Mutagenesis, 2015, 00, 1–9
doi:10.1093/mutage/gev085
Original Manuscript

OXFORD

Original Manuscript

Empirical analysis of BMD metrics in genetic toxicology part I: *in vitro* analyses to provide robust potency rankings and support MOA determinations

John W. Wills[1,†], George E. Johnson[2,*,†,‡], Shareen H. Doak[2], Lya G. Soeteman-Hernández[3], Wout Slob[3] and Paul A. White[1,*,‡]

[1]Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada, [2]Institute of Life Science, Swansea University Medical School, Swansea, UK and [3]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

*To whom correspondence should be addressed. Tel: +44 (0) 1792 295158; Fax: +44 (0) 1792 513100; Email: g.johnson@swansea.ac.uk; Tel: +1 613 941 7373; Fax: +1 613 941 8530; Email: paul.white@hc-sc.gc.ca

†These authors are joint first authors.

‡These authors are joint corresponding authors.

Received 9 July 2015; Revised 17 November 2015; Accepted 24 November 2015.

Abstract

Genetic toxicity testing has traditionally been used for hazard identification, with dichotomous classification of test results serving to identify genotoxic agents. However, the utility of genotoxicity data can be augmented by employing dose–response analysis and point of departure determination. Via interpolation from a fitted dose–response model, the benchmark dose (BMD) approach estimates the dose that elicits a specified (small) effect size. BMD metrics and their confidence intervals can be used for compound potency ranking within an endpoint, as well as potency comparisons across other factors such as cell line or exposure duration. A recently developed computational method, the BMD covariate approach, permits combined analysis of multiple dose–response data sets that are differentiated by covariates such as compound, cell type or exposure regime. The approach provides increased BMD precision for effective potency rankings across compounds and other covariates that pertain to a hypothesised mode of action (MOA). To illustrate these applications, the covariate approach was applied to the analysis of published *in vitro* micronucleus frequency dose–response data for ionising radiations, a set of aneugens, two mutagenic azo compounds and a topoisomerase II inhibitor. The ionising radiation results show that the precision of BMD estimates can be improved by employing the covariate method. The aneugen analysis provided potency groupings based on the BMD confidence intervals, and analyses of azo compound data from cells lines with differing metabolic capacity confirmed the influence of endogenous metabolism on genotoxic potency. This work, which is the first of a two-part series, shows that BMD-derived potency rankings can be employed to support MOA evaluations as well as facilitate read across to expedite chemical evaluations and regulatory decision-making. The follow-up (Part II) employs the combined covariate approach to analyse *in vivo* genetic toxicity dose–response data focussing on how improvements in BMD precision can impact the reduction and refinement of animal use in toxicological research.

Introduction

Genetic toxicology testing plays an essential role in the safety assessment of new and existing compounds, with the aim to ensure that the risk of adverse human health effects mediated by genetic damage are minimised. Genetic toxicity testing has traditionally been used only for hazard identification, with dichotomous groupings (i.e. positives *versus* negatives) being used to separate the genotoxic agents from those compounds that are unable to elicit a significant positive response in selected *in vivo* and/or *in vitro* assays. However, there is increasing recognition of the limitations of the qualitative paradigm currently employed for genetic toxicity assessment, and moreover, the need to develop a quantitative framework for the analysis and interpretation of genotoxicity test results (1, 2). An alternative, quantitative paradigm necessitates dose–response analysis and the determination of point of departure (PoD) metrics that can be used to determine margins of exposure and/or exposure limits below which the risk of adverse effect is considered to be acceptable.

To facilitate the transition from qualitative to quantitative analyses, there is an urgent need to develop rigorous methods for the analysis of genetic toxicity dose–response data, and determine PoD values that can subsequently be reviewed, interpreted, compared and/or ranked (1–3). Recent publications, including those by Gollapudi *et al.* (4) and Johnson *et al.* (5), have considered and evaluated numerous approaches for quantitative analysis of genetic toxicity dose–response data and the determination of PoD metrics (1, 2, 4, 5). A major conclusion by these authors, as well as MacGregor *et al.*, (1) is that there are distinct advantages to the benchmark dose (BMD) approach that employs computational algorithms to fit mathematical functions to dose–response data (6, 7) and subsequently interpolate the dose that corresponds to a predefined increase in response above the control (e.g. 10%). Figure 1 illustrates the BMD concept and how the range delineated between the upper (BMDU) and lower (BMDL) confidence bounds defines the precision of the estimated BMD.

Recent work conducted by researchers at the Dutch National Institute of Public Health and the Environment (RIVM) has shown that appropriate use of BMD estimates and their confidence intervals has applications for compound potency ranking within an endpoint, as well as empirical potency comparisons across endpoints (8–11). Furthermore, novel computational algorithms developed at the RIVM (e.g. the BMD covariate approach) permit combined analysis of multiple dose–response data sets for a shared endpoint. These algorithms permit multiple BMDs to be defined for dose–response relationships differentiated by a covariate (i.e. factor defining the subgroups) included in the analysis (e.g. compound, tissue, cell type, sex, exposure duration/regime, genotype, etc.), and importantly, have the potential to yield more precise BMD estimates in instances where the shapes of the normalised dose–response curves are the same at each covariate level (12).

Several of the aforementioned researchers have proposed the use of quantitative analyses of *in vivo* genetic toxicity dose–response data to determine PoD metrics (e.g. BMDL) that are in turn used to establish human exposure limits for regulatory decision-making (e.g. tolerable daily intake, permissible daily exposure (1–3, 5)). In contrast *in vitro* data are, at least for the time being, better suited to substance ranking and/or mode of action (MOA) determination. More specifically, the aforementioned MacGregor *et al.* publications, which were prepared under the auspices of the International Workgroup on Genetic Toxicology (IWGT), note that interpretation of *in vitro* results in a human health risk assessment context is hindered by, among other things, issues pertaining to toxicokinetics and *in vivo* tissue-specific metabolism (1, 2). Advanced computational methods (i.e. *in vitro* to *in vivo* extrapolation or IVIVE) will be required to reliably use *in vitro* dose–response data to determine human exposure limits (13, 14), and although such methods are currently under development, further refinement and validation will be required (15, 16).

In this work, which is Part I of a two-part series, we use previously published data for the *in vitro* micronucleus (MN) frequency endpoint to examine the utility of the BMD covariate method to increase the precision of BMD estimates and permit statistically rigorous examinations of relative potencies across compounds and/or other covariates that pertain to MOA. The former is illustrated by comparative analysis of a series of radiation exposures, whilst the latter employs examination of dose–response data for a set of mitotic spindle poisons (i.e. aneugens), as well as metabolically activated DNA-reactive substances and DNA replication inhibitors. The selected data sets represent responses for well-studied agents where the MOA underlying the dose–responses has previously been investigated and characterised (17). The follow-up publication (i.e. Part II) expands on the use of the BMD covariate method to examine study reproducibility, sex-specific differences and compound MOA *in vivo*; focussing on how improvements in BMD precision can impact the reduction and refinement of animal use in toxicological assessments and research.

Materials and methods

In vitro MN frequency dose–response data sets were obtained from the published literature. The collected data were then subjected to combined, covariate BMD analyses through combination of dose–response relationships for a series of compounds or other covariates, with study types/protocols matched as far as possible within an analysis (i.e. same laboratory or at least analogous protocols). The radiation data were collected in human lymphocytes (HL) (18). The MN data for well-characterised aneugens were combined from studies using Chinese hamster ovary (CHO-K1) cells, primary human lymphocyte (HL) cells and the human lymphoblastoid cell line AHH-1 (19–23). MN data for the clastogenic azo dyes Sudan-1 and Para Red were generated using AHH-1 cells and the transgenic daughter cell line MCL-5 that expresses five human drug metabolising enzymes (24). Finally, the DNA replication inhibition data in mouse lymphoma L5178Y cells were collected following time-dependent exposures to the clastogen etoposide (25).

PROAST version 50.9 was used to conduct the dose–response analyses (http://www.proast.nl). As necessary, dose–response data were analysed using one (exponential) or both (exponential and the Hill) nested model families that have been recommended by the European Food Safety Authority (EFSA) for the analysis of continuous data (26). For each analysis, combined data sets were analysed using the factor discriminating the data sets (e.g. study, compound, cell type) as a covariate. PROAST uses the likelihood ratio test to assess whether inclusion of additional parameters resulted in a statistically significant improvement in model fit (9, 10). Models with additional parameters are only accepted if the difference in log-likelihood exceeds the critical value at $P < 0.05$ (12). In this way, it can be established which model parameters need to be estimated for each subgroup, and which parameters may be considered as constant among the subgroups of a combined data set. In general, it was assumed that the maximum response (parameter *c*) and log-steepness (parameter *d*) (i.e. the two shape parameters) were equal for all response curves, while parameters for background response (parameter *a*), potency (parameter *b*) and *var* (i.e. within group variation) were examined for being covariate dependent (12). PROAST outputs designate potency (i.e. the BMD) as CED or critical effect dose, and the metrics BMDL and BMDU are designated CEDL and CEDU, respectively. The CED (two-sided) 90% confidence interval (CEDL, CEDU) is calculated for each level of the covariate (if they

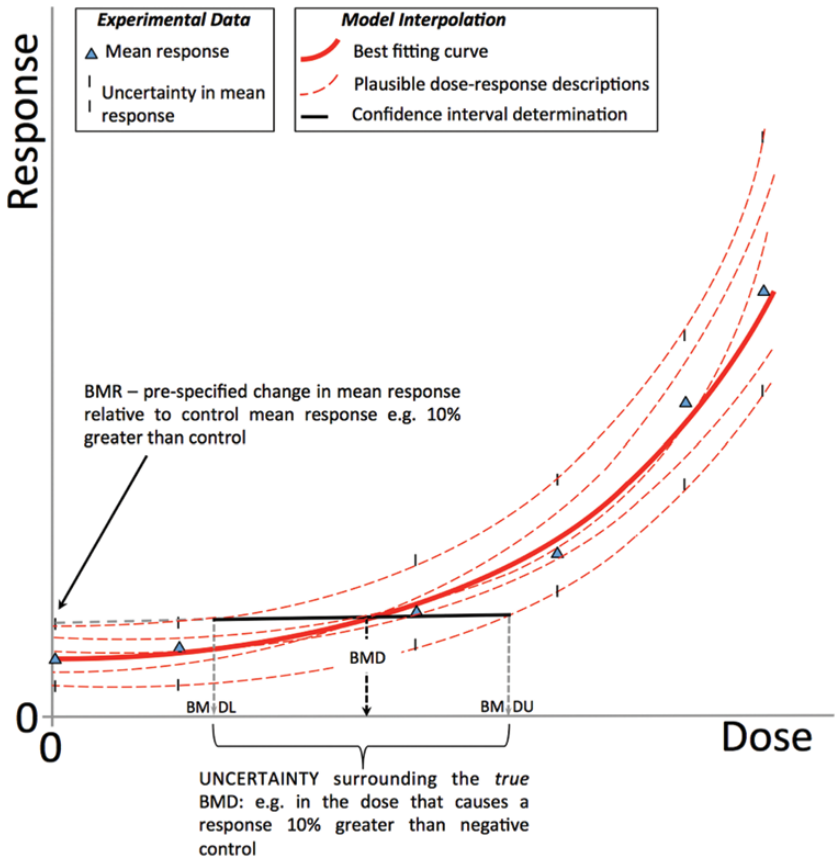

Figure 1. Schematic representation of the benchmark dose (BMD) approach for analysing dose–response data. The BMD is an estimate of the dose that will elicit the benchmark response (BMR), and is estimated by interpolation from the fitted curve. The BMR is usually defined as a percentage increase in response (e.g. 10%) relative to control; with this BMR adjustable to any desired response level. The uncertainties in the data can be taken into account by calculating a confidence interval for the BMD. Conceptually, one may imagine that, by varying the parameters in the model, different curves can be generated, and for those that are considered compatible with the data (dashed curves) BMDs could be established. Together, they comprise values that make up the confidence interval for the BMD. As an approximate conceptual illustration, the horizontal black line segment intersecting the range of plausible curves results in the BMDL and BMDU, the lower and upper 90% confidence bounds of the BMD estimate, respectively. The width of this interval (expressed as the ratio BMDU to BMDL) therefore represents the BMD precision.

are found to significantly differ). Fits of the model to the data sets of each subgroup are presented in supplementary Figures S1–S4, available at *Mutagenesis* Online, and were used to visually evaluate the validity of the assumed constant shape parameters. This approach was preferred over evaluating the assumption by statistical testing, since statistical tests on the shape parameters are highly sensitive to non-random errors in the data that are ubiquitous in experimental data, and the effect of which may even be amplified by leverage effects in dose–response data (12). Furthermore, minor non-random errors in the data might lead to rejection of the constancy of the shape parameter assumption (i.e. given the relatively high power in a combined data set), while small differences among the shape parameters would only have a small impact on the coverage of the BMD confidence interval (12). Visual inspection of the fitted curves was therefore considered a better way to determine whether any differences in parameters c and d between covariates were small enough to be ignored.

The benchmark response (BMR), also known as critical effect size (in PROAST notation), employed in the presented analyses was set at various values depending on the data available. This is justified since the aim of the analyses was to examine differences in potency rather than derive a PoD for risk assessment. Thus, the BMR can be adjusted to higher values in situations where a BMR such as 10% is deemed to be low compared to the magnitude of the observed effects in the data set, and choosing a higher value may have a favourable effect on the width of the BMD confidence intervals. The BMDL and BMDU values represent the lower and upper bounds of the two-sided 90% confidence interval of the BMD (27), with the BMDU–BMDL ratio defining the width of the confidence interval and therefore its precision. Confidence interval plots, arranged using the geometric midpoint of the BMDL–BMDU interval were employed to visually compare potencies across levels of examined covariates while taking estimation uncertainty into account (28).

Results and discussion

Use of combined BMD-covariate analyses to increase BMD precision

In their 2014 review, Slob and Setzer (12) showed that fitting a dose–response model to combined data sets related to the same endpoint/study type has the potential to yield more precise BMD estimates (i.e. the BMDU:BMDL ratio is reduced). This increase in precision is a consequence of conserved dose–response shape after correcting each constituent dose–response for background (i.e. y-axis scaling, making parameter a covariate dependent)

and potency (i.e. *x*-axis scaling, making parameter *b* covariate dependent). Therefore, after scaling, the individual dose–response curves included in a combined analysis may be regarded as having highly similar shapes if they can be adequately described using constant model values for the maximum response parameter (*c*) and the log-steepness parameter (*d*). In turn, this process improves the precision of the resulting BMD estimates as all dose–response data sets included in the combined analysis contribute information on the common shape parameters in the dose–response model (12).

To demonstrate this principle, the 'combined covariate' analysis approach was used to observe whether increased BMD precision could be achieved for *in vitro* analyses of neutron, X- and gamma-ray radiation exposures of HL (18). For each of the exposure subgroups, the use of conserved exponential model shape parameters resulted in acceptable model description of the dose–response data (supplementary Figure S1, available at *Mutagenesis* Online). The data were also subjected to independent BMD analyses carried out individually for each of the constituent dose–response data sets. As Figure 2 shows, the combined analysis assuming conserved dose–response shape resulted in substantially better precision in the BMD estimates, indicated by the smaller BMD confidence intervals. The plotted confidence intervals, here ordered by log-midpoint, provide a visually intuitive way to assess BMD precision. At the same time, they may be used to assess differences in potency across the subgroups included in an analysis (Figures 3–5). Importantly, as confidence intervals should represent the range in which the true BMD lies, alleged potency differences are only statistically defensible when there is no apparent overlap between intervals. The improvement in BMD precision resulting from the use of the combined, covariate analysis is thus extremely useful for effective BMD comparisons. The narrower confidence intervals resulting from the covariate approach confers an improved ability to observe existing differences in BMD values across levels of a covariate (i.e. lower probability of a Type II error).

Use of combined BMD-covariate analysis to permit robust potency ranking of genotoxic substances with the same MOA

In 2002 the UK Committee on Toxicity put forward a proposal (29), later reviewed by the Committee on Mutagenicity (30), to conduct combined risk assessments for selected genotoxic substances that act via a common MOA. In this proposal, the benzimidazoles were used as archetypal compounds since they are known to cause aneugenic effects through an MOA that involves mitotic spindle polymerisation (i.e. mitotic spindle poisons (31, 32)). This type of multi-substance risk assessment, which has also been conducted for substances such as chlorinated dibenzodioxins and polycyclic aromatic hydrocarbons, requires quantitative information on relative potency for the endpoint under consideration to generate equivalency factors [i.e. relative potency factors (RPFs), potency equivalence factors (PEFs), toxicity equivalence factors (TEFs)] that scale the potency of each compound to one of the compounds in that group, used as a reference compound (33–35). Here we demonstrate that the BMD approach employing compound as covariate can provide a rigorous potency ranking for a series of aneugens examined *in vitro*. The results are presented for the exponential model only since the Hill model resulted in virtually identical BMD confidence intervals.

The combined aneugen analysis yielded finite confidence intervals for all compounds except rotenone, diethylstilboestrol and thiabendazole (Figure 3), for which the underlying data were insufficient to determine an upper BMD confidence bound. Infinite BMD upper bounds indicate that there might not be any dose-related response for these chemicals (see supplementary Figure S2, available at *Mutagenesis* Online for the dose–response data of each compound and the fitted model). In these instances, the fact that the lower BMD confidence bounds could be calculated highlights one of the benefits of the BMD approach. More specifically, although the dose–response trend is non-significant, which is an inconclusive result, the BMDL provides conclusive information. It indicates that, *if* there is in reality a dose-related effect (e.g. at elevated doses compared to those examined), effects larger than the BMR (i.e. here, 10%) will most likely

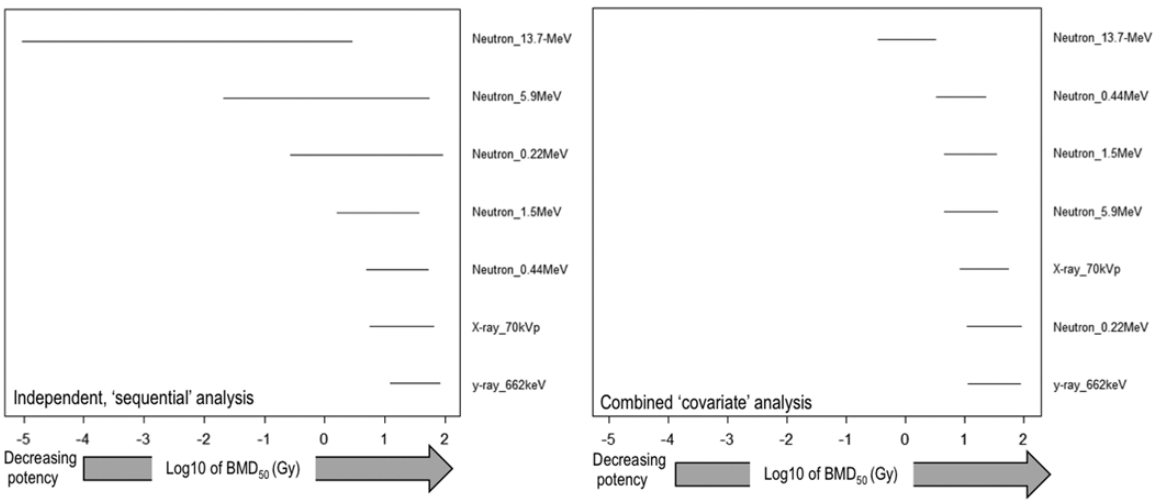

Figure 2. BMD (two-sided) 90% confidence intervals of the BMD_{50} calculated from the fitted exponential model for MN frequency responses in human lymphocytes exposed to low doses of ionising radiation (18). The left panel illustrates the results obtained when each of the seven dose–response data sets were analysed independently and sequentially. The right panel illustrates the results obtained when the data sets were combined and analysed using the covariate 'radiation type'. For the combined analysis, parameters for maximum response (*c*) and log-steepness (*d*) were assumed to be equal for all response functions, while parameters for background response (*a*), potency (*b*) and *var* (within group variation) were covariate dependent. The right panel shows that the combined covariate analysis yields increased BMD precision (i.e. narrower confidence intervals) for all estimates. The underlying dose–response data and fitted model curves are shown supplementary Figure S1, available at *Mutagenesis* Online.

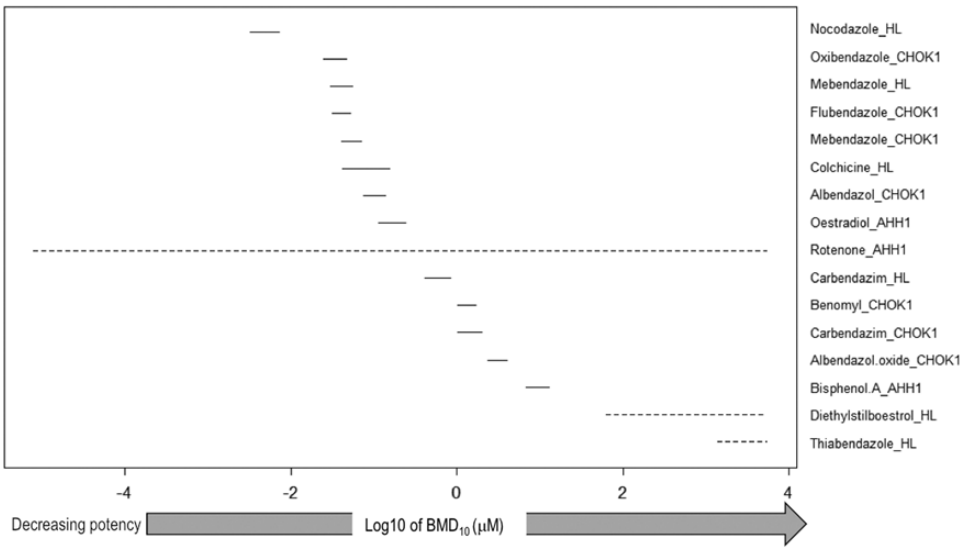

Figure 3. Illustration of the BMD covariate approach to rank the potency of selected aneugens. The panel shows the (two-sided) 90% confidence intervals for the BMD_{10} for each substance based on exponential model covariate BMD analyses of MN dose–response data. Overlapping confidence intervals mean that BMDs cannot be distinguished due to the uncertainties in the underlying dose–response data. Flubendazole, oxibendazole, mebendazole, albendazole, benomyl, carbendazim and albendazole oxide were tested in Chinese hamster ovary (CHO-K1) cells (36). Nocodazole, mebendazole, colchicine, carbendazim, diethylstilboestrol and thiabendazole were tested in human lymphocytes (HL) (19–21). Oestradiol, rotenone and bisphenol-A were tested in human lymphoblastoid (AHH-1) cells (22, 23). Individual dose–response data and fitted model curves are shown in supplementary Figure S2, available at *Mutagenesis* Online.

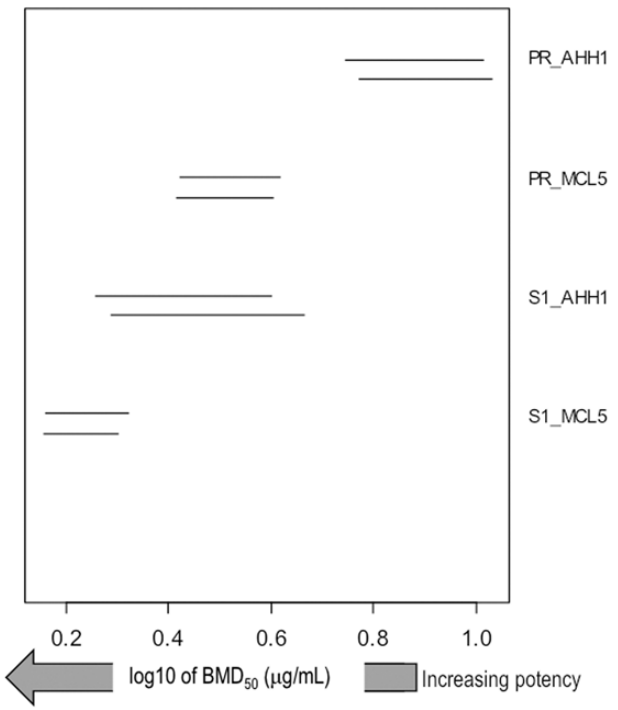

Figure 4. BMD (two-sided) 90% confidence intervals for the BMD_{50} (BMR = 50%) resulting from a covariate analysis of micronucleus dose–response data for the azo dyes Sudan-1 (S1) and Para Red (PR) in human lymphoblastoid cell lines MCL-5 and AHH-1 (24). For each subgroup, the upper interval relates to the fitted exponential model, the lower to the Hill model. The underlying dose–response data and fitted model curves are shown in supplementary Figure S3, available at *Mutagenesis* Online.

occur at doses higher than the established BMDL. Furthermore, a BMD confidence interval indicates how much information regarding the BMD is provided by the dose–response data. For most of the aneugens analysed, the BMD confidence intervals spanned a factor of ~2, showing that the information in the dose–response data examined was quite good.

The results of this potency ranking can be subsequently scrutinised in relation to chemical structure and the proposed mechanism(s) of action. For example, oxibendazole, flubendazole and mebendazole, which have large moieties in the 5-position that are absent in benomyl and carbendazim, are benzimidazole derivatives known to competitively inhibit colchicine binding (31, 32, 37). These moieties might account for the higher potency observed for the former three compounds relative to the latter two, which are known to bind to a different site than that of colchicine (37) but still act by inhibiting microtubule assembly and microtubule depolymerisation (38). Diethylstilboestrol also acts at the colchicine-binding site but in a non-specific way (39) and the structurally related compound bisphenol-A promotes spindle polymerisation through stabilisation of the microtubule (40), suggesting possible reasons for their reduced potency. Although albendazole oxide and thiabendazole act at the colchicine-binding site in a similar way to the other benzimidazoles (36), they were also determined to have lower potency, emphasising the need to consider both mechanism of action and potency for successful multi-substance assessments.

Interestingly, the choice of CHO-K1 cells or HL did not appear to substantially impact the aforementioned potency estimates, as illustrated by the results for mebendazole and carbendazim, which were studied in both cell systems. For mebendazole, the small BMD confidence intervals largely overlapped implying the difference can only be small. For carbendazim they did not overlap, but as the results in Figure 3 indicate, the difference in BMDs between cell systems could be very small, or at most differ by a factor of ~2.5 (0.4 log units). Here it is worth noting that replication error in the same cell system (and compound) might affect the potency estimates since the study replication error is currently unknown (i.e. only single replicate data set available).

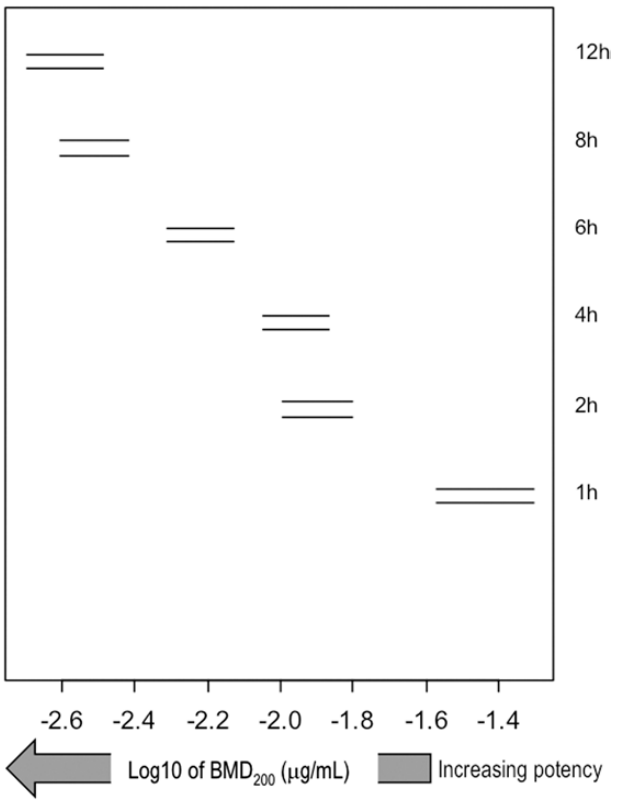

Figure 5. BMD (two-sided) 90% confidence intervals for the BMD_{200} (BMR = 200%) resulting from a covariate analysis of micronucleus dose–response data for the DNA replication inhibitor etoposide in mouse L5178Y lymphoma cells (25). For each exposure duration the upper confidence interval relates to the fitted exponential model, the lower to the Hill model. The underlying dose–responses data and fitted model curves are shown in supplementary Figure S4, available at *Mutagenesis* Online

The plotted confidence intervals (Figure 3) also permit comprehensive potency ranking across the complete family of similarly acting aneugens, revealing that nocodazole is most potent, with bisphenol-A, diethylstilboestrol and thiabendazole, also *potentially* rotenone, being least potent (note that for the compounds with infinite BMD upper bounds, potency cannot be concluded to be zero). This information could also readily be used to generate hypotheses related to the mechanistic phenomena that underlie the observed potency differences. For example, one might hypothesise that the observed differences relate to toxicokinetic variations across cell types (i.e. differences in uptake and elimination rates), differences in compound-specific metabolic requirements and metabolic capacity across cell types (i.e. rate of formation and destruction of reactive metabolites), differences in repair capacity across cell types and differences in the nature and consequence of toxicodynamic phenomena at the site of action (i.e. affinity of compound or reactive metabolites for the site of action) (2, 41–43).

The aforementioned potency rankings derived from BMD covariate analyses might be more generally employed to provide a link between substance potency and chemical structure or physical–chemical properties. Although similar analyses with more compounds would be needed to substantiate links between potency and compound properties, it is reasonable to assert that this type of analysis could provide rationales for assigning substances to sub-groupings based on evidence of potency similarity. Chemical categorisation and sub-grouping could, in turn, contribute to data gap filling via read across, trend analysis and quantitative structure-activity relationship (QSAR) modelling. Interestingly, the Organisation for Economic Cooperation and Development (OECD), the European Chemicals Agency (ECHA) and the European Commission Join Research Centre (JRC) recently published guidance documents and case study reports on the importance of establishing chemical categories to enhance hazard/risk assessment and regulatory decision-making (44–49). Categorisation permits the establishment of substance groups and/or subgroups whereby 'physicochemical and human health and/or ecotoxicological properties and/or environmental fate are likely to be similar or follow a regular pattern as a result of structural similarity' (44). Where sufficient justification can be found for the formation of categories, regulatory agencies such as those listed advocate the use of read across, either qualitative or quantitative, to fill data gaps (i.e. data for tested compounds are applied to similar untested compounds and/or analogous endpoints for the same compounds). For establishing empirically defined potency groups we therefore suggest the combined covariate BMD approach presented herein is an appropriate tool since it is an effective means of assessing and ranking chemical potencies.

Use of combined BMD-covariate analysis for potency comparisons between cell lines with different metabolic competencies

In addition to establishing robust potency rankings, the BMD covariate approach might also provide, as already suggested, information to support mechanistic hypotheses. To demonstrate this, *in vitro* MN data for the azo dyes Sudan-1 and Para Red (24) were used to compare responses across two cell lines that differ with respect to endogenous metabolic capacity. These substances and their metabolites are known to generate DNA adducts through reactive oxygen species (e.g. 8-oxo-7,8-dihydro-2′-deoxyguanosine or 8-oxo-dG) and bulky adducts that form via a reactive intermediate [i.e. benzene diazonium ion (BDI) adducts] (50). Thus, the metabolic capacity of an *in vitro* test system is an important consideration in the determination of genotoxic potency for these compounds. The dose–response analyses conducted herein used compound and cell line as covariates to examine relative potency and the influence of each cell lines' endogenous metabolic capacity (Figure 4). The analyses were highly similar using either exponential or Hill models, and revealed that both Para Red and Sudan-1 are more potent in the transgenic MCL-5 cell line, which expresses five human drug metabolising enzymes (i.e. CYP1A1, 1A2, 2A6, 2E1 and 3A4) that are absent in the parent AHH-1 cell line. This indicates that endogenous metabolism is an important determinant of genotoxic potency for these compounds. Further, the analysis showed significant differences in potency between the two azo dyes within cell lines. While the paucity of information on the metabolism of the structurally similar compounds prohibits any definitive statements regarding the cause of the observed potency pattern, the differences in potency between Para Red and Sudan-1 likely reflect differential formation of DNA-reactive metabolites such as BDI, as well as reactive oxygen species (24, 51, 52).

Use of combined BMD-covariate analysis to examine the impact of exposure duration

Disruption of DNA replication has been shown to induce chromosomal damage, with one of the best examples provided by Lynch *et al.* (25) who examined the effect of topoisomerase II inhibition on the clastogenicity (i.e. MN frequency) of etoposide *in vitro*. One of the main findings of the Lynch *et al.* study was that etoposide-induced

clastogenicity was both concentration- and time-dependent; however, the authors were not able to successfully employ bilinear dose–response modelling to evaluate their time dependency hypothesis since the dose–response data did not follow a bilinear dose–response relationship.

In contrast, our BMD reanalysis of the Lynch *et al*. etoposide data shows that the dose–response data, which are suitable for BMD analysis, accurately describe the effect of exposure duration. Using the combined BMD approach with exposure duration as the covariate, the exponential and Hill BMD confidence intervals showed a consistent decrease with exposure time (Figure 5), supporting the assertion of increasing potency for increased exposure duration. This analysis substantiates the work of Lynch *et al*. (25) by providing a quantitative analysis confirming the expected time- and dose-dependency of MN frequency.

Conclusions

The analyses presented herein show that *in vitro* data can be used for more than hazard identification and qualitative binning of genotoxicity test results. More specifically, they illustrate that quantitative dose–response analyses, and moreover, considerations of BMD confidence intervals, in particular when established by combined covariate dose–response analyses, permit robust potency determinations and potency rankings. In addition, they illustrate that rigorous comparisons of potency across compound-cell combinations can provide support for MOA hypotheses. Finally, they illustrate that BMD ranking would be the appropriate tool for developing substance categories and/or sub-categories, which in turn can be used to facilitate quantitative read across to compensate for data gaps encountered during regulatory reviews of chemicals in commerce. When sufficiently validated, the various applications illustrated here could also be employed in the development of other chemical assessment tools and procedures based on *in vitro* test results; e.g. establishment and scrutiny of key event relationships in AOPs (adverse outcome pathways (53)), and, in combination with IVIVE approaches (42), the establishment of human exposure limits.

We offer the following concluding statements regarding the utility of the combined covariate approach presented here, and more generally, comparison of *in vitro* BMDs:

1. Quantitative dose–response analysis of *in vitro* MN data using the BMD approach provides quantitative information on the potency of a compound, which is more informative than just concluding whether a given compound has genotoxic potential or not.
2. The precision of individual BMD estimates from genetic toxicity tests can be improved by covariate analysis of combined data sets for a shared endpoint when the dose–response data at each level of the covariate can be adequately described by a model that assumes constant shape parameters. Increased BMD precision affords an improved ability to rank and group potency values across covariate levels (e.g. compounds, exposure durations, cell lines, etc.).
3. Comparison of potencies requires consideration of BMD confidence intervals: plotting BMD confidence intervals provides an approach whereby both the magnitude of potential differences in BMDs and the uncertainty in the BMD estimates can be taken into account.
4. BMD confidence intervals can be employed to identify differences/similarities in potency that may be associated with a hypothesised MOA or cell line property (e.g. increased potency associated with increased metabolic capacity).
5. BMD confidence intervals would be the appropriate starting point in developing equipotent chemical groupings and/or subgroupings when supported by mechanistic information or structural similarities, thereby facilitating read across and data gap filling that could expedite regulatory evaluations and decision-making.

Supplementary data

Supplementary Figures S1–S4 are available at *Mutagenesis* Online.

Funding

G.E.J. was supported by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs, contract number NC/K500033/1) and also contract funding from the Government of Canada for the development of advanced quantitative methods for the analysis of genetic toxicology dose–response data. P.A.W.'s funding was provided by the Government of Canada's Chemicals Management Plan (CMP) research fund. J.W.W. is a Post-Doctoral Fellow supported under the Natural Sciences and Engineering Research Council of Canada's (NSERC) Visiting Fellowships program.

Acknowledgements

We wish to thank Professor Anthony Lynch (GlaxoSmithKline) and Dr Sibylle Ermler (Brunel University) for providing us with raw/numerical data from their published manuscripts. We also thank Dr Nikolai Chepelev, Mr Scott Hancock and two anonymous reviewers for providing insightful comments that contributed to marked improvements in the manuscript.

References

1. MacGregor, J. T., Frötschl, R., White, P. A. *et al*. (2015) IWGT Report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure-response relationships and points of departure (PoDs). *Mutat. Res.*, 783, 55–65.
2. MacGregor, J. T., Frötschl, R., White, P. A. *et al*. (2015) IWGT report on quantitative approaches to genotoxicity risk assessment II. Use of Point-of-Departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. *Mutat. Res.*, 783, 66–78.
3. Johnson, G. E., Slob, W., Doak, S. H. *et al*. (2015) New approaches to advance the use of genetic toxicology analyses for human health risk assessment. *Toxicol. Res.*, 4, 667–676.
4. Gollapudi, B. B., Johnson, G. E., Hernandez, L. G. *et al*. (2013) Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. *Environ. Mol. Mutagen.*, 54, 8–18.
5. Johnson, G. E., Soeteman-Hernández, L. G., Gollapudi, B. B. *et al*. (2014) Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. *Environ. Mol. Mutagen.*, 55, 609–623.
6. Slob, W. (2014) Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. *Crit. Rev. Toxicol.*, 44, 557–567.
7. Slob, W. (2014) Benchmark dose and the three Rs. Part II. Consequences for study design and animal use. *Crit. Rev. Toxicol.*, 44, 568–580.
8. Hernández, L. G., Van Benthem, J. and Slob, W. (2012) Estimating the carcinogenic potency of chemicals from the *in vivo* micronucleus test: an RIVM Report. *RIVM Report 340700007/2012*. Bilthoven, The Netherlands. http://www.rivm.nl/en/Documents_and_publications/Scientific/Reports/2013/april/Estimating_the_carcinogenic_potency_of_chemicals_from_the_in_vivo_micronucleus_test (accessed December 8, 2015).
9. Hernández, L. G., Slob, W., van Steeg, H. and van Benthem, J. (2011) Can carcinogenic potency be predicted from in vivo genotoxicity data? a meta-analysis of historical data. *Environ. Mol. Mutagen.*, 52, 518–528.

10. Hernández, L. G., Johnson, G. E., Pottenger, L. H., and van Benthem, J. (2011) Analysis of low-dose mutagenic responses and the applicability of genotoxicity tests for carcinogen potency prediction. *Environ. Mol. Mutagen.*, 52, S26.
11. Hernández, L. G., Slob, W., van Steeg, H. and van Benthem, J. (2010) Comparison of carcinogenic potency estimates to *in vivo* genotoxic potencies from the micronucleus, transgenic rodent mutation and comet assay using the benchmark dose approach. *Environ. Mol. Mutagen.*, 51, 707–707.
12. Slob, W. and Setzer, R. W. (2014) Shape and steepness of toxicological dose-response relationships of continuous endpoints. *Crit. Rev. Toxicol.*, 44, 270–297.
13. National Research Council (NRC) (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington, DC, doi:10.17226/11970
14. Crump, K. S., Chen, C. and Louis, T. A. (2010) The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. *Environ. Health Perspect.*, 118, 1350–1354.
15. Yoon, M., Campbell, J. L., Andersen, M. E. and Clewell, H. J. (2012) Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. *Crit. Rev. Toxicol.*, 42, 633–652.
16. Groothuis, F. A., Heringa, M. B., Nicol, B., Hermens, J. L., Blaauboer, B. J. and Kramer, N. I. (2015) Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology*, 332, 30–40.
17. Elhajouji, A., Lukamowicz, M., Cammerer, Z. and Kirsch-Volders, M. (2011) Potential thresholds for genotoxic effects by micronucleus scoring. *Mutagenesis*, 26, 199–204.
18. Rithidech, K. N. and Scott, B. R. (2008) Evidence for radiation hormesis after in vitro exposure of human lymphocytes to low doses of ionizing radiation. *Dose Response*, 6, 252–271.
19. Clare, M. G., Lorenzon, G., Akhurst, L. C. *et al.* (2006) SFTG international collaborative study on *in vitro* micronucleus test: II. Using human lymphocytes. *Mutat. Res.*, 607, 37–60.
20. Elhajouji, A., Tibaldi, F. and Kirsch-Volders, M. (1997) Indication for thresholds of chromosome non-disjunction versus chromosome lagging induced by spindle inhibitors in vitro in human lymphocytes. *Mutagenesis*, 12, 133–140.
21. Elhajouji, A., Van Hummelen, P. and Kirsch-Volders, M. (1995) Indications for a threshold of chemically-induced aneuploidy in vitro in human lymphocytes. *Environ. Mol. Mutagen.*, 26, 292–304.
22. Hernández, L. G., van Benthem, J. and Johnson, G. E. (2013) A mode-of-action approach for the identification of genotoxic carcinogens. *PLoS ONE*, 8, e64532, doi:10.1371/journal.pone.0064532.
23. Johnson, G. E. and Parry, E. M. (2008) Mechanistic investigations of low dose exposures to the genotoxic compounds bisphenol-A and rotenone. *Mutat. Res.*, 651, 56–63.
24. Johnson, G. E., Quick, E. L., Parry, E. M. and Parry, J. M. (2010) Metabolic influences for mutation induction curves after exposure to Sudan-1 and para red. *Mutagenesis*, 25, 327–333.
25. Lynch, A., Harvey, J., Aylott, M., Nicholas, E., Burman, M., Siddiqui, A., Walker, S. and Rees, R. (2003) Investigations into the concept of a threshold for topoisomerase inhibitor-induced clastogenicity. *Mutagenesis*, 18, 345–353.
26. European Food Safety Authority (EFSA) (2009) Use of benchmark dose approach in risk assessment: Guidance of the Scientific Committee. EFSA, http://www.efsa.europa.eu/en/efsajournal/pub/1150. Vol. 1150, pp. 1–72. doi:10.2903/j.efsa.2009.1150
27. Soeteman-Hernandez, L. G., Johnson, G. E. and Slob, W. (2015) Estimating the carcinogenic potency of chemicals from the *in vivo* micronucleus test. *Mutagenesis*, doi:10.1093/mutage/gev043
28. Bemis, J. C., Wills, J. W., Bryce, S. M., Torous, D. K., Dertinger, S. D. and Slob, W. (2015) Comparison of *in vitro* and *in vivo* clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. *Mutagenesis*, doi:10.1093/mutage/gev041
29. Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT) (2002) Risk assessment of mixtures of pesticides and similar substances. *A Report*. the Food Standards Agency (FSA). London UK. http://cot.food.gov.uk/sites/default/files/cot/reportindexed.pdf (accessed December 8, 2015).
30. Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT). (2007) Benzimidazoles: An approach to defining a common aneugenic grouping. *Annual Report*. 2007 Annual Report of the Committee on Mutagenicity Chiton, UK. http://cot.food.gov.uk/sites/default/files/cot/comsection07.pdf (accessed December 8, 2015).
31. Friedman, P. A. and Platzer, E. G. (1980) Interaction of anthelmintic benzimidazoles with Ascaris suum embryonic tubulin. *Biochim. Biophys. Acta*, 630, 271–278.
32. Friedman, P. A. and Platzer, E. G. (1978) Interaction of anthelmintic benzimidazoles and benzimidazole derivatives with bovine brain tubulin. *Biochim. Biophys. Acta*, 544, 605–614.
33. Lemieux, C. L., Long, A. S., Lambert, I. B., Lundstedt, S., Tysklind, M. and White, P. A. (2015) Cancer risk assessment of polycyclic aromatic hydrocarbon contaminated soils determined using bioassay-derived levels of benzo[a]pyrene equivalents. *Environ. Sci. Technol.*, 49, 1797–1805.
34. Environmental Protection Agency (EPA) Risk Assessment Forum (2000) Supplementary guidance for conducting health risk assessment of chemical mixtures. *A Report. EPA/630/R-00/002*, Washington DC.
35. Van den Berg, M., Birnbaum, L. S., Denison, M. *et al.* (2006) The 2005 World Health Organization reevaluation of human and Mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicol. Sci.*, 93, 223–241.
36. Ermler, S., Scholze, M. and Kortenkamp, A. (2013) Seven benzimidazole pesticides combined at sub-threshold levels induce micronuclei in vitro. *Mutagenesis*, 28, 417–426.
37. Yenjerla, M., Cox, C., Wilson, L. and Jordan, M. A. (2009) Carbendazim inhibits cancer cell proliferation by suppressing microtubule dynamics. *J. Pharmacol. Exp. Ther.*, 328, 390–398.
38. Singh, P., Rathinasamy, K., Mohan, R. and Panda, D. (2008) Microtubule assembly dynamics: an attractive target for anticancer drugs. *IUBMB Life*, 60, 368–375.
39. Sharp, D. C. and Parry, J. M. (1985) Diethylstilboestrol: the binding and effects of diethylstilboestrol upon the polymerisation and depolymerisation of purified microtubule protein in vitro. *Carcinogenesis*, 6, 865–871.
40. George, O., Bryant, B. K., Chinnasamy, R., Corona, C., Arterburn, J. B. and Shuster, C. B. (2008) Bisphenol A directly targets tubulin to disrupt spindle organization in embryonic and somatic cells. *ACS Chem. Biol.*, 3, 167–179.
41. Ikeda, N., Fujii, K., Sarada, M. *et al.* (2012) Genotoxicity studies of glycidol fatty acid ester (glycidol linoleate) and glycidol. *Food Chem. Toxicol.*, 50, 3927–3933.
42. Thomas, R. S., Philbert, M. A., Auerbach, S. S. *et al.* (2013) Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. *Toxicol. Sci.*, 136, 4–18.
43. Honma, M. and Hayashi, M. (2011) Comparison of in vitro micronucleus and gene mutation assay results for p53-competent versus p53-deficient human lymphoblastoid cells. *Environ. Mol. Mutagen.*, 52, 373–384.
44. European Commission Joint Research Centre's European Union Reference Laboratory for Alternative to Animal Testing (EURL ECVAM) (2015) Chemical categories and read across. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/background/chemical-categories-and-read-across (accessed October 20, 2015).
45. European Commission Joint Research Centre's European Union Reference Laboratory for Alternative to Animal Testing (EURL ECVAM) (2015) Compendium of case studies that helped to shape the REACH guidance on chemical categories and read across. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_22481_EN.pdf (accessed October 20, 2015).
46. European Commission Joint Research Centre's European Union Reference Laboratory for Alternative to Animal Testing (EURL ECVAM) (2015) *The Use of Computational Methods in the Grouping and Assessment of Chemicals - Preliminary Investigations*. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_22941_EN.pdf (accessed October 20, 2015).
47. European Chemicals Agency (ECHA) (2015) *Guidance on Information Requirements and Chemical Safety Assessment*. http://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment (accessed October 20, 2015).

48. European Chemicals Agency (ECHA) (2015) *How to Report Read-Across and Categories*. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/pg_report_readacross_categ.pdf (accessed October 20, 2015).
49. Organisation for Economic Co-operation and Development (OECD) (2004) *Manual for Investigation of HPV Chemicals*. http://www.oecd.org/chemicalsafety/risk-assessment/31177672.pdf (accessed October 20, 2015).
50. Stiborova, M., Schmeiser, H. H., Frei, E., Hodek, P. and Martinek, V. (2014) Enzymes oxidizing the azo dye 1-phenylazo-2-naphthol (Sudan I) and their contribution to its genotoxicity and carcinogenicity. *Curr. Drug Metab*., 15, 829–840.
51. Stiborová, M., Martínek, V., Schmeiser, H. H. and Frei, E. (2006) Modulation of CYP1A1-mediated oxidation of carcinogenic azo dye Sudan I and its binding to DNA by cytochrome b5. *Neuro Endocrinol. Lett*., 27 (Suppl. 2), 35–39.
52. Stiborová, M., Asfaw, B., Frei, E., Schmeiser, H. H. and Wiessler, M. (1995) Benzenediazonium ion derived from Sudan I forms an 8-(phenylazo)guanine adduct in DNA. *Chem. Res. Toxicol*., 8, 489–498.
53. Yauk, C. L., Lambert, I. B., Meek, M. E., Douglas, G. R. and Marchetti, F. (2015) Development of the adverse outcome pathway "alkylation of DNA in male premeiotic germ cells leading to heritable mutations" using the OECD's users' handbook supplement. *Environ. Mol. Mutagen*. 56, 724–750. doi:10.1002/em.21954.