101	+S9
90	Sodium diclofenac	NaDIC	E. coli WP2 uvrA pKM101	-S9
91	p-Chloroaniline	PCAN	E. coli WP2 uvrA pKM101	-S9
92	p-Chloroaniline	PCAN	TA98	+S9
93	Proflavine	PF	TA1535	+S9
94	Proflavine	PF	TA1535	-S9
95	Phenanthrene	Phen	TA100	+S9
96	Phenanthrene	Phen	TA98	-S9
97	Phenformin HCl	PhenF	E. coli WP2 uvrA pKM101	-S9
98	Phenformin HCl	PhenF	TA98	-S9

99	Pyrene	Pyr	TA98	-S9
100	Pyridine	Pyri	E. coli WP2 uvrA pKM101	+S9
101	Pyridine	Pyri	E. coli WP2 uvrA pKM101	-S9
102	Pyridine	Pyri	TA1535	-S9
103	Pyridine	Pyri	TA1537	-S9
104	Pyridine	Pyri	TA98	-S9
105	Quinoline	Quin	TA98	+S9
106	Sodium azide	SA	TA100	-S9
107	Sulfisoxazole	SULX	E. coli WP2 uvrA pKM101	-S9
108	Sulfisoxazole	SULX	TA1535	+S9
109	tert-Butyl alcohol	TBA	TA98	+S9
110	Tris(2-ethylhexyl) phosphate	TEHP	E. coli WP2 uvrA pKM101	+S9
111	Tris(2-ethylhexyl) phosphate	TEHP	TA100	+S9
112	Tris(2-ethylhexyl) phosphate	TEHP	TA98	+S9
113	Tetraethylthiuram disulfide	TETD	TA100	+S9
114	Tetrakis(hydroxymethyl) phosphonium chloride	TKPC	TA98	-S9
115	N,N,N,N-Tetramethyl-3,6-acridinediamine HCl hydrate	TMAD	TA100	+S9
116		Test11	TA100	-S9
117		Test11	TA98	+S9
118		Test2	TA1535	+S9
119		Test2	TA98	+S9
120	Urea	UREA	E. coli WP2 uvrA pKM101	-S9
121	Urea	UREA	TA100	+S9
122	o-Anthranilic acid	oAA	TA1537	-S9
123	p-Nitrophenol	pNP	TA1537	-S9
124	p-Nitrophenol	pNP	TA98	+S9

Table 52. Test chemicals that elicited equivocal responses on the 24-well plate assay. All entries reflect the call provided by the data submitter.

Obs #	Chemical name	Abbreviation	Strain	S9
1	4-Methoxycarbonyl phenylboronic acid	4MCPB	E. coli WP2 uvrA pKM101	+S9
2	4-Nitroquinoline-N-oxide	4NQO	E. coli WP2 uvrA pKM101	+S9
3	6-Mercaptopurine	6MP	E. coli WP2 uvrA	-S9
4	6-Mercaptopurine	6MP	E. coli WP2 uvrA pKM101	-S9
5	6-Mercaptopurine	6MP	TA100	-S9
6	Methylmethanesulfonate	MMS	TA97	+S9

Table 53. Test chemicals that elicited equivocal responses on the Ames II assay. All entries reflect the call provided by the data submitter.

Obs #	Chemical name	Abbreviation	Strain	S9
1	1,2-Dichloroethane	12DIC	TAMix	-S9
2	2,4-Diaminotoluene	24DAT	TA98	-S9
3	2-Nitropropane	2NP	TA98	+S9
4	2-Nitropropane	2NP	TAMix	-S9
5	Acrylonitrile	ACRY	TA98	+S9
6	Chloroacetaldehyde	CAC	TAMix	-S9
7	Diaminodiphenyl ether	DDPE	TAMix	-S9
8	Dimethylcarbamoyl chloride	DMCC	TA98	-S9
9	Glyoxal	GLY	TA98	+S9
10	Hexamethylphosphoramide	HMP	TAMix	-S9
11	Methylmethanesulfonate	MMS	TAMix	-S9
12	N-Methylaniline	NMAN	TA98	+S9
13	p-Toluidine	PTOL	TAMix	+S9
14	Phthalic acid	PTacid	TA98	-S9
15	Trimethyl phosphate	TMP	TAMix	+S9
16	O-Toluidine	TOL	TAMix	+S9
17	Urea	UREA	TAMix	-S9
18	o-Dianisidine	oD	TAMix	+S9

Table 54. Summary of correspondence mismatches; 6-well plate assay versus standard assay, individual assessments. Equivocal calls not included in the analysis.

#	Chemical Name	Abbreviation	CAS Number	Bacterial Strain	S9	Standard Call	6-Well Call
1	2,4-Diaminotoluene	24DAT	95-80-7	TA98	-S9	Negative	Positive
2	4-Methoxycarbonyl phenylboronic acid	4MCPB	99768-12-4	E. coli WP2 uvrA pKM101	-S9	Negative	Positive
3	6-Mercaptopurine	6MP	50-44-2	E. coli WP2 uvrA	-S9	Negative	Positive
4		C_2	NA	TA98	+S9	Negative	Positive
5	7,12-Dimethylbenz[a]anthracene	DMBA	57-97-6	TA97	-S9	Negative	Positive
6	Dimethyl yellow	DMY	60-11-7	E. coli WP2 uvrA pKM101	-S9	Negative	Positive
7	Dimethyl yellow	DMY	60-11-7	TA100	-S9	Negative	Positive
8	Dimethyl yellow	DMY	60-11-7	TA97	-S9	Negative	Positive
9	Dimethyl yellow	DMY	60-11-7	TA98	-S9	Negative	Positive
10	1-H-Pyrazole-4-boronic acid	PBA	763120-58-7	TA100	+S9	Negative	Positive
11	Sodium azide	SA	26628-22-8	TA100	+S9	Negative	Positive
12	tert-Butyl carbazate	TBZ	870-46-2	E. coli WP2 uvrA pKM101	+S9	Negative	Positive
1	4-Nitroquinoline-N-oxide	4NQO	56-57-5	TA1537	-S9	Positive	Negative
2		B2_009		TA1535	-S9	Positive	Negative
3		B2_012		TA98	+S9	Positive	Negative
4		B_11		E. coli WP2 uvrA	-S9	Positive	Negative
5	Cyclophosphamide	CPA	50-18-0	TA100	-S9	Positive	Negative
6	7,12-Dimethylbenz[a]anthracene	DMBA	57-97-6	TA1537	-S9	Positive	Negative
7		N2_004		TA1535	+S9	Positive	Negative
8		N2_006		TA1535	-S9	Positive	Negative
9		N2_010		TA97	+S9	Positive	Negative
10		N2_016		TA1535	+S9	Positive	Negative
11		N2_016		TA1535	-S9	Positive	Negative

Table 55. Summary of correspondence mismatches; 24-well plate assay versus standard assay, individual assessments. Equivocal calls not included in the analysis.

#	Chemical Name	Abbreviation	CAS Number	Bacterial Strain	S9	Standard Call	24-well Call
1	2-Nitrofluorene	2NF	607-57-8	TA1535	-S9	Negative	Positive
2	4-Methoxycarbonyl phenylboronic acid	4MCPB	99768-12-4	E. coli WP2 uvrA	-S9	Negative	Positive
3	4-Methoxycarbonyl phenylboronic acid	4MCPB	99768-12-4	E. coli WP2 uvrA pKM101	-S9	Negative	Positive
4	4-Methoxycarbonyl phenylboronic acid	4MCPB	99768-12-4	TA97	-S9	Negative	Positive
5	8-Hydroxyquinoline	8HQ	148-24-3	TA100	+S9	Negative	Positive
1	2-Aminoanthracene	2AA	613-13-8	TA97	-S9	Positive	Negative
2	4-Nitroquinoline-N-oxide	4NQO	56-57-5	TA98	+S9	Positive	Negative
3	Benzo[a]pyrene	BAP	50-32-8	E. coli WP2 uvrA pKM101	+S9	Positive	Negative
4	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0	TA100	+S9	Positive	Negative
5	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0	TA98	+S9	Positive	Negative
6	Methylmethanesulfonate	MMS	66-27-3	TA1535	+S9	Positive	Negative

Table 56. Summary of correspondence mismatches; Ames MPF™ assay versus standard, individual assessments. Equivocal calls not included in the analysis.

#	Chemical Name	Abbreviation	CAS Number	Bacterial Strain	S9	Standard Call	Ames MPF™ Call
1	2,4-Diaminotoluene	24DAT	95-80-7	TA1535	+S9	Negative	Positive
2	2-Acetylaminofluorene	2AAF	53-96-3	TA1535	+S9	Negative	Positive
3	2-Naphthylamine-1-sulfonic acid	2NAS	81-16-3	TA98	+S9	Negative	Positive
4		CmpdQ7		TA100	+S9	Negative	Positive
5	Cumene hydroperoxide	CuHyp	80-15-9	TA100	-S9	Negative	Positive
6	Diphenylnitrosamine	DPN	86-30-6	TA98	+S9	Negative	Positive
7	Diphenylnitrosamine	DPN	86-30-6	TA98	-S9	Negative	Positive
8	Glutaraldehyde	Glut	111-30-8	TA98	+S9	Negative	Positive
9	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-	ICR191	17070-45-0	TA1535	+S9	Negative	Positive

	methoxyacridine dihydrochloride						
10	Tris(2-ethylhexyl) phosphate	TEHP	78-42-2	TA1537	+S9	Negative	Positive
11		Test13		TA98	+S9	Negative	Positive
12		Test14		TA98	+S9	Negative	Positive
13		Test20		TA100	-S9	Negative	Positive
1	2-Aminoanthracene	2AA	613-13-8	TA97	-S9	Positive	Negative
2	6-Mercaptopurine	6MP	50-44-2	TA100	+S9	Positive	Negative
3	6-Mercaptopurine	6MP	50-44-2	TA100	-S9	Positive	Negative
4	2-(2-Furyl)-3-(5-nitro-2-furyl)acrylamide	AF2	3688-53-7	TA1537	-S9	Positive	Negative
5	Acetaldehyde oxime	AcOx	107-29-9	TA1535	+S9	Positive	Negative
6	Cyclophosphamide	CPA	50-18-0	TA100	-S9	Positive	Negative
7		CmpdQ5		TA98	-S9	Positive	Negative
8		CmpdQ6		TA1537	+S9	Positive	Negative
9		CmpdQ8		TA1535	+S9	Positive	Negative
10	7,12-Dimethylbenz[a]anthracene	DMBA	57-97-6	E. coli WP2 uvrA pKM101	+S9	Positive	Negative
11	7,12-Dimethylbenz[a]anthracene	DMBA	57-97-6	TA98	-S9	Positive	Negative
12	Epichlorohydrin	ECH	106-89-8	TA1537	+S9	Positive	Negative
13	Epichlorohydrin	ECH	106-89-8	TA1537	-S9	Positive	Negative
14	Epichlorohydrin	ECH	106-89-8	TA97	-S9	Positive	Negative
15	Epichlorohydrin	ECH	106-89-8	TA98	+S9	Positive	Negative
16	Epichlorohydrin	ECH	106-89-8	TA98	-S9	Positive	Negative
17	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0	TA1537	+S9	Positive	Negative
18	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0	TA97	+S9	Positive	Negative
19	m-Phenylenediamine	MPD	108-45-2	TA100	+S9	Positive	Negative
20	m-Phenylenediamine	MPD	108-45-2	TA1535	+S9	Positive	Negative
21	m-Phenylenediamine	MPD	108-45-2	TA98	-S9	Positive	Negative
22	Quinoline	Quin	91-22-5	E. coli WP2 uvrA	+S9	Positive	Negative
23	Streptomycin sulfate	STREP	3810-74-0	TA98	+S9	Positive	Negative
24		Test13		TA100	-S9	Positive	Negative
25		Test14		TA100	-S9	Positive	Negative
26		Test22		TA1535	-S9	Positive	Negative
27		Test4		TA100	+S9	Positive	Negative
28		Test5		TA1535	+S9	Positive	Negative

29		Test5		TA1535	-S9	Positive	Negative
30		Test6		TA100	+S9	Positive	Negative
31		Test6		TA98	+S9	Positive	Negative
32		Test7		TA100	+S9	Positive	Negative
33		Test7		TA98	+S9	Positive	Negative
34		Test7		TA98	-S9	Positive	Negative
35		Test8		TA100	+S9	Positive	Negative
36		Test8		TA100	-S9	Positive	Negative
37		Test8		TA98	+S9	Positive	Negative
38		Test9		TA100	+S9	Positive	Negative
39		Test9		TA98	+S9	Positive	Negative

Table 57. Summary of correspondence mismatches; Ames II assay versus standard, individual assessments. Equivocal calls not included in the analysis.

#	Chemical Name	Abbreviation	CAS Number	Bacterial Strain	S9	Standard Call	Ames II Call
1	2,4-Dinitrotoluene	24DNT	121-14-2	TA98	-S9	Negative	Positive
2	Benzo[a]pyrene	BAP	50-32-8	TA98	-S9	Negative	Positive
3	Cyclophosphamide	CPA	50-18-0	TA98	+S9	Negative	Positive
4	Hexamethylphosphoramide	HMP	680-31-9	TA98	+S9	Negative	Positive
5		L_17		TA98	+S9	Negative	Positive
6		L_17		TA98	-S9	Negative	Positive
7	Phenol	PHL	108-95-2	TA98	+S9	Negative	Positive
8	Tris(2-chloroethyl) phosphate	TCEP	115-96-8	TA98	+S9	Negative	Positive
1	2,4-Dinitrotoluene	24DNT	121-14-2	TA98	+S9	Positive	Negative
2	Glyoxal	GLY	107-22-2	TA98	-S9	Positive	Negative
3	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	ICR191	17070-45-0	TA98	+S9	Positive	Negative
4		L_19		TA98	+S9	Positive	Negative
5	Michler's ketone	MK	90-94-8	TA98	+S9	Positive	Negative

Table 58. Results of the test chemicals assessed in the 6-well plate assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold.

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	6-well 5 strains	Standard 2 strains	6-well 2 strains
3	2CP	2-Chloropyrimidine	1722-12-9	Negative	Negative	Negative	Negative
8	B2_001			Negative	Negative	Negative	Negative
9	B2_002			Negative	Negative	Negative	Negative
10	B2_003			Negative	Negative	Negative	Negative
11	B2_004			Negative	Negative	Negative	Negative
12	B2_005			Negative	Negative	Negative	Negative
13	B2_006			Negative	Negative	Negative	Negative
14	B2_007			Negative	Negative	Negative	Negative
15	B2_008			Negative	Negative	Negative	Negative
17	B2_010			Negative	Negative	Negative	Negative
30	CAP	Chloramphenicol	56-75-7	Negative	Negative	Negative	Negative
31	CH	Cyclohexanone	108-94-1	Negative	Negative	Negative	Negative
34	C_3			Negative	Negative	Negative	Negative
35	C_4			Negative	Negative	Negative	Negative
36	C_5			Negative	Negative	Negative	Negative
37	C_6			Negative	Negative	Negative	Negative
38	C_7			Negative	Negative	Negative	Negative
39	C_8			Negative	Negative	Negative	Negative
40	C_9			Negative	Negative	Negative	Negative
45	CmpdF			Negative	Negative	Negative	Negative
49	HQ	Hydroquinone	123-31-9	Negative	Negative	Negative	Negative
51	H_10			Negative	Negative	Negative	Negative

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	6-well 5 strains	Standard 2 strains	6-well 2 strains
55	H_14			Negative	Negative	Negative	Negative
57	H_16			Negative	Negative	Negative	Negative
58	H_17			Negative	Negative	Negative	Negative
60	H_19			Negative	Negative	Negative	Negative
61	H_20			Negative	Negative	Negative	Negative
62	H_21			Negative	Negative	Negative	Negative
68	H_7			Negative	Negative	Negative	Negative
72	LMETH	L-Methionine	63-68-3	Negative	Negative	Negative	Negative
76	N2_001			Negative	Negative	Negative	Negative
77	N2_002			Negative	Negative	Negative	Negative
78	N2_003			Negative	Negative	Negative	Negative
82	N2_007			Negative	Negative	Negative	Negative
83	N2_008			Negative	Negative	Negative	Negative
84	N2_009			Negative	Negative	Negative	Negative
86	N2_011			Negative	Negative	Negative	Negative
87	N2_012			Negative	Negative	Negative	Negative
89	N2_014			Negative	Negative	Negative	Negative
97	U_9			Negative	Negative	Negative	Negative
85	N2_010			Positive	Negative	Negative	Negative
7	AGE	Allyl glycidyl ether	106-92-3	Positive	Positive	Positive	Positive
19	B2_012			Positive	Positive	Positive	Positive
33	C_2			Positive	Positive	Positive	Positive
44	CmpdE			Positive	Positive	Positive	Positive
63	H_22			Positive	Positive	Positive	Positive

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	6-well 5 strains	Standard 2 strains	6-well 2 strains
71	ICR191	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	17070-45-0	Positive	Positive	Positive	Positive
74	MBAM	(S)-1-[(R)-alpha-Methylbenzyl]-2-aziridinemethanol	173143-73-2	Positive	Positive	Positive	Positive
80	N2_005			Positive	Positive	Positive	Positive
91	N2_016			Positive	Positive	Positive	Positive
92	NB	Nitrobenzene	98-95-3	Positive	Positive	Positive	Positive
94	SA	Sodium azide	26628-22-8	Positive	Positive	Positive	Positive
16	B2_009			Positive	Positive	Negative	Negative
20	BHAHCl	O-Benzylhydroxylamine HCl	2687-43-6	Positive	Positive	Negative	Negative
41	CmpdB			Positive	Positive	Negative	Negative
42	CmpdC			Positive	Positive	Negative	Negative
46	CmpdG			Positive	Positive	Negative	Negative
47	DIMET	Dimethoate	60-51-5	Positive	Positive	Negative	Negative
64	H_3			Positive	Positive	Negative	Negative
79	N2_004			Positive	Positive	Negative	Negative
88	N2_013			Positive	Positive	Negative	Negative
96	TBZ	tert-Butyl carbazate	870-46-2	Positive	Positive	Negative	Negative
4	2NAS	2-Naphthylamine-1-sulfonic acid	81-16-3			Negative	Negative
21	B_10					Negative	Negative
22	B_11					Negative	Negative
23	B_3					Negative	Negative
24	B_4					Negative	Negative
25	B_5					Negative	Negative

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	6-well 5 strains	Standard 2 strains	6-well 2 strains
26	B_6					Negative	Negative
27	B_7					Negative	Negative
28	B_8					Negative	Negative
29	B_9					Negative	Negative
43	CmpdD					Negative	Negative
50	H_1					Negative	Negative
52	H_11					Negative	Negative
53	H_12					Negative	Negative
54	H_13					Negative	Negative
56	H_15					Negative	Negative
59	H_18					Negative	Negative
65	H_4					Negative	Negative
66	H_5					Negative	Negative
67	H_6					Negative	Negative
69	H_8					Negative	Negative
70	H_9					Negative	Negative
73	MAN	D-Mannitol	69-65-8			Negative	Negative
75	MEN	DL-Menthol	1490-04-6			Negative	Negative
81	N2_006					Negative	Negative
95	TA1205-A					Negative	Negative
1	24DAT	2,4-Diaminotoluene	95-80-7			Positive	Positive
2	2AAF	2-Acetylaminofluorene	53-96-3			Positive	Positive
5	2NF	2-Nitrofluorene	607-57-8			Positive	Positive
6	6MP	6-Mercaptopurine	50-44-2			Positive	Positive

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	6-well 5 strains	Standard 2 strains	6-well 2 strains
18	B2_011					Positive	Positive
32	CPA	Cyclophosphamide	50-18-0			Positive	Positive
48	ENU	N-Ethyl-N-nitrosourea	759-73-9			Positive	Positive
90	N2_015					Positive	Positive
93	OQ	Olaquindox	23696-28-8			Positive	Positive

Table 59. Results of the test chemicals assessed in the 24-well plate assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold.

#	Abbreviation	Full Chemical Name	CAS Number	Standard 5 strains	24-well 5 strains	Standard 2 strains	24-well 2 strains
6	LMETH	L-Methionine	63-68-3	Negative	Negative	Negative	Negative
7	SA	Sodium azide	26628-22-8	Positive	Positive	Positive	Positive
1	2AA	2-Aminoanthracene	613-13-8			Positive	Positive
2	2NF	2-Nitrofluorene	607-57-8			Positive	Positive
3	4NQO	4-Nitroquinoline-N-oxide	56-57-5			Positive	Positive
4	BAP	Benzo[a]pyrene	50-32-8			Positive	Positive
5	ICR191	6-Chloro-9-[3-(2-chloroethylamino)propylamino]-2-methoxyacridine dihydrochloride	17070-45-0			Positive	Positive

Table 60. Results of the test chemicals assessed in the Ames MPF™ assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold.

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames MPF™ 5 strains	Standard 2 strains	Ames MPF™ 2 strains
1	2HBA	2-Hydroxybenzoic acid	69-72-7	Negative	Negative	Negative	Negative
9	CmpdQ1			Negative	Negative	Negative	Negative
10	CmpdQ2			Negative	Negative	Negative	Negative
11	CmpdQ3			Negative	Negative	Negative	Negative
12	CmpdQ4			Negative	Negative	Negative	Negative
18	EGDE	Ethylene glycol diethyl ether	629-14-1	Negative	Negative	Negative	Negative
27	RBFPS	Riboflavine-5-phosphate sodium	130-40-5	Negative	Negative	Negative	Negative

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames MPF™ 5 strains	Standard 2 strains	Ames MPF™ 2 strains
16	CmpdQ8			Positive	Negative	Negative	Negative
13	CmpdQ5			Positive	Positive	Positive	Positive
14	CmpdQ6			Positive	Positive	Positive	Positive
15	CmpdQ7			Positive	Positive	Positive	Positive
24	MPD	m-Phenylenediamine	108-45-2	Positive	Positive	Positive	Positive
28	SN	Sodium nitrite	7632-00-0	Positive	Positive	Positive	Positive
4	AMTr	Amitrol	61-82-5			Negative	Negative
5	AZT	Azidothymidine	30516-87-1			Negative	Negative
6	AmpT	Ampicillin trihydrate	7177-48-2			Negative	Negative
8	BenOH	Benzyl alcohol	100-51-6			Negative	Negative
19	EtiA	Ethionamide	536-33-4			Negative	Negative
20	IsoBut	Isobutyraldehyde	78-84-2			Negative	Negative
21	MAN	D-Mannitol	69-65-8			Negative	Negative
22	MELA	Melamine	108-78-1			Negative	Negative
23	MEN	DL-Menthol	1490-04-6			Negative	Negative
25	NBC	n-Butyl chloride	109-69-3			Negative	Negative
26	NOR	Norfloxacin	70458-96-7			Negative	Negative
30	SULX	Sulfisoxazole	127-69-5			Negative	Negative
32	Test1					Negative	Negative
33	Test12					Negative	Negative
34	Test16					Negative	Negative
35	Test17					Negative	Negative
36	Test18					Negative	Negative
37	Test19					Negative	Negative
39	Test22					Negative	Negative

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames MPF™ 5 strains	Standard 2 strains	Ames MPF™ 2 strains
42	Test5					Negative	Negative
47	oAA	o-Anthranilic acid	118-92-3			Negative	Negative
2	2NAS	2-Naphthylamine-1-sulfonic acid	81-16-3			Negative	Positive
29	STREP	Streptomycin sulfate	3810-74-0			Positive	Negative
43	Test6					Positive	Negative
45	Test8					Positive	Negative
46	Test9					Positive	Negative
3	AF2	2-(2-Furyl)-3-(5-nitro-2-furyl)acrylamide	3688-53-7			Positive	Positive
7	BAP	Benzo[a]pyrene	50-32-8			Positive	Positive
17	ECH	Epichlorohydrin	106-89-8			Positive	Positive
31	TCP	1,2,3-Trichloropropane	96-18-4			Positive	Positive
38	Test20					Positive	Positive
40	Test3					Positive	Positive
41	Test4					Positive	Positive
44	Test7					Positive	Positive

Table 61. Results of the test chemicals assessed in the Ames II assay and the standard assay with 5 or 2 strains (i.e., overall calls). Mismatches are indicated in bold.

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames II 5 strains	Standard 2 strains	Ames II 2 strains
6	HQ	Hydroquinone	123-31-9	Negative	Negative	Negative	Negative
7	L_1			Negative	Negative	Negative	Negative
8	L_10			Negative	Negative	Negative	Negative

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames II 5 strains	Standard 2 strains	Ames II 2 strains
9	L_11			Negative	Negative	Negative	Negative
10	L_12			Negative	Negative	Negative	Negative
11	L_13			Negative	Negative	Negative	Negative
12	L_14			Negative	Negative	Negative	Negative
13	L_15			Negative	Negative	Negative	Negative
14	L_16			Negative	Negative	Negative	Negative
15	L_2			Negative	Negative	Negative	Negative
16	L_3			Negative	Negative	Negative	Negative
17	L_4			Negative	Negative	Negative	Negative
18	L_5			Negative	Negative	Negative	Negative
19	L_6			Negative	Negative	Negative	Negative
20	L_7			Negative	Negative	Negative	Negative
21	L_8			Negative	Negative	Negative	Negative
22	L_9			Negative	Negative	Negative	Negative
25	U_1			Negative	Negative	Negative	Negative
26	U_11			Negative	Negative	Negative	Negative
27	U_3			Negative	Negative	Negative	Negative
28	U_4			Negative	Negative	Negative	Negative
29	U_5			Negative	Negative	Negative	Negative
30	U_6			Negative	Negative	Negative	Negative
31	U_9			Negative	Negative	Negative	Negative
24	SA	Sodium azide	26628-22-8	Positive	Negative	Positive	Negative
23	MCarb	Methyl carbamate	598-55-0			Negative	Negative
32	oAA	o-Anthranilic acid	118-92-3			Negative	Negative
2	AMTr	Amitrol	61-82-5			Negative	Positive

#	Abbreviation	Full Test Chemical Name	CAS Number	Standard 5 strains	Ames II 5 strains	Standard 2 strains	Ames II 2 strains
1	2AAF	2-Acetylaminofluorene	53-96-3			Positive	Positive
3	BAP	Benzo[a]pyrene	50-32-8			Positive	Positive
4	CPA	Cyclophosphamide	50-18-0			Positive	Positive
5	ENU	N-Ethyl-N-nitrosourea	759-73-9			Positive	Positive

Table 62. Overall summary of performance metrics and agreement statistics

Mini Assay	Individual Assessments		5-strain Overall	2-strain Overall		Pre-incubation Only	Plate Incorp. Only		TA98 Only	TA100 Only	TA1535 Only	TA1537 Only	AT Strains Only
	False – (%)		False – (%)	False – (%)		False – (%)	False – (%)		False – (%)	False – (%)	False – (%)	False – (%)	False – (%)
6-Well	7.3 (3.7-12.7)		4.5 (0.1-23.0)	0 (0.0-16.8)			7.5 (3.5-13.8)		4.0 (0.1-20.4)	2.9 (0.1-14.9)	14.7 (5.0-31.1)	11.4 (3.2-26.7)	3.4 (0.1-17.8)
Ames MPF™	26.7 (19.7-34.7)			23.5 (6.8-49.9)		16.5 (10.4-24.4)	38.3 (26.1-51.8)		28.2 (15.0-44.9)	32.4 (18.0-49.8)	30.0 (11.9-54.3)	24.3 (11.8-41.2)	24.0 (9.4-45.1)
Ames II	20.8 (7.1-42.2)			20.0 (0.5-71.6)		15.0 (3.2-37.9)	33.3 (4.3-77.7)		20.8 (7.1-42.2)				
24-Well	7.7 (2.9-16.0)						14.5 (6.9-25.8)			5.0 (0.1-24.9)	11.1 (0.3-48.2)	4.5 (0.1-22.8)	10.5 (1.3-33.1)
Mini Assay	**Individual Assessments**		**5-strain Overall**	**2-strain Overall**		**Pre-incubation Only**	**Plate Incorp. Only**		**TA98 Only**	**TA100 Only**	**TA1535 Only**	**TA1537 Only**	**AT Strains Only**
	False + (%)		False + (%)	False + (%)		False + (%)	False + (%)		False + (%)	False + (%)	False + (%)	False + (%)	False + (%)
6-Well	1.6 (0.8-2.7)		0 (0.0-8.8)	0 (0.0-4.7)			1.9 (1.1-3.2)		1.6 (0.3-4.5)	1.7 (0.3-4.8)	0 (0.0-2.9)	1.3 (0.2-4.6)	3.7 (1.4-7.8)
Ames MPF™	2.5 (1.3-4.2)			3.3 (0.08-17.2)		1.6 (0.7-3.3)	9.4 (4.9-15.8)		4.3 (1.6-9.1)	2.3 (0.5-6.7)	2.2 (0.5-6.4)	0.9 (0.0-4.7)	0 (0.0-7.1)
Ames II	5.4 (2.4-10.3)			3.7 (0.09-19.0)		5.9 (2.9-10.6)	5.9 (1.2-16.2)		5.4 (2.3-10.3)				
24-Well	6.8 (2.2-15.1)						8.1 (2.7-17.8)			6.2 (0.2-30.2)	9.1 (0.2-41.3)	4.5 (0.1-22.8)	12.9 (3.6-29.8)

Mini Assay	Individual Assessments		5-strain Overall	2-strain Overall		Pre-incubation Only	Plate Incorp. Only		TA98 Only	TA100 Only	TA1535 Only	TA1537 Only	AT Strains Only
	Sensitivity (%)		Sensitivity (%)	Sensitivity (%)		Sensitivity (%)	Sensitivity (%)		Sensitivity (%)	Sensitivity (%)	Sensitivity (%)	Sensitivity (%)	Sensitivity (%)
6-well	92.7 (87.3-96.3)		95.4 (77.2-99.9)	100 (83.2-100)			92.5 (86.2-96.5)		96 (79.6-99.9)	97.1 (85.1-99.9)	85.3 (68.9-95.0)	88.6 (73.3-96.8)	96.6 (82.2-99.9)
Ames MPF™	73.3 (65.3-80.3)			76.5 (50.1-93.2)		83.5 (75.6-89.6)	61.7 (48.2-73.9)		71.8 (55.1-85.0)	67.6 (50.2-82.0)	70 (45.7-88.1)	75.7 (58.8-88.2)	76 (54.9-90.6)
Ames II	79.2 (57.8-92.9)			80.0 (28.4-99.5)		85.0 (62.1-96.8)	66.7 (22.3-95.7)		79.2 (57.8-92.9)				
24-well	92.3 (84.0-97.1)						85.5 (74.2-93.1)			95.0 (75.1-99.9)	88.9 (51.8-99.7)	95.4 (77.2-99.9)	89.5 (66.9-98.7)
Mini Assay	**Individual Assessments**		**5-strain Overall**	**2-strain Overall**		**Pre-incubation Only**	**Plate Incorp. Only**		**TA98 Only**	**TA100 Only**	**TA1535 Only**	**TA1537 Only**	**AT Strains Only**
	Specificity (%)		Specificity (%)	Specificity (%)		Specificity (%)	Specificity (%)		Specificity (%)	Specificity (%)	Specificity (%)	Specificity (%)	Specificity (%)
6-well	98.4 (97.3-99.2)		100 (91.2-100)	100 (95.3-100)			98.1 (96.8-98.9)		98.4 (95.5-99.7)	98.3 (95.2-99.7)	100 (97.1-100)	98.7 (95.4-99.8)	96.3 (92.2-98.6)
Ames MPF™	97.5 (95.8-98.7)			96.7 (82.8-99.9)		98.4 (96.7-99.3)	90.6 (84.2-95.1)		95.7 (90.9-98.4)	97.7 (93.3-99.5)	97.8 (93.6-99.5)	99.1 (95.3-100)	100 (92.9-100)
Ames II	94.6 (89.7-97.6)			96.3 (81.0-99.9)		94.1 (89.4-97.1)	94.1 (83.8-98.8)		94.6 (89.7-97.7)				
24-well	93.2 (84.9-97.8)						91.9 (82.2-97.3)			93.8 (69.8-99.8)	90.9 (58.7-99.8)	95.4 (77.2-99.9)	87.1 (70.2-96.4)

Mini Assay	Individual Assessments		5-strain Overall	2-strain Overall		Pre-incubation Only	Plate Incorp. Only		TA98 Only	TA100 Only	TA1535 Only	TA1537 Only	AT Strains Only
	PPV (%)[1]		PPV (%)[1]	PPV (%)[1]		PPV (%)[1]	PPV (%)[1]		PPV (%)[1]	PPV (%)[1]	PPV (%)[1]	PPV (%)[1]	PPV (%)[1]
6-well	92.1 (86.5-95.8)		100 (83.9-100)	100 (83.2-100)			88.8 (81.9-93.7)		88.9 (70.8-97.6)	91.9 (78.1-98.3)	100 (0.88-100)	93.9 (79.8-99.3)	82.4 (65.5-93.2)
Ames MPF™	89.2 (82.2-94.1)			92.9 (66.1-99.8)		93.5 (87.1-97.4)	75.5 (61.1-86.7)		82.4 (65.5-93.2)	89.3 (71.8-97.7)	82.4 (56.6-96.2)	96.6 (82.2-99.9)	100 (82.4-100)
Ames II	70.4 (49.8-86.2)			80.0 (28.4-99.5)		63.0 (42.4-80.6)	57.1 (18.4-90.1)		70.4 (49.8-86.2)				
24-well	93.5 (85.5-97.9)						91.4 (81.0-97.1)			95.0 (75.1-99.9)	88.9 (51.8-99.7)	95.4 (77.2-99.9)	81.0 (58.1-94.6)
Mini Assay	Individual Assessments		5-strain Overall	2-strain Overall		Pre-incubation Only	Plate Incorp. Only		TA98 Only	TA100 Only	TA1535 Only	TA1537 Only	AT Strains Only
	NPV (%)[2]		NPV (%)[2]	NPV (%)[2]		NPV (%)[2]	NPV (%)[2]		NPV (%)[2]	NPV (%)[2]	NPV (%)[2]	NPV (%)[2]	NPV (%)[2]
6-well	98.6 (97.4-99.3)		97.6 (87.1-99.9)	100 (95.3-100)			98.7 (97.6-99.4)		99.5 (97.1-100)	99.4 (96.9-100)	96.2 (91.3-98.7)	97.4 (93.6-99.3)	99.4 (96.5-100)
Ames MPF™	93.0 (90.5-94.9)			87.9 (71.8-96.6)		95.5 (93.1-97.2)	83.5 (76.2-89.2)		92.4 (86.8-96.2)	91.2 (85.2-95.4)	95.6 (90.7-98.4)	92.7 (86.7-96.6)	89.3 (78.1-96.0)
Ames II	96.6 (92.2-98.9)			96.3 (81.0-99.9)		98.1 (94.7-99.6)	96.0 (86.3-99.5)		96.6 (92.2-98.9)				
24-well	92.0 (83.4-97.0)						86.4 (75.7-93.6)			93.8 (69.8-99.8)	90.9 (58.7-99.8)	95.4 (77.2-99.9)	93.1 (77.2-99.2)

Mini Assay	Individual Assessments		5-strain Overall	2-strain Overall		Pre-incubation Only	Plate Incorp. Only		TA98 Only	TA100 Only	TA1535 Only	TA1537 Only	AT Strains Only
	PABAK		**PABAK**	**PABAK**		**PABAK**	**PABAK**		**PABAK**	**PABAK**	**PABAK**	**PABAK**	**PABAK**
6-Well	0.95 (0.92-0.97)		0.97 (0.83-1.00)	1.0 (0.92-1.00)			0.94 (0.92-0.96)		0.96 (0.91-0.99)	0.96 (0.91-0.99)	0.94 (0.86-0.98)	0.94 (0.86-0.98)	0.93 (0.85-0.97)
Ames MPF™	0.85 (0.80-0.88)			0.79 (0.54-0.93)		0.9 (0.86-0.94)	0.63 (0.50-0.73)		0.81 (0.71-0.89)	0.82 (0.71-0.90)	0.88 (0.78-0.95)	0.87 (0.77-0.94)	0.84 (0.67-0.94)
Ames II	0.85 (0.75-0.92)			0.88 (0.58-0.98)		0.86 (0.77-0.93)	0.82 (0.61-0.94)		0.85 (0.75-0.92)				
24-Well	0.86 (0.75-0.93)						0.77 (0.64-0.87)			0.89 (0.63-0.99)	0.80 (0.37-0.98)	0.91 (0.69-0.99)	0.76 (0.51-0.91)

[1]Positive Predictive Value
[2]Negative Predictive Value