Swansea University
Prifysgol
Abertawe

Application of Benchmark Dose Analysis to *in vitro* Genotoxicity Data for Compound Risk Characterisation

Ryan Paul Wheeldon BSc (Hons), ERT

Submitted to Swansea University in fulfilment of the requirements for the degree of Doctor of Philosophy.

2021

Acknowledgments

Summary

Genotoxic risk from exposure to pharmaceutical compounds has historically been focussed on dichotomous hazard characterisation, with little regulatory acceptance of risk assessment paradigms. The regulations focus on testing novel compounds with outdated genotoxicity test systems. Recent overwhelming support of the Benchmark Dose (BMD) methodology provides the baseline for advanced exposure risk assessments. Novel flow cytometric *in vitro* DNA damage response assays (MultiFlow and ToxTracker) have been developed that provide quantitative dose-response information that can be used in a high-throughput screening environment. In the following work, BMD modelling is applied to the MultiFlow and ToxTracker biomarker dose-response datasets. This work demonstrates that the MultiFlow dose-response biomarker datasets are amenable to BMD analysis for a set of clastogens and aneugens, and that the biomarker dose-responses correlate with dose-responses from the gold-standard *in vitro* micronucleus assay. A detailed appraisal of BMD confidence intervals (CIs) is provided for a selection of 10 clastogens requiring metabolic activation (with S9), demonstrating the criticality of using BMD uncertainty measures in comparative potency analysis. A comparative potency algorithm is developed and utilised in machine learning to distinguish four S9-dependent groupings: high and low-level potentiation, no effect, and diminution. A deep dive case study is presented for MultiFlow and ToxTracker analysis of Topoisomerase II Poisons, where BMD CI potency ranks are shown to correlate broadly with compound structural information. The Adverse Outcome Pathway (AOP) for Topoisomerase-II Poisoning is developed upon, and the Lhasa Derek Nexus alerts are mapped to the AOP. A Quantitative Structural Activity Relationship model is developed using Topoisomerase-II Poison molecular descriptors and BMD measurements from MultiFlow and ToxTracker biomarkers that correspond to Key Events relative to the Topoisomerase-II Poison AOP. This thesis provides an all-encompassing report of *in vitro* DNA damage response biomarker BMD analysis for compound potency ranking and read across.

Table of Contents

List of Tables

List of Figures

xviii

Abbreviations

AOP: Adverse Outcome Pathway

BMD: Benchmark Dose

BMDL: Benchmark Dose Lower Confidence Limit

BMDU: Benchmark Dose Upper Confidence Limit

CED: Critical Effect Dose

CEDL: Critical Effect Dose Lower Confidence Limit

CEDU: Critical Effect Dose Upper Confidence Limit

CES: Critical Effect Size

DNA: Deoxyribonucleic Acid

KE: Key Event

KNN: K-Nearest Neighbours

ML: Machine Learning

MOA: Mode of Action

MW: Molecular Weight

NOEL: No Observed Effect Level

NO(G)EL: No Observed (Genotoxic) Effect Level

PoD: Point of Departure

PSA: Polar Surface Area

QSAR: Quantitative Structural Activity Relationship

RNA: Ribonucleic Acid

Chapter 1. General Introduction

Genotoxicity Assessment in Industry

1.1 The Practice of Genetic Toxicology

Genetic toxicology is a subset of toxicology that evaluates the effects of chemical and physical agents on DNA. Permanent changes to DNA through mutations can have deleterious effects to the organism or offspring. Broader damage to DNA, for example unscheduled DNA synthesis, sister chromatid exchanges, and DNA strand breaks, can manifest into adverse health effects if left unchecked or unrepaired. Since its inception, genotoxicity has served as a proxy for cancer, as genotoxins are often viewed as carcinogens and cancer as a genetic disease characterised by mutations (Jenkins et al., 2010). Genetic toxicology testing covers a broad spectrum of endpoints that are assessed indirectly through measuring DNA repair or measuring the direct production of gene mutations or chromosome alterations. Consequentially, the new chemical entities that are designed by the pharmaceutical, biotechnology, chemical and cosmetic industries require assessment and regulation for genotoxic risk (Dearfield et al., 2017).

1.1.1 DNA Damage Mechanisms

There are many and varied types of DNA damage that are induced by exogenous agents including chemicals and ionising and non-ionising radiations. Such damage includes single- and double- strand breaks in the DNA backbone, crosslinks between DNA bases and proteins, and addition of chemical groups to the DNA bases (adducts). As a consequence of the induced DNA damage, mutations can form as a result of errors in DNA repair or replication (Klaassen, 2007, Miller and Miller, 1971). In somatic cells, mutations in critical genes such as tumour suppressor genes can lead to unregulated cell replication and cancer (Loeb and Loeb, 2000). Broadly speaking, chemicals can induce DNA damage through two categories: 1) By reacting directly with DNA as adducts at specific base locations (Heflich and Neft, 1994); or 2) indirectly in a non-DNA reactive fashion, typically through reactivity with the DNA replication machinery and/or mechanisms related to gene expression, for example interference with DNA synthesis pathways (Galloway et al., 1998), effects of

reactive oxygen species (Cooke et al., 2003, Evans et al., 2004), or Topoisomerase II inhibition (McClendon and Osheroff, 2007).

The following drawing by Klaassen (2007) (**Figure *1.1***) proficiently outlines the spectrum of DNA damage that is induced by different physical and chemical agents. The focus of this thesis is on genotoxicity risk from industrially produced chemical agents, and so the types of DNA damage induced by irradiation will not be expanded upon.
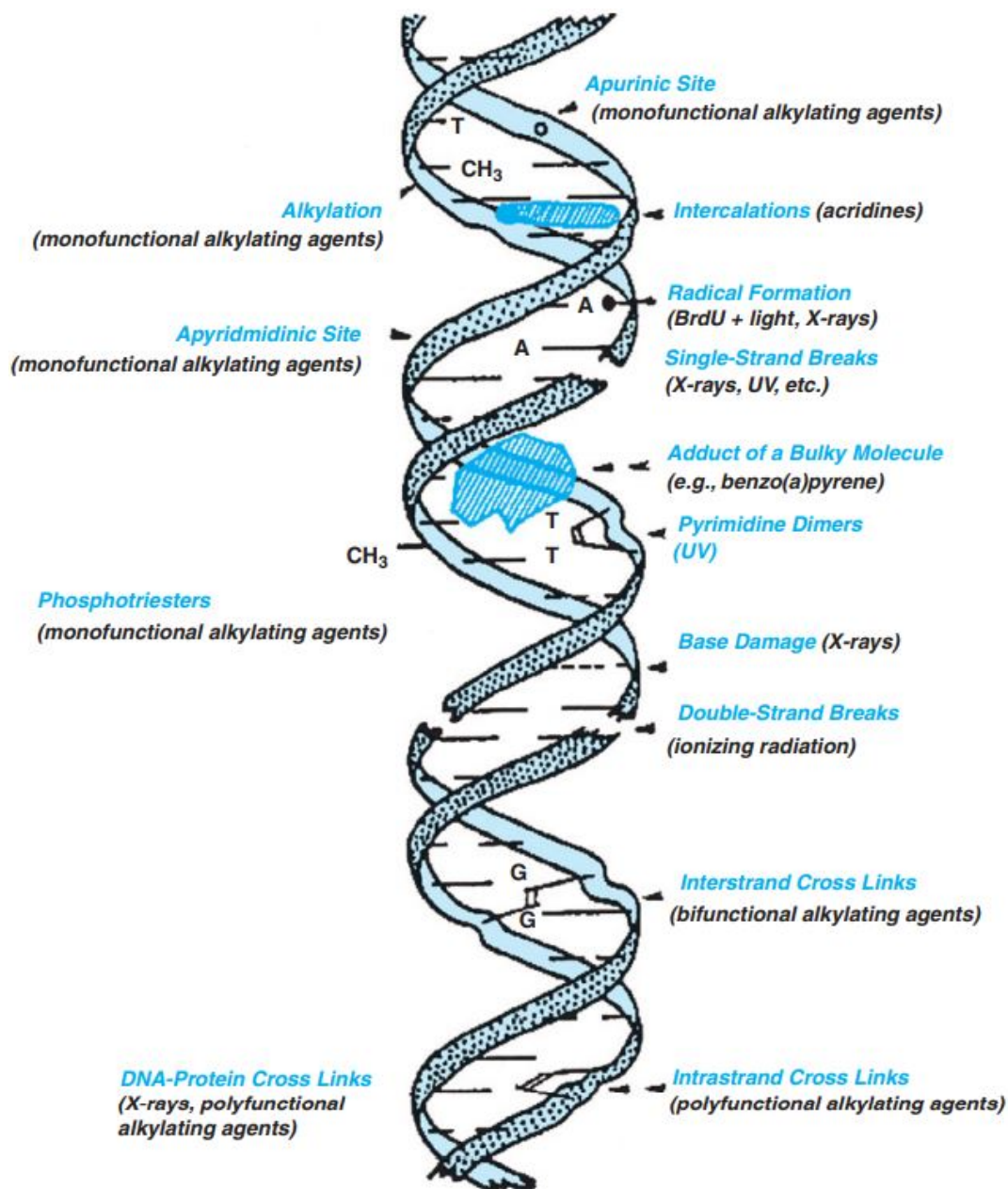
Figure 1.1. Types of DNA damage induced by physical and chemical agents. Figure adapted from Klaassen (2007).

1.1.1.1 DNA-reactive genotoxic compounds

Most DNA-reactive compounds are electrophilic and hence react with proteins and DNA's electron-rich nucleophilic centres. The Swain Scott constant (*s* value) serves as a standard measure of electrophilicity of alkylating agents (Swain and Scott, 1953). High *s* value compounds are able to target highly nucleophilic centres (the N^7-position of Guanine (N7G), N^3-position of Adenine (N3A)), contrasting low *s* value compounds are less sensitive to the nucleophilicity of their target; thus, do not discriminate between high and low centres of nucleophilicity and are able to efficiently alkylate Oxygen atoms (e.g., 0^6-position of Guanine (O6G)). Such compounds include potent alkylating agents that can react with DNA in either a mono- or bi-functional mechanism. Monofunctional alkylating agents target the N7G, N3A and O6G reactive sites, and include the alkane sulfonates methyl methanesulfonate (MMS) and ethyl methanesulfonate (EMS), and the N-Nitrosamine group of compounds (Beranek, 1990). Bifunctional alkylating agents include nitrogen mustards that can target multiple DNA active sites simultaneously leading to DNA adduct formation or inter-strand cross-linking (Shahin et al., 2001). These DNA reactive compounds usually lead to mutations in the DNA sequence, and/or clastogenicity where the integrity of the chromosomes is impacted.

1.1.1.2 Non-DNA Reactive Genotoxic Compounds

Not all genotoxic compounds interact directly with the DNA backbone. Such agents that lead to indirect genotoxicity can be equally as potent as direct DNA reactive compounds (Lutz, 1998). Some classical examples include a subset of aneugens that are known to interact with the cellular mitotic machinery, such as the vinca alkaloid tubulin poisons Vincristine and Vinblastine that inhibit tubulin production leading to potentially uneven chromosome separation during mitosis (Owellen et al., 1972; Dhamodharan et al., 1995). Another mechanism of indirect DNA damage includes that influenced by the Topoisomerase II poisons, including etoposide that covalently interacts with the Topoisomerase II enzyme active site (Hande, 1998; Baldwin and Osheroff, 2005), leading to an overwhelming increase in enzyme mediated DNA double strand breaks and replication fork collapse (McClendon and Osheroff, 2007). Topoisomerase II Poisons can induce deleterious cellular perturbation, and inherited mutations and chromosome rearrangements can be inherited from and cells that do not induce apoptosis (Montecucco et al., 2015). The latter has been implicated in the

development of treatment-related acute myeloid leukaemia (AML) in 2-3% of cancer patients treated with etoposide (Bueno et al., 2009; Pendleton et al., 2014).

1.2 Industrial Genotoxicity Assessment Regulatory Landscape

Originally, exogenous agents were thought to be responsible for approximately 85% of human cancers (Wood, 1969). In as long ago as 1775, Percival Pott fascinatingly observed an increase in scrotal cancers in chimney sweeps (Androutsos, 2006). Pott's findings can be considered an early epidemiological study. Of course, exposure to chimney Soot is limited to the few individuals with historical occupational exposure; however, Pott's work highlights the risk of exposure to hazardous substances. The advancements in access to healthcare and the broader use of consumer products means that more of the world's population are exposed to exogenous compounds that are industrially synthesised which could pose a genotoxic risk. In the regulated industry of compound development, including the pharmaceutical, biotechnology, and chemical sectors; genotoxicity is conservatively assessed through hazard identification which merely identifies if a compound falls into the binary classification of genotoxic or not. Specific regulatory genotoxicity guidelines describe and standardise the assays which are used in industry to hazard assess genotoxic compounds. Specifically, the pharmaceutical sector adheres to guidelines set by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICHS2R1). Assessment for genotoxic hazard follows a 2-step process of performing *in vitro* (stage 1) and *in vivo* (stage 2) assays.

1.2.1 Stage 1: The Standard *In Vitro* Test Battery

The work by Kirkland et al. (2005) has been key to the inclusion of the core *in vitro* test battery in the regulatory assessment of a compound's genotoxic potential. The authors concluded that the use of a combination of three common assays (Ames bacterial reverse mutation assay, *in vitro* mouse lymphoma assay, *in vitro* chromosome aberration assay and/or *in vitro* micronucleus assay) increases the sensitivity and predictability to detect a rodent carcinogen. The three-test battery is mostly applicable to cosmetic products, whereas pharmaceutical products are advised to follow a two-test battery (Kirkland et al., 2011). This is practical since no single test can detect all genotoxic mechanisms relevant to carcinogenesis. The ICHS2 guideline on genotoxicity testing and data interpretation for pharmaceutical intended for human use recommends the use of a bacterial reverse mutation

assay, with the addition of an *in vitro* mammalian test (ICH, 2008). Kirkland et al. (2005) highlighted the poor specificity of the assays, where the rate of false positives is high (75-95%), which can be reduced by the inclusion of multiple assays in the battery. Well characterised compounds with negative results in the stage 1 tests are generally concluded to be non-genotoxic. Based on the predicate that the *in vitro* standard battery is predictive of *in vivo* genotoxicity, compounds which elucidate a positive response in any of the stage 1 tests are required to undergo follow-up *in vivo* genotoxicity testing in stage 2 tests. The United Kingdom Committee on Mutagenicity Guidance on a Strategy for Genotoxicity Testing of Chemical Substances (Committee on Mutagenicity of Chemicals in Food, 2011) outline the stages involved in testing of compounds for regulatory assessment of genotoxic potential. A flow chart covering the stage 1 approach is shown in **Figure** *1.2*.

Stage 1 Core Tests:
1. Bacterial gene mutation (Ames test)
2. Clastogenicity and aneugenicity (*in vitro* micronucleus test)

NEGATIVE results in all Stage 1 tests

EQUIVOCAL result in any Stage 1 test

POSITIVE result in any Stage 1 test

Consider other factors that indicate additional evaluation is required:
- Structural alerts,
- Results of other tests (e.g. rodent tumours)

Consider:
- Reproducibility
- Historical control data
- Mode of Genotoxic Action (MoGA)
- Results of stage 0
- Additional *in vitro* tests (e.g. Mammalian cell mutation, mammalian cytogenetics, Comet assay, or human reconstituted skin)

Consider:
- Mode of Genotoxic Action (MoGA)
- Results of stage 0
- Misleading positive results (e.g. bacterial specific metabolism, or excessive mammalian cell cytotoxicity)

Substance is **not mutagenic**

Insufficient evidence to assess the mutagenicity of the substance

Substance should be considered to be an *in vitro* **mutagen**.

If high, or moderate and prolonged exposure, consider proceeding to Stage 2 only where *in vivo* testing is permitted*

Proceed to Stage 2 only where *in vivo* testing is permitted*

* In situations where *in vivo* testing is prohibited, further *in vitro* testing should be considered

Figure 1.2. Flow chart showing the stage 1 *in vitro* genotoxicity testing process involved in compound hazard identification. Figure adapted from Committee on Mutagenicity of Chemicals in Food (2011).

1.2.2 Stage 2: Inclusion of an *In Vivo* Genotoxicity Test

Due to differences in compound toxicokinetic properties between *in vitro* and *in vivo* test systems, the inclusion of an *in vivo* assay is recommended in the standard battery, and as a follow up to a positive *in vitro* test result. Various guidelines including ICHS2, recommends that an *in vivo* follow up assay should be performed in a relevant test species

using appropriate tissues based on a weight of evidence approach and expert judgement (Thybaud et al., 2007). Some common assays include the rodent bone marrow chromosome aberration assay to detect clastogenicity, the transgenic rodent mutation (TGR) assay to detect gene mutations, and the rodent comet assay to detect a wide range of DNA damage effects. A more recent assay includes the *in vivo* Pig-a (phosphatidylinositol glycan, Class A) assay (Dobrovolsky et al., 2010), which offers an efficient follow up to *in vitro* mutagenicity positive results (Aardema, 2013). Although specificity is higher in the *in vivo* assays than the standard battery *in vitro* assays, the question of the relevance of the assays persists. For example, (Tweats et al., 2007) highlights critical compounds that are positive i*n vivo* without demonstrating evidence of genotoxicity in the standard *in vitro* assays. Physiological issues including disturbances in animal body temperature and increases in erythropoiesis were hypothesised to contribute to the observations made by (Tweats et al., 2007). The following flow chart (**Figure 1.3**) from (Committee on Mutagenicity of Chemicals in Food, 2011) summarises the step 2 process involved in current genotoxicity *in vivo* follow up testing.
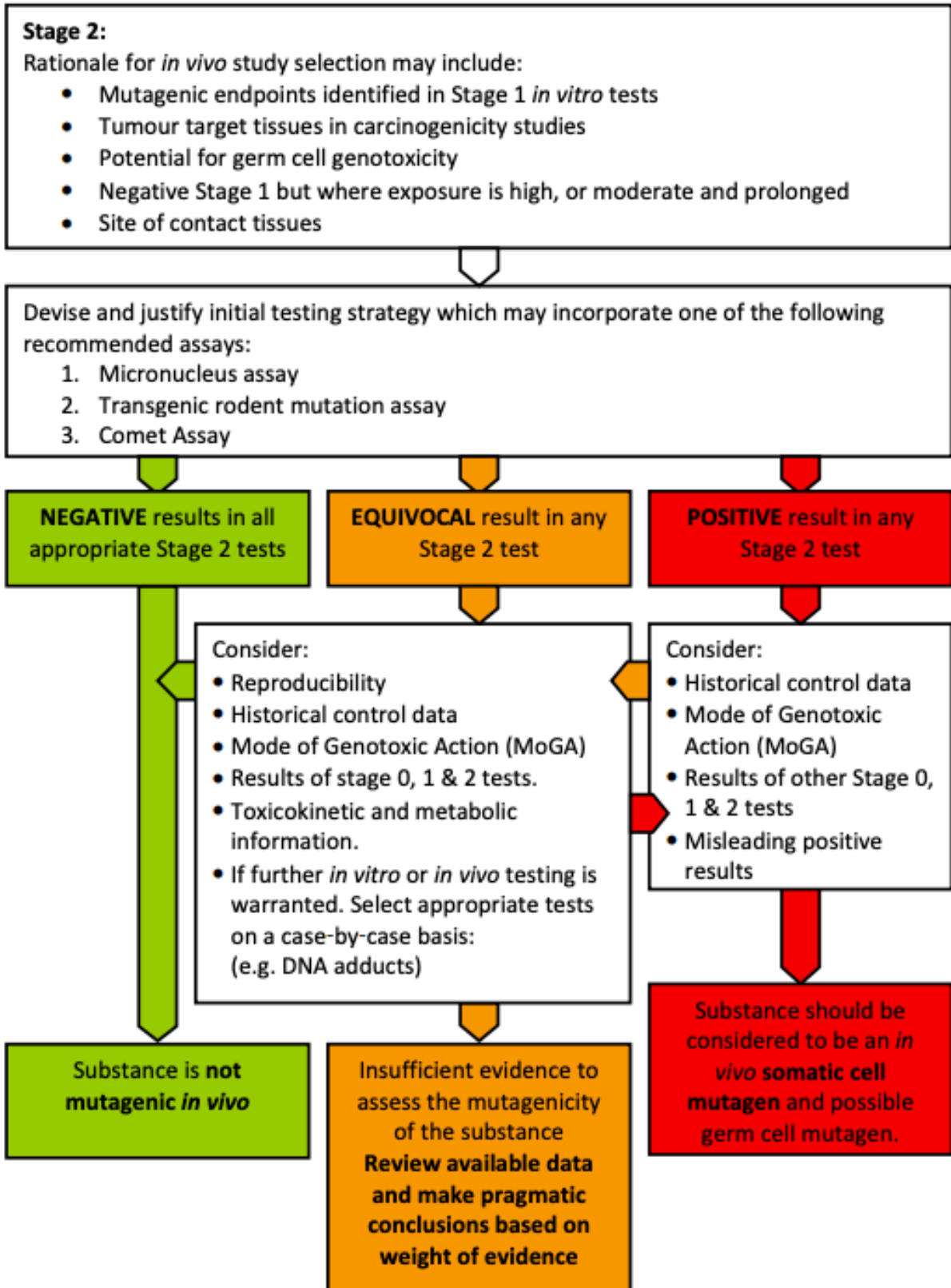
Figure 1.3. Flow chart showing the stage 2 *in vivo* genotoxicity testing process involved in compound hazard identification. Figure adapted from Committee on Mutagenicity of Chemicals in Food (2011).

1.2.3 Hazard and Risk Assessment Paradigms

Despite some of the technical drawbacks associated with the standard genotoxicity test assays – of which expansion upon is out of scope of this thesis – the main issue associated with the application of the standard battery is that the results assign a dichotomous outcome to the compounds genotoxic potential. Generally, a single high concentration of test compound is applied to the *in vitro* genotoxicity assays, and high (maximum tolerated dose) doses for *in vivo* studies (Cimino, 2006), which is consistent with the hazard assessment approach, dubbed *"screen and bin"* (Pottenger and Gollapudi, 2010). However, there is a growing interest in understanding a substance's potential genotoxic risk to exposed persons, of which a detailed dose-response analysis is required (Gollapudi et al., 2013b). Over the past two decades, several researchers have focussed their efforts on understanding the detailed dose-response relationships of genotoxic compounds in various tests systems, which better encompasses the essential aspects of risk assessment. The next section focusses on dose-response relationships in genetic toxicology which is the central theme of this thesis.

1.3 Dose Response Relationships in Genetic Toxicology

"All things are poison, and nothing is without
poison; the dosage alone makes it so a thing
is not a poison."

Paracelsus, c. 1564

1.3.1 Linear and Threshold Dose Responses

Direct acting genotoxic agents have long been assumed to follow a linear dose-response relationship (Henderson et al., 2000). DNA damage induction in the linear model is believed to be directly proportional to dose; implying that there are no genotoxic doses, however low, that are without finite risk of genetic damage and hence cancer. The linear concept has been implemented due to a combination of historic experimental data and precautionary postulate (Jenkins et al., 2010).

In practical terms, low dose linear extrapolation has been implicated in regulatory limits for controlling potentially genotoxic impurities in pharmaceutical products (ICH, 2013). Briefly, the allowable levels for lifetime exposure to genotoxicants was developed by the US FDA by using the TD_{50} (dose of statistically significant increase in tumours in 50% of

animals) from over 700 carcinogens in chronic animal carcinogenicity studies, where carcinogenicity was concluded to be the most sensitive endpoint from exposure to chemicals (Gold et al., 1948; 1989; 1995). The FDA concluded that for combined food and drink consumption, 0.025 μg/kg bw/day, or 1.5 μg/person/day (for a 60 kg individual), presents negligible concern for carcinogenic substances. The regulatory limit concept was termed the "Threshold of Toxicological Concern", and later applied to pharmaceutical genotoxic impurities where a TTC value of 1.5 μg/person/day is associated with a 1 in 10^6 increased lifetime risk of cancer (Boobis et al., 2017). Although the TTC concept is useful in screening and prioritization, the limit presents practical challenges since it is inherently highly conservative.

The linear model has been challenged by several researchers (Hengstler et al., 2003; Pottenger and Gollapudi, 2010; Gollapudi et al., 2013a; Guerard et al., 2015; Calabrese, 2019) as fallacious since the model assumes a dichotomous situation where a compound is either genotoxic or not, without considering the effect of dose. The role of dose in mutagenicity testing is a critical issue, and inappropriately high doses are suggested to be responsible for false positive results in *in vitro* genotoxicity tests (Kirkland et al., 2007). High doses are often used to maximise exposure to the low sensitivity genotoxicity test systems to increase the chance of detecting a genotoxic agent. However, this has far-reaching consequences when assuming the effects are induced in a linear fashion, and high doses are extrapolated down to the low dose region.

There are theoretical arguments against linear dose-response relationships for human genotoxic exposure, including epithelial barrier exclusion; water soluble genotoxin detoxification and excretion; cellular and nuclear membrane barriers reducing access of genotoxins to the nucleus; DNA repair mechanisms; DNA redundancy; and apoptosis/autophagy (Jenkins et al., 2010). DNA repair is crucial since it is central to genome stability, and DNA repair pathways have been shown to repair specific DNA damage types (Riches et al., 2008). Hence, it is plausible that DNA repair has a direct impact on the linearity of genotoxicity, particularly at low doses where DNA repair mechanisms may not be overwhelmed.

Experimental evidence to attempt to dispel genotoxic linear dose relationships include a pivotal study in 2007 by scientists at Swansea University titled "Mechanistic Influences for Mutation Induction Curves after Exposure to DNA-Reactive Carcinogens" (Doak et al.,

2007). The researchers generated *in vitro* genotoxicity study data for the model DNA reactive alkylating agents' methyl methanesulfonate (MMS), ethyl methanesulfonate (EMS), methylnitrosourea (MNU) and ethylnitrosourea (ENU). Human lymphoblastic AHH-1 cells were exposed to varying concentrations of the test compounds and analysed for chromosome damage in the cytokinesis blocked *in vitro* micronucleus assay and point mutations in the hypoxanthine phosphoribosyl transferase (*HPRT*) assay. **Figure** *1.4* and **Figure** *1.5* show the dose-response relationships for MMS, EMS, MNU and ENU responses in the *in vitro* micronucleus assay and the *HPRT* forward mutation assay, respectively.
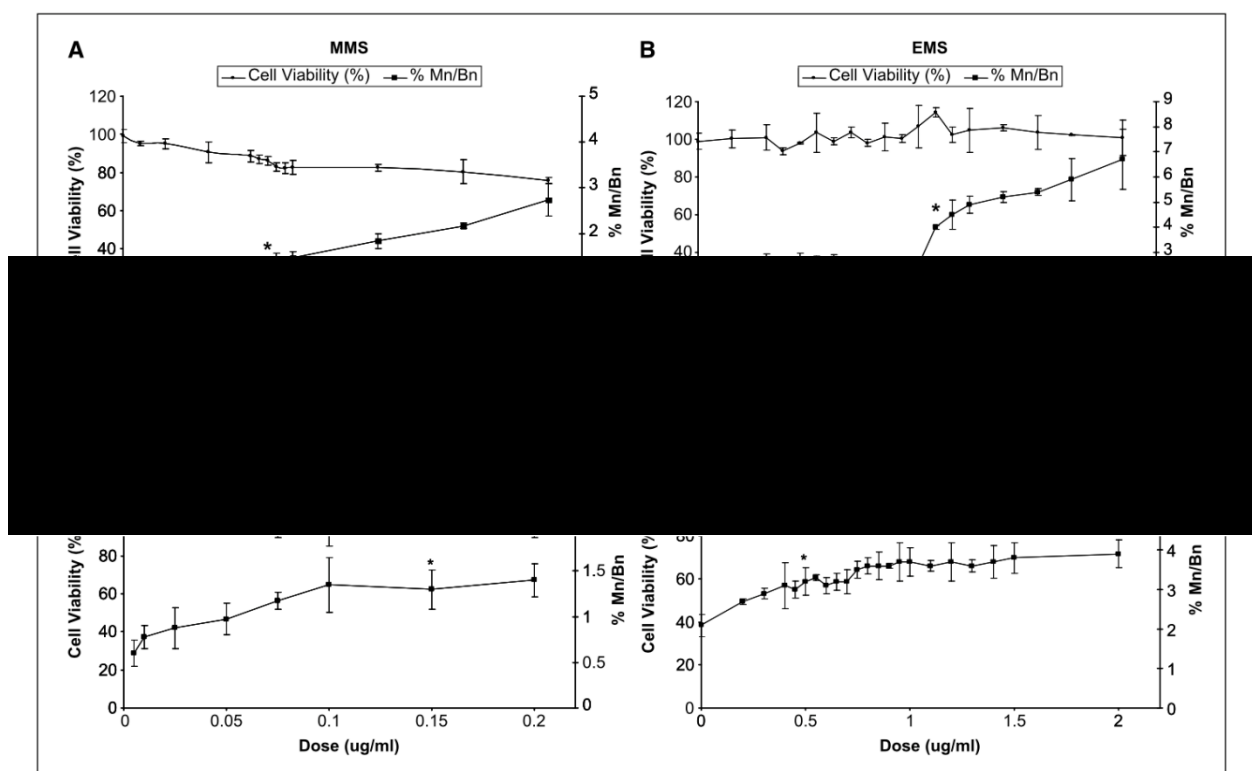
Figure 1.4. Dose-response relationships of MMS (A), EMS (B), MNU (C), and ENU (D) for micronucleus frequency in the AHH-1 cell line. Points represent the mean of duplicate treatments. Bars represent SD. *, indicates the first statistically significant increases in chromosome damage at 0.85 μg/mL MMS (A), 1.40 μg/mL EMS (B), 0.15 μg/mL NMU (C), and 0.50 μg/mL ENU (D). %Mn/Bn, percentage of binucleated cells containing one or more micronuclei. Image taken directly from Doak et al. (2007).
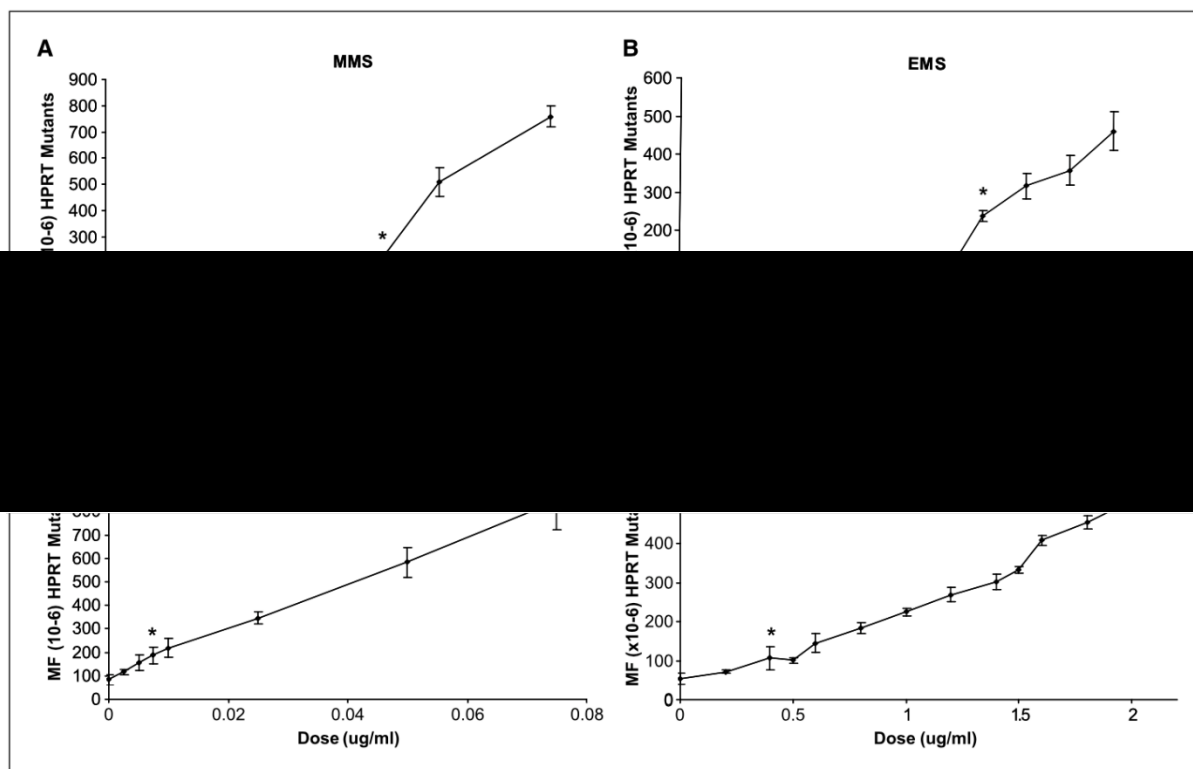
Figure 1.5. Dose-response relationships of MMS (A), EMS (B), MNU (C), and ENU (D) for *HPRT* gene mutation frequency (MF) – the number of 6-thioguanine resistant clones/10^6 clone-forming cells. Points represent the mean mutation frequency calculated from 100x 96-well plates where each dose exposure was performed in triplicate. Bars represent SD. *, indicates the first statistically significant increases in MF at 1.25 µg/mL MMS (A), 1.40 µg/mL (B), 0.0075 µg/mL MNU (C), and 0.40 µg/mL ENU (D). Image taken directly from (Doak et al., 2007).

Examination of the dose-response relationships with respect to chromosome damage in **Figure *1.4*** indicates regions of the curve where there is no statistically significant increase in response compared to background, ranging between 0 to 0.80 µg/mL for MMS and 0.135 µg/mL for EMS. Sharp statistically significant increases in micronuclei frequency were then observed corresponding to concentrations of 0.85 and 1.4 µg/mL for MMS and EMS, respectively. The alkylnitrosoureas, MNU and ENU, exhibited contrasting dose-responses compared to the alkylmethanesulfonates, MMS and EMS, where no clear statistically significant increase in response was observed at either tested concentration. Similar dose-response relationships were observed in the curves for *HPRT* MF for the alkylmethanesulfonates, with where there is no statistically significant increase in response compared to background both MMS and EMS within 0 to 1.0 and 0 to 1.2 µg/mL,

12

respectively. Statistically significant increases occurred at 1.25 µg/mL for MMS and 1.4 µg/mL for EMS.

The linear-dose responses observed for MNU and ENU findings were consistent with the prevailing belief that DNA reactive agents induce DNA adducts in a linear manner, where a single DNA-reactive interaction has the potential to form a neoplasm (Kirsch-Volders et al., 2000). However, most importantly the studies demonstrated that a threshold dose-response exists *in vitro* for the alkyl methanesulfonates MMS and EMS. An understanding of the genotoxic mode of action of the alkylsulfonates (MMS, EMS) and nitrosoureas (MNU, ENU) was critical to the study's conclusions. The study by Doak et al. (2007) took into consideration the potential for cytoprotective mechanisms to counteract DNA insult after exposure to genotoxic agents. The authors discussed that the doses that did not induce a statistically significance increase in micronucleus frequency relative to background, observed in the micronucleus assay with MMS and EMS exposure, are likely due to the repair of N^7-alkyl-Guanine adducts, and the repair of O^6-alkyl-Guanine adducts could yield the same dose response for these compounds in the HPRT assay. Doak et al. (2007) describe that at the low concentration EMS and MMS exposure level, the rate of DNA adduct repair is faster than the rate of formation; however, at concentrations equal to and greater than the statistically significant increases, repair pathways such as error-free Base Excision Repair (BER) and Methyl Guanine Methyl Transferase (MGMT) are working at their full capacity and adducts start to escape repair, yielding adverse effects.

The research by Doak et al. (2007) was the first published instance of the *in vitro* genotoxic dose response characteristics of low concentration alkylating agents. Moreover, the data demonstrated the plausibility of a threshold mechanism for EMS and MMS exposure to mammalian cells. These datasets lead to a paradigm shift in genetic toxicology and have led to the advances in genotoxicity assessment since 2007. Other researchers have confirmed the presence of threshold responses in a variety of test systems (Gocke and Wall, 2009; Pottenger et al., 2009; Bryce et al., 2010). Experimental data of non-DNA reactive agents such as spindle poisons supported theories of non-linear threshold dose-response relationships (Johnson and Parry, 2008) for said class of compounds.

1.3.2 The No Observed Effect Level (NOEL)

As experimental evidence emerged that some genotoxins display non-linear dose responses, attention turned to understanding how dose-responses could be used to derive biologically meaningful points of departure (PoD) to inform regulatory decision making. In this context, the PoD is defined as the dose corresponding to the first observed biological response in the dose-response curve (Sturla, 2018). In other fields of toxicological sciences since the early 20th century, sigmoidal (S-shaped) dose response curves have been implied as standard, where a marked increase in response between 20-80%, and asymptotically approaching 0 or 100% at the tails of the distribution (Calabrese, 2005). These sigmoidal curves provided data driven support that thresholds exist at low doses in classical toxicity endpoints. Thus, the no-observed-effect-level (NOEL) [or no-observed-adverse-effect-level, (NOAEL), where the effect is adverse] serves as the PoD along the dose response curve and have formed an integral part of non-clinical risk assessment to determine acceptable intakes to non-carcinogenic compounds. Support for use of the NOEL/NOAEL approach is primarily driven by its ease of use, as depicted by the simplicity in determining the PoD as highlighted in **Figure *1.6*** below. The investigating scientist simply has to test the statistical significance of each dose/concentration relative to the preceding dose or concomitant background response.
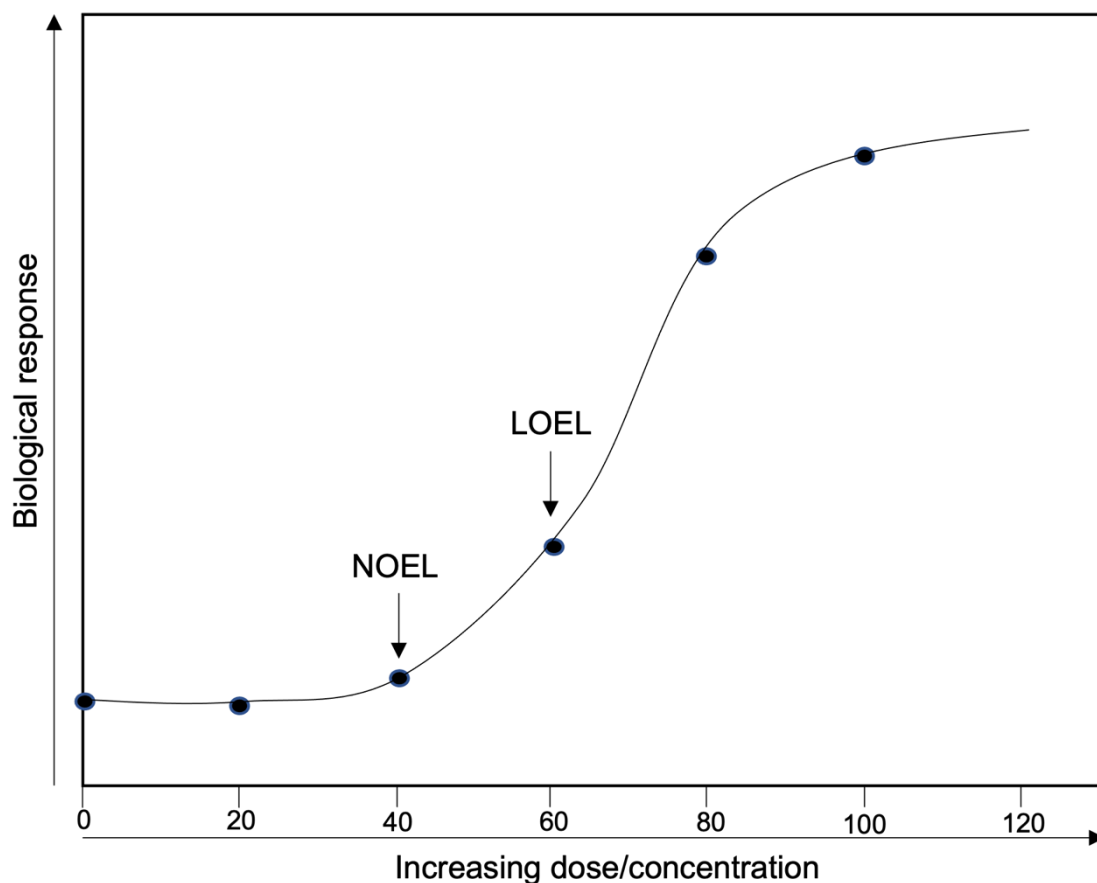
Figure 1.6. Depiction of a non-linear dose-response curve. Hypothetical doses/concentrations are plotted on the X axis, with an arbitrary increase in biological response on the Y axis. The highest dose that corresponds to a non-statistically significant increase in biological response compared to background serves as the NOEL. The lowest dose that corresponds to a statistically significant increase in biological response compared to background serves as the LOEL.

The shape of the dose-response curve draws parallels with that of the *in vitro* micronucleus and HPRT MF curves for EMS and MMS shown in section 1.3.1 (**Figure *1.4*** and **Figure *1.5***). In fact, the low dose region of said curves can be considered the NOEL, with the first concentration corresponding to a statistically significant increase in response being regarded as the LOEL. Since these studies evaluate genotoxic endpoints, some researchers refer to associated PoDs as NOGELs or LOGELs, where the acronym's 'G' stands for genotoxic.

Opposition for the NOEL/LOEL approach has been postulated by Crump (1984). His research posed serious objections including difficulties with the effect of sample size, and the

restriction to the design of the toxicological study. Firstly, with respect to the effect of sample size, Crump (1984) hypothesised that larger studies naturally demonstrate a higher probability of showing a statistically significant result compared to smaller studies, thus yielding a skewed lower PoD. In practical terms, study directors are penalized by using larger studies to demonstrate safety, and so regulatory agencies set minimum requirements for group size. Secondly, with respect to study design, Crump (1984) describes that dose-responses are under-utilized. For example, a NOEL is determined based on whether an effect is observed or not, and generally disregards the magnitude of the effect and the relationship amongst the effect observed at the other dose levels. Put another way, the dose-response trend is ignored. Crump (1984) outlined a hypothetical model for two experiments A, and B, to aid in presentation of his reasoning, as shown in **Figure *1.7***. Experiment A shows a sharp increase in dose response with increasing dose, whereas experiment B shows a flatter dose-response which is consistent with a linear model with intercept at the origin. Since there is a sharp decrease in response with decreasing dose in experiment A, compared to experiment B, the NOEL for experiment A should be larger than the NOEL for experiment B. However, since the response at d_{2A} was barely significant, the NOEL for experiment A is d_{1A}, which is lower than the NOEL d_{2B} for experiment B.
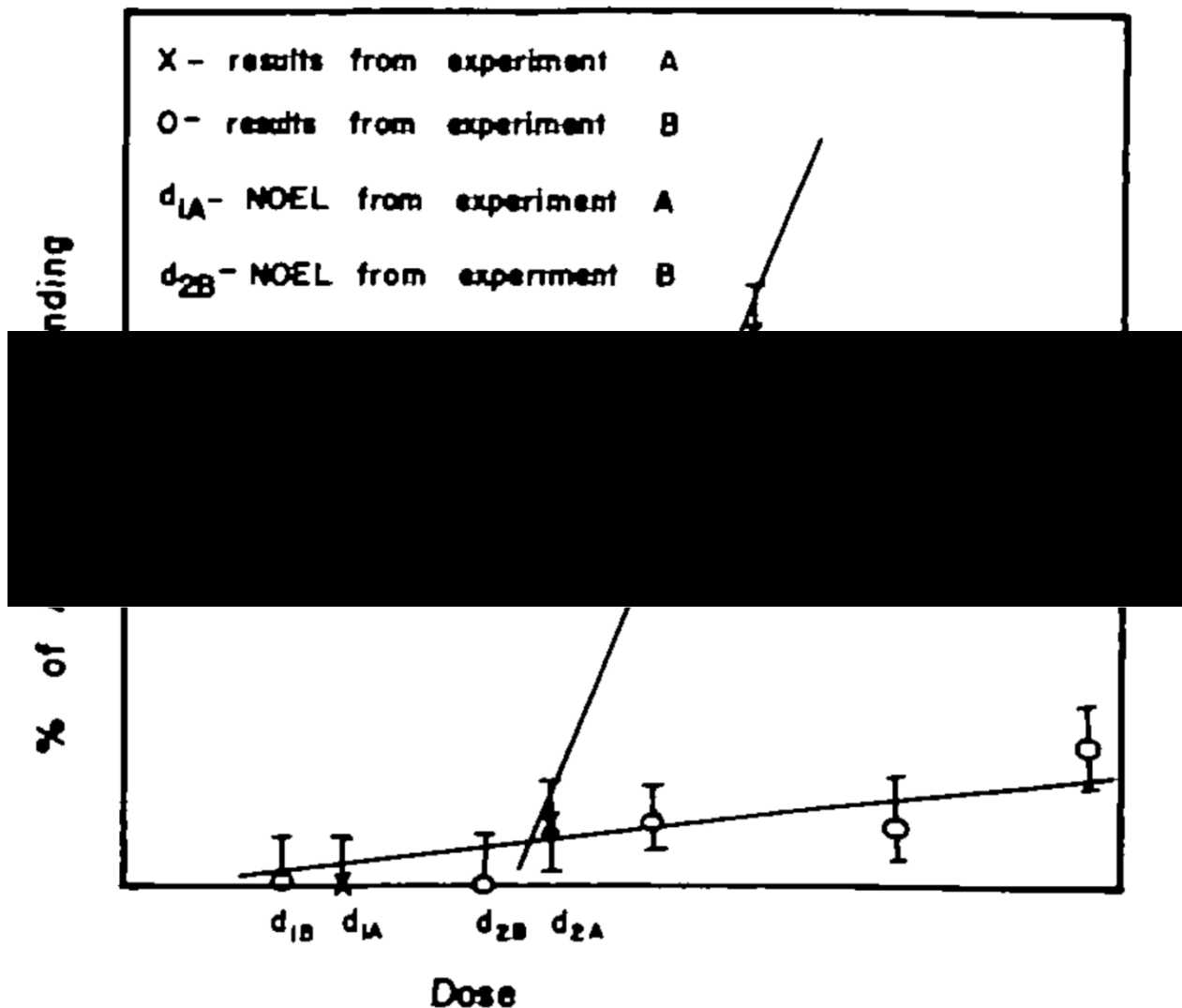
Figure 1.7. Hypothetical data for two animal studies A, and B with 95% confidence limits shown. Figure taken directly from Crump (1984).

The same dose-response constrictions are applicable to genotoxicity tests. The emerging evidence that some genotoxic agents may display threshold mechanisms, and the need for accurate dose-response analysis led to the formation of the Health and Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GTTC) Quantitative Analysis Workgroup (QAW) in 2010-2011 (Gollapudi et al., 2011).

The subgroup of the QAW assembled to assess the use of the NOEL approach using data for the induction of micronuclei and gene mutations *in vitro* and *in vivo* for the alkylating agents MMS and EMS (Gollapudi et al., 2013b). The workgroup tested their proof of concept using published *in vitro* data from the Doak et al. (2007) pivotal study, and *in vivo*

data from Lebaron et al. (2008). The data were selected since the *in vitro* endpoints contained additional replicates at the four dose levels surrounding the NOEL, allowing the scoring of 10,000+ cells at the same PoD dose region, whilst the *in vivo* studies also evaluated >10,000 cells per replicate. Gollapudi and colleagues performed simple statistical evaluations of the datasets including analysis of variance (ANOVA), and one-sided Dunnett's testing, to determine the NOGELs and LOGELs. The researchers outlined the same critical limitations that were highlighted by Crump (1984), primarily that the derivation of a NOEL is influenced by study design restricted to the dose levels that are tested. But what about the doses between dose levels? The NOEL (and NOAEL) dose-response model effectively ignores the inter-dose regions and fails to account for measures of uncertainty that are associated with the experimental results (Gollapudi et al., 2013b).

One could argue that the NOEL approach is simple to use and does not require the analyst to have an advanced understanding of statistics; however, when test results do not provide evidence of an obvious dose-response curve, the data is unsuitable for determining a NOEL, and hence fails to establish a PoD to determine health-based guidance values for human exposure scenarios. Gollapudi et al. (2013) highlight that the main consequence is that compounds that could have tremendous societal value cease development.

1.3.3 Threshold Dose Response Models

The limitations associated with the use of NOEL determination, including restrictions to the tested doses, and lack of transparency around uncertainty measurements, lead to the development of mathematical models to determine the Threshold Effect Level (Td) (Lynch et al., 2003; Lutz and Lutz, 2009). The Td was defined as the statistically identified dose (importantly not limited to one of the tested doses) at which the effect is indistinguishable from the background level, and at which it is possible to observe an increase in response above the background level (Gollapudi et al., 2013b).

The Td models are based upon decision rules that are applied to the statistical analyses of the dose response datasets. Two models are described in the literature, the bi-linear hockey stick model (Lutz and Lutz, 2009), and the broken stick model (Lynch et al., 2003). Put simply, the models include hypothesis testing to confirm or reject the presence of a linear dose-response. If the model rejects the linear dose-response argument, a more advanced statistical model to determine the Td is needed. The bilinear hockey stick model

developed by Lutz and Lutz (2009) (**Figure *1.8***) is a relatively simple mathematical model, with the function $y(d) = a + b \times (b - td) \times 1[d > td]$ [summand $= 0$ for $d \le td$]. This function is explained by the dose-response starting with a straight line with slope 0 followed through to the threshold dose *td* and continues with a straight line with slope $b > 0$. The parameters *a, td,* and *b* are estimated for best fit according to a 'hockey stick' shape along the dose response, as shown in **Figure *1.8***. An advantage of this model is that confidence intervals can be calculated, with a level of significance usually equal to 90%. The lower limit of the confidence interval associated with the threshold dose (LLCI td) is the inflection point where the change in gradient is at its maximum. One of the benefits of the mathematical model is that the associated uncertainty of the measurements is included.
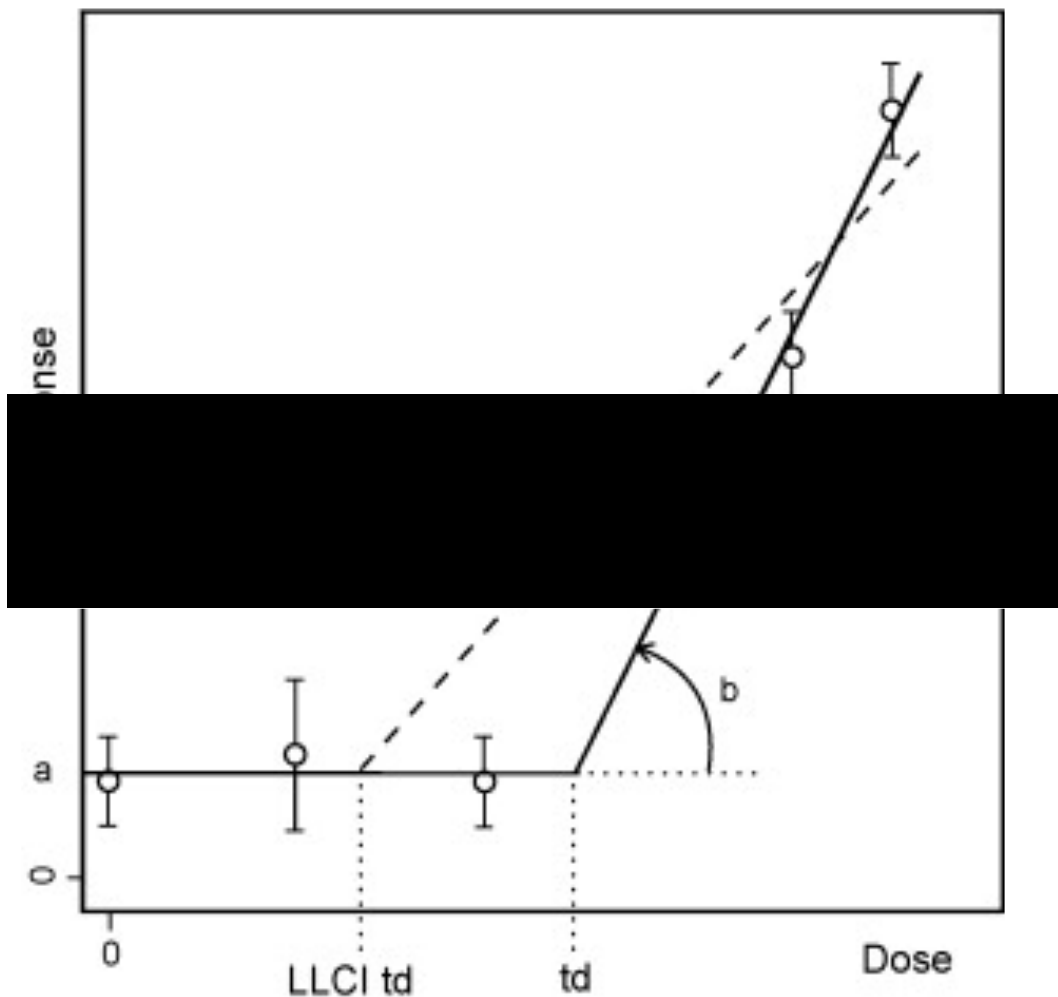
Figure 1.8. Depiction of a hypothetical sublinear dose-response dataset which is mathematically fit according to the 'hockey stick' function $y(d) = a + b \times (b - td) \times 1[d > td]$. 90% confidence intervals are indicated. The lower limit of the confidence

19

interval (LLCI) associated with the threshold dose (td) yields a dashed line, termed the 'broken hockey stick' dose-response. Figure taken directly from Lutz and Lutz (2009).

The hockey stick model introduced by Lutz and colleagues (2009) was not intended to support a mathematic threshold. However, the intent was to provide a toxicologist with a tool to test the degree of deviation from linearity for a given dose response. Johnson and colleagues (2009) used the hockey stick model to further evaluate the *in vitro* micronucleus induction and *HPRT* mutation data for MMS, EMS, MNU and ENU data previously published by Doak et al. (2007). The results of the analysis showed that the hockey stick model best described the dose-responses for both MMS and EMS in the micronucleus assay; and MMS, EMS and surprisingly ENU in the HPRT assay. However, the mathematical model showed that MNU was linear in both assays. One of the major discussion points in the Johnson et al. (2009) paper was that it is imperative for toxicologists have tools to use that provide objectivity in dose-response analysis. Without such tools, the toxicologist defaults to an unreliably subjective (sometimes by eye) interpretation of the shape of a dose-response curve. The hockey stick model does provide a useful tool for determining if a threshold dose-response exist, and it provides a measure of uncertainty, which is an advantage over the NOEL method; however, there are limitations in the data requirements which can make its use disadvantageous. An input prerequisite is that there are multiple dose levels above and below the threshold dose (Gollapudi et al., 2013b), which would usually be derived from dose-range finding experiments up to maximum tolerated dose, followed by the definitive study with an additional equally spaced five dose groups plus a solvent control group (Lutz and Lutz, 2009). Additionally, the Td from hockey stick model is influenced by effects of the top dose, indicative of sub- or supra-linearity effects (saturation or high dose toxicity) (Gollapudi et al., 2013b).

A follow up study by Johnson et al. (2014) compared the Lutz hockey stick model with other bi-linear dose-response models. Johnson et al. (2014) introduced a similar bilinear model to the Lutz hockey stick model, called the segmented model (Muggeo, 2008), which also determines the best-fitting bi-segmented linear model where the first segment from zero dose to inflection point is horizontal, and the second segment with a positive slope. However, unlike the Lutz hockey stick model, the segmented model does not require iterative removal of the top dose due to supra-linearity (Johnson et al., 2014a). With the segmented approach, the inflection points where the dose-response changes are called the 'breakpoint dose' (BPD)

– synonymous to the threshold dose introduced by Lutz. Similarly, 90% confidence intervals can be calculated associated with the BPD that can be interpolated to the PoD. Johnson et al. (2014) also utilized penalized smoothing splines, which is a set of flexible statistical techniques that does not assume linearity or any specific non-linear function. The smoothing splines analysis is advantageous because it can be applied to datasets such as cell or colony counts and incidence frequencies that are generally non-normally distributed data with heterogenous variances (Johnson et al., 2014a).

Johnson and colleagues (2014) used the methods to compare the PoDs that were derived from each bi-linear statistical model and concluded that the PoDs are heavily influenced by study design including dose selection and spacing, consistent with other researchers findings. An advantage is that the PoDs are more conservative than those derived from the NOEL models, which may complement available mechanistic data to support a threshold effect. However, the influence of study design on the derivation of the PoD is far from ideal and fails to address the same concern that Crump (1984) expressed with the NOEL approach. Moreover, biostatisticians argue that dose-thresholds in a strict sense are implausible in toxicological dose-response relationships, and that the question toxicologists should be asking is "what size of an effect can the organism subsist, and when is the effect adverse?" (Slob, 1999).

1.3.3.1 Use of the Threshold Dose-Response Model in Industry: The Viracept Story

The pretext for modern genotoxicity risk assessment was set with a high-profile case in the first decade of this millennium. This introduction chapter would be incomplete without acknowledging the contamination of the HIV protease inhibitor, Viracept (nelfinavir) with a well-known DNA-alkylating mutagen, ethyl methanesulfonate (EMS) which led to a global recall by the manufacturer and legal Market Authorisation Holder (MAH), pharmaceutical company Hoffman-La Roche (herein referred to as Roche) in June 2007 (Müller and Singer, 2009). During the time of the contamination issue, all compounds in the DNA alkylating agent chemical class were generally assumed to demonstrate a linear dose-effect relationship for carcinogenic risk, where virtually no "safe-dose" of exposure exists (Bos et al., 2004; Müller et al., 2006). Regulators considered that levels of the EMS impurity had to be controlled at a level below the Threshold of Toxicological Concern (TTC) for lifetime excess carcinogenic risk not greater than 1 in 100,000 to 1 in 1 million (Lutz, 2009; ICH, 2013). At the time, EMS was considered a Class 2 impurity where genotoxic potential was understood,

however carcinogenic potential was unknown (ICH, 2013). Subsequently, regulatory authorities considered that Viracept tablets contained more than 1,000 times the permissible levels of EMS and hence posed an unacceptable carcinogenic risk to human health (Müller and Singer, 2009).

However, the TTC limits imposed by the regulatory authorities are based on linear dose-response assumptions. The *in vitro* genotoxicity research that emerged from Swansea University academics that refuted linear dose-response relationships for EMS and MMS (discussed in detail in section 1.2.1 of this introduction) encouraged scientists at Roche to pursue a quantitative risk assessment of Viracept EMS contamination. Thus, Roche received regulatory authority support to initiate *in vivo* studies to define a threshold for EMS mutagenicity and clastogenicity suitable extrapolation to human exposure assessment. To this end, EMS was studied alongside ENU by daily oral-gavage exposure in an acute 7-day mouse bone marrow micronucleus study, and in a 2-armed subchronic 28-day daily repeat exposure and acute single exposure lacZ gene mutation test using CD2-lacZ80/HazfBR mice (Muta^{TM}Mouse assay) (Gocke et al., 2009). Bone marrow preparations were made on day 8, approximately 24-hours following the terminal exposure in the 7-day mouse micronucleus test. Bone marrow, liver and small intestine (duodenum/ileum) tissue were taken for analysis following necropsy on either day 31 of the 28-day exposure, or either days 7 or 31 of the acute exposure in the Muta^{TM}Mouse study.

The results of the studies were analysed using the bilinear hockey stick model (discussed earlier) to determine a potential threshold effect level (Td) (Gocke and Wall, 2009). The *in vivo* genotoxicity study results were consistent with the findings from Doak et al. (2007), where EMS displayed a distinct threshold dose response for genotoxicity. Ultimately, the experimentally derived mouse NOELs were used to calculate a Permissible Daily Exposure (PDE) of 104 μg/day (Müller and Gocke, 2009). Scientists at Roche were able to demonstrate that patients potentially exposed to 0.055 mg/kg/day EMS by oral administration of Viracept, were exposed to levels of EMS clearly below the threshold exposure level for genetic damage (Müller et al., 2009).

The series of studies performed by scientists at Roche and Swansea University were early examples of successfully utilising genotoxic mode of action information together with dose-response information to quantitatively risk assess human exposure to DNA reactive compounds.

1.3.4 The Benchmark Dose-Response Model

When Crump originally raised valid objections to the NOEL approach, he responded by introducing a new mathematical dose-response model called the "Benchmark Dose" (BMD) as an alternative to the NOEL (Crump, 1984). The BMD was defined by Crump as "a statistical lower confidence limit on the dose corresponding to a small increase in effect over the background level, somewhat in the order of a 10% change".

1.3.4.1 Data Types

It is important to note the main types of data that can be produced by toxicological studies for dose response analysis. Quantal data in toxicological studies are typically animal incidence data, for example the number of animals with tumours are examples of quantal data. Continuous data types in toxicological studies specify the degree of the effect, for example organ weights or relative percent of micronucelated cells in the bone marrow are examples of continuous data. Genotoxicity data is of the continuous data type, so the focus of this section is on the same data type.

Continuous dose-response data includes the dose-levels and the associated response levels, where there will be inherent variation about a nonzero value in the control group (Crump, 1984). The BMD model supposes that the response values are normally distributed, following the Central Limit Theorem of probability. This has practical implications in the data analysis of results from toxicological studies since if a non-normal distribution were assumed, knowledge of the individual response values in the test systems (e.g., animals, cells, etc.) would be required, which is seldom possible. Moreover, Crump explains that the choice of the normal distribution is unimportant since it only governs the error structure, without influencing the dose response (Crump, 1984). This is impactful since it implies that the dose-response relationship can be defined independently from the experimental error.

1.3.4.2 Dose-Response Modelling of Continuous Endpoints

As previously mentioned, the BMD model includes fitting a dose-response model to the toxicological data and estimating the dose at a given level of response. Factors such as sample size and shape of the dose-response curve are accounted for in the BMD approaches mathematical models (Crump, 1984), which is advantageous over the NOEL and Td approaches. The BMD approach estimates the measurable effect (benchmark response,

BMR) and the critical effect (benchmark dose, BMD), where the BMR for continuous data corresponds to a biologically relevant percent change compared to the background response level (e.g., 50%, 10%) (Macgregor et al., 2015). The BMR is a pre-requisite for BMD modelling, and is a value set by the toxicologist that is determined to be a biologically meaningful change in the size of the effect that is acceptable or non-adverse (Slob, 2002). The BMR and BMD terminology is often interchangeable with CER (critical effect response) or CES (critical effect size), and CED (critical effect dose) that was introduced by (Slob, 2002) to be specific to BMD modelling of continuous data types. As mentioned, since this thesis presents continuous genotoxicity data, the acronym CES will be used from hereon in. BMD will be used interchangeable with CED throughout this thesis, since CED appears as an acronym in the BMD modelling software described in the general materials and methods section (chapter 2.3), and BMD is commonly referred to in literature and guidance.

Before the BMD (or CED) is estimated, the BMD approach fits a family of mathematical models that best describe the dose-response. It has been postulated that no single dose-response model can fully describe a given dose-response according to the underlying biological mechanisms, and hence descriptive models are more appropriate. The mathematical models are described by (Slob, 2002) and summarised in the general materials and methods section of this thesis (chapter 2.3).

1.3.4.3 The BMD Approach

As previously mentioned, the BMD is the dose level estimated from the fitted dose response curve of the family of nested exponential or hill models. The BMD is associated with a biologically relevant and specified change in response, the critical effect size (CES). Model uncertainty is accounted for when fitting the models to the dose response dataset, which is the statistical confidence level, usually the 95% confidence (EFSA, 2017). The term 'BMDL' corresponds to the lower limit of the one-sided 95% confidence interval of the BMD (Gollapudi et al., 2013b). For example, when related to a CES of 10%, the $BMDL_{10}$ for continuous endpoints corresponds to the lower 95% confidence limit of the dose that produces a 10% increase in response compared to the concomitant background response, or control values (Gollapudi et al., 2013b). These key concepts are depicted in **Figure *1.9*** and further described in the figure legend.
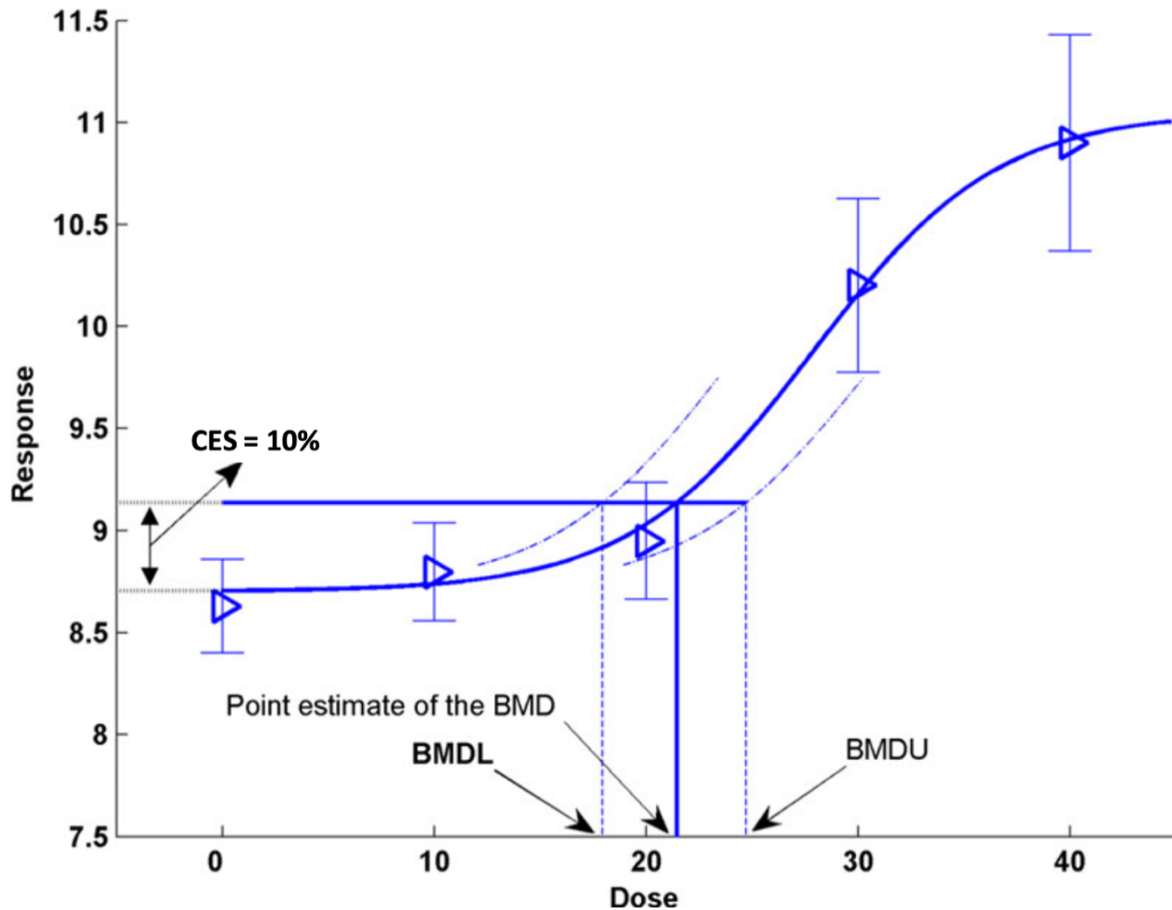
Figure 1.9. The key concepts of the BMD approach are shown on a hypothetical fitted dose response curve. The mean response values are shown as triangles, with their associated confidence intervals. The solid blue curve is the fitted dose response model. The point estimate of the BMD is calculated associated with the given 10% CES. The dashed curves represent the model uncertainty and correspond to the lower and upper 95% one-sided confidence bounds, respectively. The intersection of the dashed curves with the horizontal line forms the lower and upper bounds of the BMD, corresponding to the BMDL and BMDU, respectively. Image modified from EFSA (2017) to change the acronym BMR to CES with a value of 10%.

The confidence limits around the BMD will be tighter for better studies and models (Haber et al., 2018), and the associated BMDL will reward better designed studies. The BMDL can be used as the PoD in lieu of other dose response metrics (Td, NOEL/NOAEL) in reference dose setting in safety assessments. The previously cited publications by Gollapudi et al. (2013b) and Johnson et al. (2014a), as well as a publication by Macgregor et al. (2015) unanimously support the BMD approach as the favoured dose response modelling approach

for quantitative datasets. Their research compared the PoDs derived from the different available dose response models (NOEL, Td, and BMD) and concluded that the BMD approach is optimal for analysing genotoxicity dose response datasets; and the authors re-emphasised the advantages stated by biostatisticians (Slob, 2002; Slob and Setzer, 2014) that the BMD approach has additional benefits over other methods, including the fact that the BMD can be any dose when considering the full dose response, conveyance of uncertainty measurements, and flexibility in parameterisation.

Despite overwhelming support from expert groups, the BMD approach suffers from lack of adaption. Whilst the BMD approach is conceptually simple, as shown by **Figure *1.9***, several reasons impede its universal use. Firstly, calculation of a BMDL requires mathematical interpretation using dose response modelling software. Although the BMD modelling softwares are user friendly and freely available (e.g., BMDS, PROAST), advocates for identifying a NOEL rely on manually visualising the dataset (Haber et al., 2018), which is a rather unsophisticated approach and poor reason against using a more statistically robust method. Another issue relates to some study datasets not being amenable to BMD analysis, where only mean values are reported without the group-size or measurements of variability. Haber et al. (2018) reasons that said studies are inherently low quality and unworthy of use in risk assessment. For most genotoxicity continuous datasets of high quality, metrics are generated that allow the application of the BMD approach (Gollapudi et al., 2013b; Johnson et al., 2014a).

A more worthy debate over the universal implementation of the BMD approach includes the choice of the CES. This is a crucial issue which is the focus of most hesitancy to rely on the approach for reference dose setting. Regulatory bodies avoid providing fixed CES values, however EFSA (2017) and the US EPA (2012) agree that the choice of CES for continuous data shall be biologically meaningful and defendable. However, the two regulatory agencies disagree on what change in response is determinably biologically adverse. The US EPA (2012) recommends that a change corresponding to 1 standard deviation (SD) in the control response may be an appropriate CES value. Haber et al. (2018) explains that this is based on the work by Crump (1995) who demonstrated that when responses are normally distributed with control variation within 1%, and exposure results in a shift of the normal distribution, that changing the mean response by 1.1 SDs results in abnormal response levels in an additional 10% of the test animals. EFSA are somewhat less

flexible, with a recommendation of a 5% change in the mean response being biologically relevant. Haber et al. (2018) indicate that the 5% value is since the $BMDL_{05}$ approximates to the NOAEL from a re-analysis of National Toxicology Program (NTP) studies. EFSA postulates that a higher CES value may be appropriate for endpoints with larger within-group variability, which is similar to the US EPA 1 SD approach (Haber et al., 2018). These recommendations have been criticized, as summarized by White et al. (2020). The authors state that in endpoints with low response variability, it is unlikely that a 1SD change from control can be deemed adverse. On the other hand, 1SD in high control variability endpoints could yield large CESs and higher corresponding BMDs that are unfavourable from a regulatory standpoint. The same concerns are applicable to genotoxicity endpoints, and some authors note that 10% is too low for said endpoints. Zeller et al. (2016) proposed a novel method to determine endpoint specific CESs by scrutinizing the historical background response values before defining what a biologically meaningful response would be. The researchers determined *in vivo* chromosome aberration and mutagenicity endpoint specific CESs in the range of 34-76%. A similar theory presented by Slob (2017) determines how endpoint specific CESs can be determined by utilizing the dose response models parameters, specifically *c* (maximal fold change), and the within-group variation. Slob (2017) presented results of endpoint specific effect size determination from 27 endpoints, which included *in vivo* micronuclei frequency, Pig-a mutation frequency, and transgene mutation frequency in BigBlue® rat liver, and found that all CESs should be much higher than the values recommended by EFSA and the US EPA. Slob (2016) found endpoint specific CESs in the range of 45-80% are more statistically defensible. Although the approaches proposed by Slob (2016) and Zeller et al. (2017) offer statistically robust options to determine an appropriate CES, White et al. (2020) states that it requires the analysis of large datasets containing dose response data from a variety of studies and compounds, which to date, does not exist.

The choice of CES for determining a reference dose for use in safety assessment is therefore crucial and should be justified. To date, only data from *in vivo* studies are suitable for extrapolation to a human based safety guidance value since there are too many inconsistencies in *in vitro* studies including differences in toxicokinetic and tissue-specific metabolism (Macgregor et al., 2015). However, the extent to which the CES impacts the outcome of BMD modelling of *in vitro* endpoints remains understudied. Since *in vitro* endpoints are not considered applicable to PoD extrapolation, it is reasonable to surmise that the choice of CES is less important if the tested hypothesis is not impacted to a determinable

extent. One published study by Bemis et al. (2016) performed BMD modelling to compare micronucleus frequency responses between *in vitro* and *in vivo* systems. The researchers utilized various CES values including 5, 50, 100, 200, and 400% to inform the BMD values at various points along the dose-response curves. The resulting BMDs and their confidence intervals were used to rank the compounds and provide a comparison between the two systems, and the authors demonstrated that the choice of CES did not influence the outcome of the system cross-correlations.

1.3.4.4 BMD Modelling of *In Vitro* Genotoxicity Endpoints

Most of the focus of the BMD approach has been on the advantages and disadvantages of using the approach to determine a PoD from *in vivo* toxicology/genotoxicity studies for application to human risk assessment. However, the use of BMD confidence intervals has been used to define relative compound potency. The term 'potency' in compound genotoxicity ranking refers to the comparison of BMD metrics between compounds, or test systems, depending on the hypothesis being tested.

Research performed by individuals at the Dutch National Institute of Public Health and the Environment (RIVM) utilised the potency information derived from BMD estimates and their confidence intervals (CIs) and applied it to compound potency ranking within endpoints, as well as comparative potency analyses across endpoints (Hernandez et al., 2010; Hernández et al., 2011a; Hernández et al., 2011b; Soeteman-Hernández et al., 2015; Soeteman-Hernández et al., 2016). These studies by RIVM highlight that comparative BMD potency correlations are a statistically robust and efficient mechanism to compare compound potencies between systems. The same researchers built computational algorithms that permit combined analysis of multiple dose-response datasets for a shared endpoint. In doing so, multiple BMDs can be defined for dose-response relationships that vary by a covariate (e.g., compound, sex, cell type, exposure scenario, etc.), known as the BMD combined covariate approach (Slob and Setzer, 2014). The referenced studies demonstrate that comparative potency analysis of *in vivo* genotoxicity endpoints such as *in vivo* micronucleus assay may be used as a predictor for carcinogenic potency of the same compound.

With respect to performing BMD analysis of *in vitro* genotoxicity studies results, several publications report on using the BMD confidence intervals to inform *in vitro* genotoxicity compound potency. Different studies by Soeteman-Hernández et al. (2015) and

Bemis et al. (2016) compared BMD confidence intervals of compounds with varying genotoxic MOAs from the *in vitro* and *in vivo* micronucleus endpoints and concluded that the potency of a compound from the *in vitro* micronucleus endpoint may be used to predict the potency of the same compound and endpoint for the *in vivo* assay. A prominent study by Wills et al. (2016a) used the BMD combined covariate approach to analyse, of interest, *in vitro* micronucleus frequencies for Chinese hamster ovary (CHO-K1) cells, primary human lymphocyte (HL) cells and human lymphoblastoid cell line AHH-1 cells exposed to well-characterised aneugens. The authors demonstrated that the BMD combined covariate approach provides a rigorous potency ranking for the series of aneugens that were examined *in vitro*, where the selected aneugens yielded finite confidence intervals except for rotenone, diethylstilbesterol, and thiabendazole, which exhibited infinite upper bounds. The confidence interval potency rank plot obtained by Wills et al. (2016a) is shown in **Figure *1.10*** below.
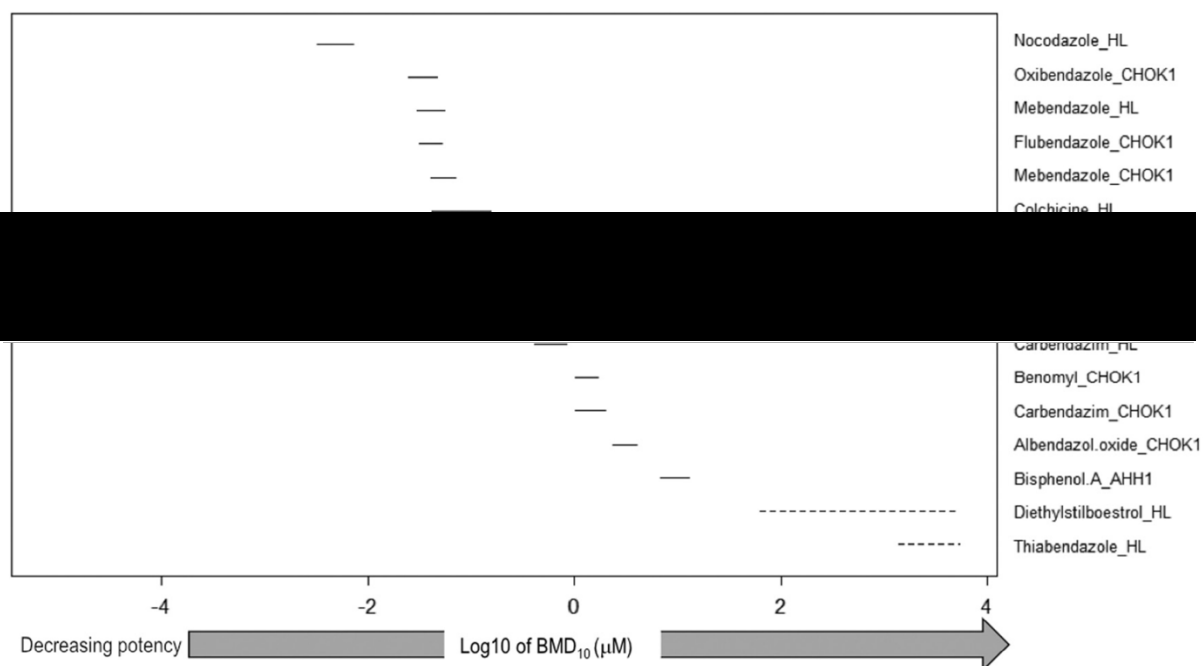
Figure 1.10. BMD confidence interval potency ranks of *in vitro* micronucleus responses from the benzimidazole compounds relative to colchicine and bisphenol-A. Finite confidence intervals for most compounds bar rotenone, diethylstilbesterol and thiabendazole. Compounds with infinite confidence interval bounds are shown with dashed lines across the potency plot. Abbreviations: HL = human lymphocytes; CHOK1 = Chinese hamster ovary cells; AHH1 = human lymphoblastoid. Figure taken directly from Wills et al. (2016a).

One of the observations that Wills et al. (2016) made was the degree of overlap/non-overlap present in the potency plot for the aneugens. The authors highlighted that the confidence interval represents the range in which the true BMD lies, and hence one can only statistically defend potency differences when there is no apparent overlap between intervals. For the aneugens presented in **Figure *1.10***, Wills et al. (2016) scrutinised the potency ranks in relation to the compound structure and the mechanism of action of the compound. As an example, the benzimidazole compounds, oxibendazole, flubendazole, and mebendazole all have large moieties in the 5-position that can competitively inhibit colchicine binding, that the other benzimidazoles lack. The authors suggest that the differences in compound structure relate to the apparent differences in potencies observed from BMD combine covariate comparative potency analysis. This study was the first published example of using BMD analysis to relate compound mode of action to genotoxic potency and supports that *in vitro* genotoxicity data maybe suitable for potency ranking and mode of action determination.

One of Wills et al. (2016) major concluding remarks highlights that *in vitro* genotoxicity data is more useful than providing a binary risk characterisation of whether a compound is genotoxic or not. This is due to the suitability of the data for the interpolation of *in vitro* genotoxicity potency metrics, which may be useful in screening and prioritisation. Additionally, the authors conclude that BMD confidence intervals can serve as a starting point for developing compound groupings when supported by mechanistic data for read across and data gap filling. However, the study uses good quality data from a well characterised assay and uncertainties exist on how the same theoretical approach could be expanded to other assay types, and for other classes of compounds.

1.4 Modern Genotoxicity *In Vitro* Test Systems

Johnson et al. (2014b) reviewed the status of genotoxicity test systems and highlighted the critical limitations of standard assays used in genetic toxicology, including low sensitivity and specificity, and how the assays are used to dichotomise genotoxic hazard. Considering the growing appreciation of quantitative analysis and interpretation of genotoxicity test data, including quantitative dose-response analysis endorsed by the same authors, Johnson et al. (2014b) underlined the acute need for more reliable experimental methods which produce high quality data suitable for quantitative dose-response analysis. To this end, Johnson et al. (2014b) noted that the most promising advancements include the development of flow cytometric assays that benefit from miniaturisation and high throughput

automation, allowing the rapid scoring of high quantities of cellular content. The specifics of some select assays will be focussed on in the next few sections, particularly those that are cellular based and use novel biomarkers to quantify different types of DNA damage. Other modern assays have been developed which evaluate toxicogenomic signatures (Li et al., 2015; 2017), however these assay types are out of scope of this thesis.

1.4.1 High Content Flow Cytometric Assays

Three broad types of assays have added enormous value to the genotoxicity scientific literature over the past one to two decades. They include Litron's *in vitro* micronucleus assay (named MicroFlow®), and the *in vitro* DNA damage response assay (named MultiFlow®) by the same company; and Toxys' *in vitro* stem-cell gene-expression based assay for the detection of DNA damage, oxidative, and cellular stress (named ToxTracker). The assays all have one major advantage in common: that fluorescent biomarker responses are scored and analysed using automated flow cytometry, which offers the benefit of being able to score thousands of cells almost instantaneously (Adan et al., 2017) compared to the low throughput manual read out genotoxicity test methods.

The basics of flow cytometry have been reviewed in (Henel and Schmitz, 2007) where it is described that the essence of flow cytometry is the coordinated use of 3 components: the fluidics, the optics, and the computer system. Firstly, the fluidics system carries cells through the cytometer. Hydrodynamic focussing involves a pressurized stream of fluid into which cells are injected and sent single file as they pass through a flow cell. Secondly, as the cells pass through the flow cell, they intersect 1 or 2 lasers. Upon reaching this interrogation point, the cell's attributes are analysed based on the ability of the cells to scatter light and emit fluorescent light if the cells are labelled with fluorescently tagged reagents. The cell scatters the laser light which is collected along the same axis of the laser (forward scatter), and at a 90° angle to the beam (side scatter). In addition to the lasers, the optics system also incorporates several other elements including a forward scatter photodiode that collects forward scatter laser light, and a series of filter and mirrors directed to multiple photomultiplier tubes that detects side scatter laser light. The flow cytometers can be configured with combinations of detectors and filter sets that are specific to wavelengths of light. The photomultiplier tubes collect photons of laser light and generate an electrical signal that is sent to the flow cytometers computer for processing – where the generated signals are

proportional to the amount of scattered light or fluorescence detected by the photomultiplier tubes. The basic structure of the flow cytometer is shown in **Figure *1.11***.
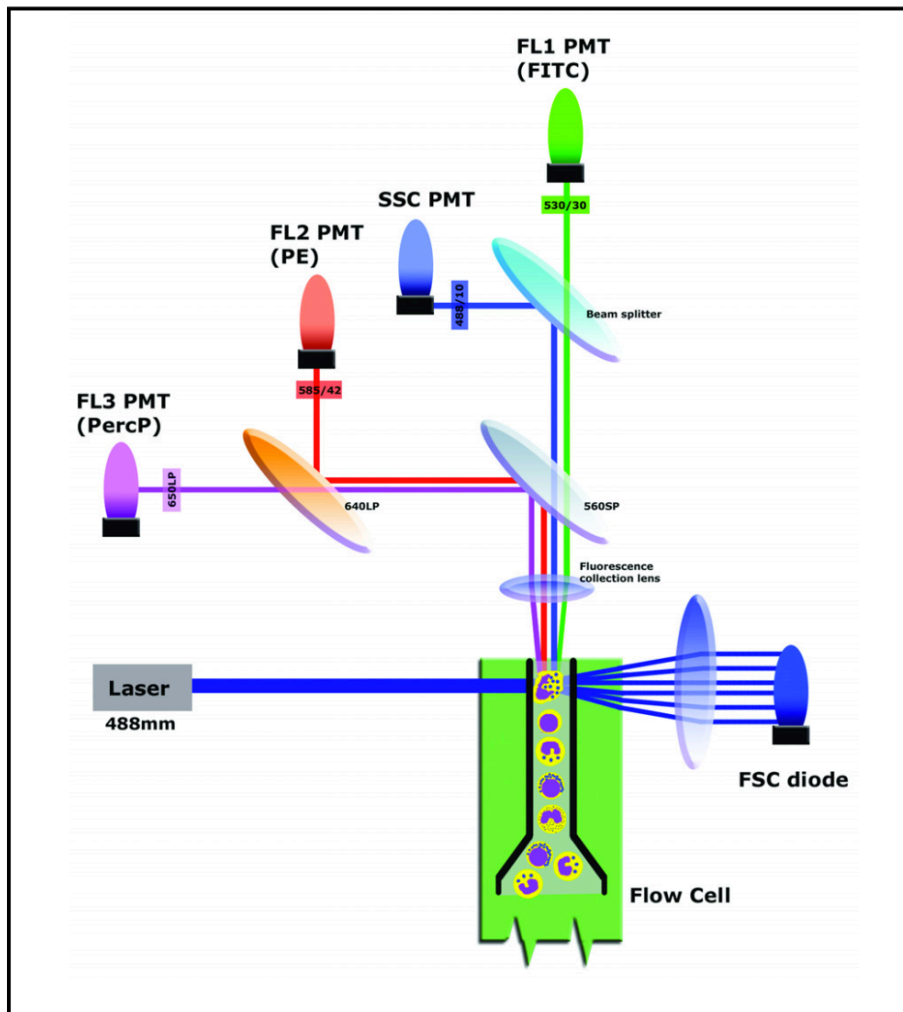
Figure 1.11. Depiction of a flow cytometer's fluid cell and optics system with a single laser. The cell scatters light in the forward direction to a forward scatter (FSC) diode, and at a 90° angle through a series of filters to photomultiplier tubes (PMT) specific to different wavelengths of light (SSC, FL1, 2, and 3). Figure adapted from Henel and Schmitz (2007).

The forward and side scatter, and fluorescent data are stored as parameters and referred to as 'list mode'. A process called 'gating' involves the identification of a subset of a population of cells based on the expression of markers. The data output is generally displayed as single-parameter histograms, or dual-parameter (bivariate) dot plots. **Figure *1.12*** shows an example of a bivariate output of list mode and gating data of flow cytometric analysis of white blood cells to identify and quantify subsets of the cellular population, which is a common use for flow cytometry in clinical haematology laboratories.
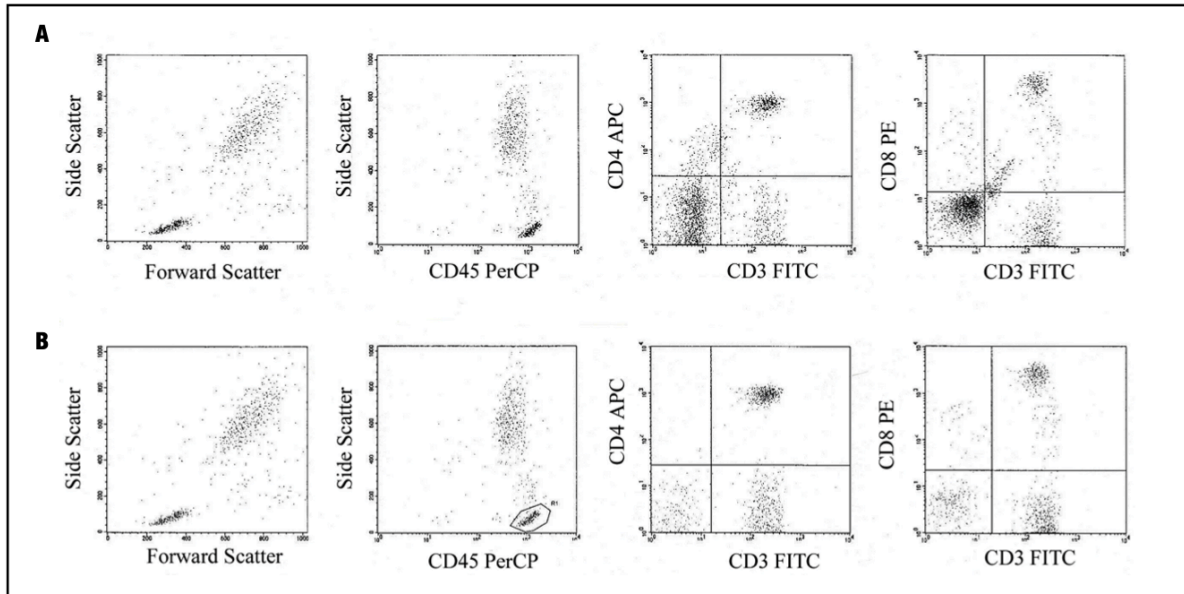
Figure 1.12. Bivariate plots from flow cytometric analysis of white blood cells. List mode and 'gating' identifies subsets of the population. For example, subsets including CD3, CD4 and CD45 positive cells have been quantified using the responses fluorescence-based labelling. Figure adapted from Henel and Schmitz (2007).

The three flow cytometric assays (MicroFlow, MultiFlow, and ToxTracker) are introduced in the next three subsections, and data from the assays are used to test the hypotheses presented in this thesis.

1.4.2 Litron's MicroFlow *In Vitro* Micronucleus Assay

Flow cytometric methods that analyse micronuclei frequency offer numerous advantages over the manual microscopic method that preceded their inception. Litron laboratories introduced their enhanced method for detecting *in vitro* micronuclei *via* flow cytometric analysis from L5178Y mouse cells in 2006 (Avlasevich et al., 2006), and extended the same method in 2007 to metabolically competent human lymphoblastoid TK6 cells (Bryce et al., 2007). The flow cytometric method is described by (Bryce et al., 2007) which labels necrotic and mid/late-stage apoptotic cells with the ethidium monoazide fluorescent dye. Following a series of washing and cytoplasmic membrane stripping, cells are incubated with RNase and a nucleic acid dye (SYTOX Green), which provides a suspension of free nuclei and micronuclei that are differentially stained (EMA-/CYTOX+) relative to chromatin associated with dead/dying cells (EMA+/CYTOX+). Briefly, the gating strategy discriminates between nuclei and micronuclei from apoptotic chromatin by excluding EMA+

chromatin from analysis (Avlasevich et al., 2011). The method also incorporates latex beads which aim to serve as a measure of cellular survival (or cytotoxicity) and are readily distinguishable by discriminating between side scatter from forward scatter in the data output bivariate plot (Bryce et al., 2007). Cytotoxicity limits have been shown to be an important element in the micronucleus assays, as some agents induce spurious genotoxic effects due to their cytotoxicity (Lorge et al., 2008). In addition to the ease of scoring micronuclei, another advantage associated with the flow cytometric method include SYTOX fluorescence's ability discriminate cell cycle effects as shown in an example with etoposide exposure, where an increase in forward scatter is associated with different cell cycle stages as show in the histogram presented in **Figure** *1.13*.
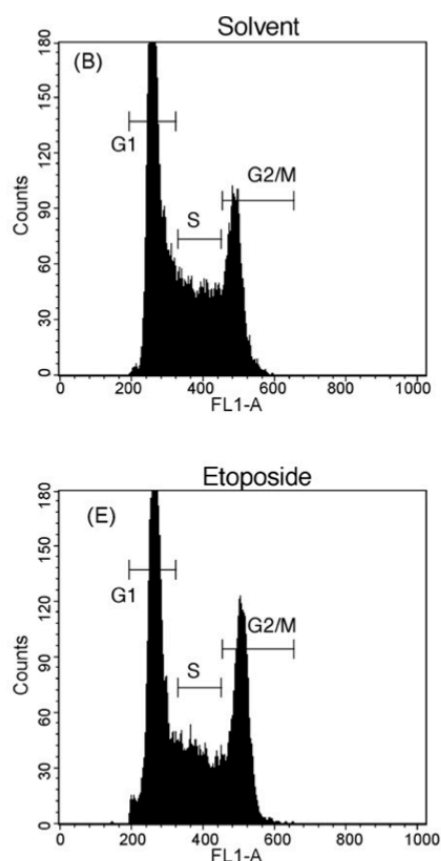
Figure 1.13. Quantitative determination of cell cycle distribution on a CYTOX fluorescence histogram. Figure adapted from Bryce et al. (2007).

Following the assays development, an international multi-lab trial for scoring micronuclei in mouse lymphoma L5178Y cells was conducted in 2008 including labs from Litron, Novartis, Johnson & Johnson and GlaxoSmithKline (Bryce et al., 2008). The laboratories studied a limited number of compounds including the genotoxicants mitomycin

c, etoposide, vinblastine, and the non-genotoxicants sucrose, staurosporine, and dexamethasone; with the latter two being challenging compounds due to their potent apoptotic potential. Key findings included 1) good concordance between flow cytometric results compared with manual scoring results of dose-dependent increases in micronuclei for the tested compounds; and 2) discrepancies amongst different methods for determining measures of cytotoxicity and the top dose to select for optimal sensitivity for micronuclei detection. However, the labs agreed that the latex counting beads performed the best compared to different measures of cytotoxicity.

Avlasevich et al. (2011) outlined some of the knowledge gaps associated with the use of flow cytometric analysis of micronuclei. One major challenge that the authors highlighted includes the relatively high false positive micronucleus events, caused by chromatin fragments produced during apoptosis. The authors credit that EMA+ events can filter out DNA originating from apoptotic cells, however it fails to differentiate between chromatin fragments that originate before membrane integrity has been compromised. The authors concluded that EMA labelling is not absolute in overcoming the false positive rate, and refinement of the cytotoxicity measures may improve the situation, including assessing cell-to-bead ratios pre- and post-treatment rather than nuclei to bead ratios. Avlasevich et al. (2011) continued to expand that for mammalian cell lines, the EMA/SYTOX staining strategy is well suited, however it was unclear at the time of the compatibility for the approach with primary cell lines. For example, the microscopic-based protocol includes treating cells with cytochalasin B to ensure cells have completed one mitosis, which benefits primary cells like human lymphocytes where only a small fraction of the population undergoes division upon mitogen stimulation. The recommendation by Avlasevich et al. (2011) was to include an additional fluorescent label to discriminate chromatin from divided cells, which may help overcome the issue related to low sensitivity from cell lysis.

Despite some of the valid concerns relating to the potential for false positive events, the MicroFlow assay offers a high content, high throughput and high precision platform (Johnson et al., 2014b) which adds tremendous value to laboratories that may screen high number of compounds for genotoxic potential (Lukamowicz et al., 2011).

1.4.3 Litron's MultiFlow *In Vitro* Cellular DNA Damage Assay

Although the *in vitro* micronucleus assay is regarded as a gold standard assay, including endorsement from OECD as an accepted test method (OECD, 2016), the issues surrounding false positivity rate remain a concern that has not fully been resolved in flow cytometric test systems for the same endpoint. Bryce et al. (2014) introduced a hypothesis that the *in vitro* micronucleus assay would benefit from alternative measures of cytotoxicity and the application of additional biomarkers that can distinguish between clastogenic or aneugenic activity. To this end, the same laboratory developed a separate high-throughput flow cytometric follow-up assay that includes a panel of biomarkers that are multiplexed into a single add-and-read assay in lysed TK6 cells.

Several publications from the assay developers describe the multiplexed flow cytometric assay in detail and explain that the combination of biomarkers are relevant to DNA damage response pathways and aneuploidy induction (Bryce et al., 2014; 2016; Bernacki et al., 2016; Bryce at al., 2017; 2018). Specifically, the so called MultiFlow DNA Damage Assay includes the following biomarkers: 1) the detection of DNA double strand breaks through phosphorylation of H2AX at serine 139 (γH2AX); 2) the identification of mitotic cells through the phosphorylation of histone H3 at serine 10 (p-H3); and 3) nuclear p53 content as an indication of nuclei counts which could be interpreted as treatment related cytotoxicity. Additionally, the use of gating logic allows the interpretation of polyploidy events (Bryce et al., 2016). Like the MicroFlow assay, MultiFlow also includes latex beads that aim to provide information related to direct cytotoxicity measures. The assay is compatible with microtiter wells including 96-well plates allowing for the study of multiple concentrations of several compounds in one single analysis.

Original validation studies with a diverse set of compounds including 31 clastogens (of which 4 require exogenous metabolic activation with rat liver S9), 14 aneugens, and 39 non-genotoxicants was published in 2016 (Bryce et al., 2016). The study evaluated exposure to test compounds over two timepoints, 4 hours, and 24 hours. Based on the *a-priori* mode of action of the training set compounds, increased γH2AX fluorescence was characteristic of a clastogen exposure with generally similar responses at the 4- and 24-hour timepoints. Interestingly, some *a-priori* non-genotoxic compounds were responsive at the 24-hour γH2AX timepoint. Aneugenic agents displayed an increase in p-H3 positive events at both the 4- and 24-hour timepoints, with a higher magnitude in the 4-hour timepoint. The p53

biomarker exhibited a time-dependent response profile where both clastogen and aneugen exposure resulted in modest increases in p53-associated fluorescence at 4-hours, dramatically increasing at 24-hours. The proportion of polyploidy nuclei follow the expected pattern where the early timepoint is unresponsive to a wide variety of compounds, however aneugens displaying a strong signal at the 24-hour timepoint.

The assay developers employed advanced data analysis strategies that make use of machine learning tools to create models that can predict the mode of action of the test compound (Bryce et al., 2016; 2018). Specifically, training sets with known mode of action information were incorporated into logistic regression, The laboratory concluded that the combination of biomarker responses can be used to predict the genotoxic mode of action of a compound. Despite advantages associated with the use of pattern recognition tools of the multiplexed assay, the original design of the assay included applications that maintain dichotomisation of genotoxicity to an extent, i.e., the compound is either a clastogen, an aneugen, or not genotoxic.

1.4.4 Toxys' ToxTracker *In Vitro* Fluorescent Reporters of DNA Damage

A parallel between the MultiFlow and the ToxTracker assays is that they both make use of biomarkers that respond to indirect measures of DNA damage. In the ToxTracker system, six genetically stable mouse embryonic stem cell lines are used which contain green fluorescent protein constructs that report on a distinct aspect of cellular perturbation. The assay's endpoints can be loosely categorised into the modes of action that they describe – namely DNA damage, p53 mediated cellular stress, oxidative stress, and protein stress (Hendriks et al., 2011; 2015). The cell lines eloquently make use of cell cycle control to report on the types of damage induced as shown in **Figure *1.14***.
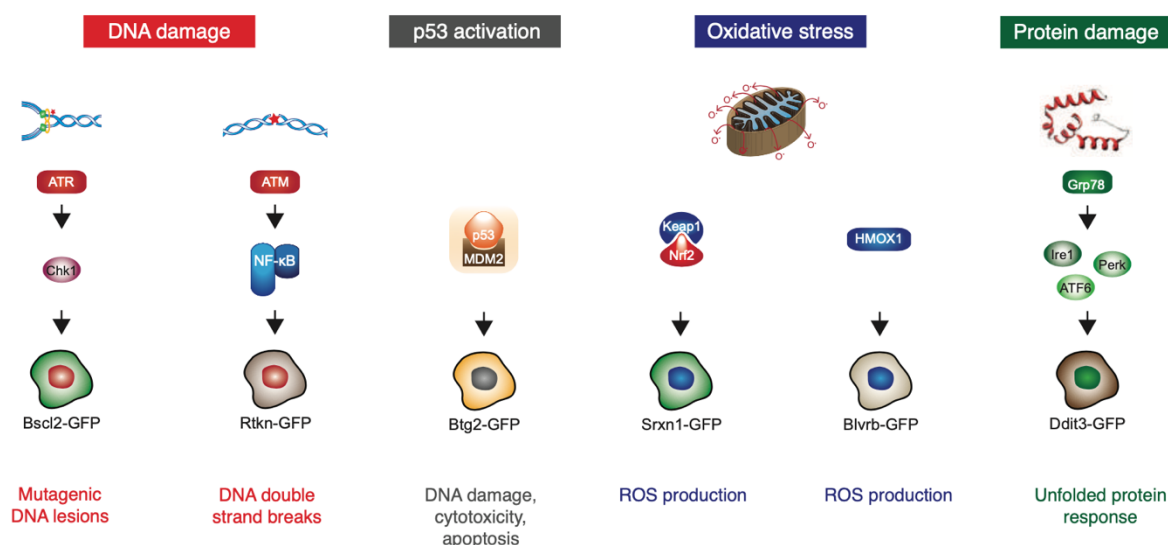
Figure 1.14. ToxTracker reporter cell lines and associated endpoints. Figure adapted from (Hendriks et al., 2015).

Briefly, with respect to DNA damage, two major signalling proteins associated with DNA damage response include the ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3-related protein (ATR) serine/threonine protein kinases. DNA double strand breaks recruit ATM for efficient repair, whereas ATR kinase becomes activated upon DNA replication stalling (Abraham, 2001; Reinhardt et al., 2009). Phosphorylation of checkpoint kinase 2 (Chk2) is ensued upon activation of ATM, whereas Chk1 is activated by ATR. Both Chk1 and Chk2 are involved in inhibiting cell cycle progression leading to apoptosis by inhibiting CDC25 phosphatase and the p53 tumour suppressor (Meek, 2009). The Berardinelli-Seip congenital lipodystrophy 2 (Bscl2) gene encodes a protein that was identified in Berardinelli-Seip congenital lipodystrophy patients that completely lack adipocyte differentiation. Hendriks et al. (2015) describes how induction of Bscl2 in mES cells in response to DNA damage is associated with the ATR-Chk1 DNA damage signalling pathway. A second biomarker, Rtkn, is involved in modulating the Rhotekin protein which is involved in cytokinesis, cell growth and transformation (Thumkeo et al., 2013). The Rtkn gene is upregulated upon cellular exposure to DNA damaging agents. Activation of Rtkn is associated with the NF-κB cytokine signalling pathway, where preferential upregulation versus ATR related DNA damage maybe indicative of aneugenic modes of action, demonstrated by the activation of NF-κB by mitotic spindle poisons. The B-cell translocation gene 2 (Btg2) expression is associated with the p53 response to DNA damage, which has a

pivotal role in cellular stress response. Btg2 is associated with regulation of the G1/S cell cycle checkpoint (Cortes et al., 2000).

The sulfiredoxin-1 protein is encoded by the Srxn1 gene and has an important role in the defence against oxidative stress by reducing oxidised cysteines in peroxiredoxins in the peroxisomes (Hendriks et al., 2011). Srxn1 is modulated by the nuclear factor (erythroid-derived 2)-like 2 (Nrf2) signalling pathway. Nrf2 is normally retained in the cytoplasm by interaction with the Keap1 protein, however upon increased levels of reactive oxygen species, Nrf2 becomes dissociated from Keap1 where it is translocated to the nucleus and is subsequently associated with antioxidant response element promoters. Conversely, Biliverdin reductase B is preferentially expressed upon oxidative stress by another related gene, Bvlrb, that does not contain an Nrf2 binding motif. Finally, a sixth marker monitors activation of the unfolded protein response. DNA damage-inducible transcript 3 (Ddit3) is associated with cell cycle arrest, apoptosis, and the endoplasmic reticulum stress response.

The combination of these biomarkers elucidates the genotoxic mode of action of tested compounds. (Hendriks et al., 2015) describes how whole genome transcriptional data of mES cells after exposure to various genotoxic agents associates with the responsiveness of the individual biomarkers. Stable cell lines are formed from the tagging of green fluorescent protein (GFP) reporters for Rtkn, Btg2, Blvrb, and Ddit3. Via flow-cytometry, the fold-change in GFP reporter induction is determined after exposure of test compounds to the reporter cell lines for 24 hours. As with the MultiFlow assay, the ToxTracker system offers the opportunity to generate high quality mechanistic data for with a variety of compound concentration exposures.

1.5 The 21st Century Genotoxicity Risk Assessment Framework

Despite the efforts of genotoxicity experts in the above referenced works to convey state-of-the-art science, the regulatory landscape regarding genotoxic hazard remains risk averse as reflected in regulatory guidance. This is in part likely due to the level of public concern pertaining to genotoxicity and potential cancer risk. The standard genotoxicity test battery remains common practice in drug development. Additionally, despite growing appreciation of statistically robust dose-response modelling, regulatory authorities continue to enforce manufacturers to control mutagenic impurities below compound specific Acceptable Intake (AI) levels, Permissible Daily Exposure (PDE) levels, or TTC limits (ICH,

2013). The Viracept® contamination case studies presented above concerns compounds with well characterized genotoxic modes of action that are electrophilic and "direct" DNA reactive molecules. Following workshops at the 6th International Workshop on Genotoxicity Testing (Iguazu, Brazil, 2013) and the Health and Environmental Science Genetic Toxicology Testing Committee annual meeting (Washington DC, USA, 2015), a pivotal article by Dearfield et al. (2017) outlined the "Next Generation Testing Strategy for Assessment of Genomic Damage". The authors acknowledge that novel small molecules, gene editing tools, and biologics are being developed because of advances in drug development technology and the need to treat life-threatening pathologies. The authors describe that relating to genotoxicity, novel molecules are being engineered that specifically lack electrophilic DNA reactive moieties, however non-electrophilic compounds may induce genotoxic damage by mechanisms that are not mutagenic, clastogenic or aneugenic. The field of genotoxicity must advance synergistically with regulatory acceptance. To this end, Dearfield et al. (2017) outlined a framework that involves a paradigm shift from a "one size fits all" dichotomous genotoxic hazard identification approach, to a more flexible approach that realises the underlying genotoxic mechanisms that affect adverse outcomes. The framework encompasses already accepted risk assessment approaches that are commonplace to other disciplines in toxicology. Key elements of risk assessment include planning and scoping, quantitative assessment of dose-response results, and exposure assessment. The overall aim is to realise the risk of humanly relevant exposures, with preference to data obtained from *in vitro* studies or *in vivo* studies where the minimum number of animals are used. The goal is ambitious, and the next 3 subsections evaluate some of the key elements proposed in the next generation risk assessment approach.

1.5.1 Quantitative Structural Activity Relationships and Read Across

(Quantitative) Structure Activity Relationships [(Q)SAR] utilise *in silico* systems to predict potential compound toxicity. Within the available (Q)SAR software models, some use expert-based rules to predict toxicity in a non-quantitative SAR manner, whilst others assign a probability of an adverse outcome – responsible for the quantitative definition of the model (Barber et al., 2017).

Expert SAR models are a form of artificial intelligence where toxicity information is collated and organised into a knowledge base and used to replicate the decision-making that human experts would make (Tripathi, 2011; Barber et al., 2017). Successful expert SAR

models are advantageous since the output is a transparent format that can be understood, challenged, and judged by the user. Such systems have been applied with success for the prediction of mutagenicity by the collation of 'structural alerts' into software that identifies the moieties and presents the user with a prediction and reason. Freely available software examples include Toxtree (Benigni et al., 2008), and the OECD QSAR Toolbox (Dimitrov et al., 2016). Examples of proprietary QSAR software used in industry are Lhasa Limited's Derek Nexus and Sarah Nexus, that perform expert rule based, and statistical probabilistic model-based estimation of genotoxic potential, respectively (Barber et al., 2017).

Some of the benefits that (Q)SAR models have brought to industry include an unbiased read-out for genotoxic potential, and the ability to translate the results into screening and prioritisation. On the latter topic, Dearfield et al. (2017) describes that the (Q)SAR prioritisation outcome could be used concomitantly with *in vitro* testing, and more specifically identify the appropriate *in vitro* assays that would establish a detailed quantitative understanding of genotoxic potential. Dearfield et al. (2017) elaborates more on this point and highlights specific examples of where (Q)SAR models have successfully positively impacted industry. One specific case exists for the pharmaceutical industry where impurities can be screened *in silico* for bacterial mutagenic potential using results from two complementary (Q)SAR models, one being an expert rule-based, and another statistical based (Amberg et al., 2016). However, since bacterial mutagenicity is only applicable to compounds that are primarily direct DNA-reactive, Dearfield et al. (2017) implies that models built upon compounds that are positive in bacterial mutagenicity assay do not provide sufficient coverage of the totality of potential DNA damaging effects. On a limited case-by-case basis, the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) regulatory framework accepts information from (Q)SAR models where the weight of evidence, supplemented with *in vitro* test results may waive the necessity to perform *in vivo* testing (REACH, 2007).

Another application of (Q)SAR results is to the domain of analogue searching (read across) that relies on relating a compound of interest to structures and functional groups of related compounds with a toxicological endpoint of interest (Grace et al., 2017). In other words, if toxicity information is available for a similar compound, said toxicity information can apply to the compound of interest. Dearfield et al. (2017) states that read across methods can be qualitative (i.e., related substance A is mutagenic due to X functional group, so

substance B is also likely to be mutagenic due to the presence of the same functional group), or quantitative where the known potency of a related compound is 'read across' to the compound under investigation. Such quantitative approach presumes that compounds with similar functional groups exert similar toxic potency and would share a similar PoD on a dose-response curve.

Although it is encouraging to see the adaption of QSAR models in industry, most read-across approaches are applicable to ecotoxicology and human health hazards including skin sensitization (REACH, 2007). The practical application of the models to genotoxicity assessment still dichotomises genotoxic potential into the hazard assessment framework, with restrictive low dose control limits (ICH, 2013). Although Dearfield et al. (2017) theorises the use of QSAR models for genotoxic potency read-across, little has been published where a QSAR model makes use of dose-response information to inform relative genotoxic potency of compounds within a compound class of interest.

1.5.2 Mode of Action Information

An integral piece of the next generation approach hypothesised by Dearfield et al. (2017) is the use of mechanistic understanding of biological pathways involved in toxicity, dubbed Mode of Action (MOA). The International Programme on Chemical Safety (IPCS) from the World Health Organisation (WHO) and the ILSI Risk Sciences Institute (ILSI RSI) provides a framework for establishing and using MOA information in a structured approach involving a series of key events along a causal pathway to a pathological or disease outcome (Boobis et al., 2006; 2008).

MOA is closely related to the growing field of adverse outcome pathways (AOPs) in genetic toxicology, with the two approaches being distinct where MOA is chemical or agent specific, whereas AOPs can apply to groups or sub-groups of compounds (Dearfield et al., 2017). The AOP programme was launched by the OECD in 2012 and defines AOPs as structured sequences of events that lead to an adverse health effect (Yauk et al., 2015). The application of AOPs to genotoxic MOAs have been reviewed by Sasaki et al. (2020). The publication was a result of a 2017 workshop at a HESI GTTC workshop held in Delaware (USA) which established work groups that would develop and publish AOPs relevant to the following genotoxic MOAs: aurora kinase inhibition and tubulin binding that lead to aneuploidy; and inhibition of topoisomerase II, production of cellular reactive oxygen

species, and inhibition of DNA synthesis, that lead to mutation and/or chromosome breakage by indirect DNA interaction. For each of the AOPs, the starting point is the Molecular Initiating Event (MIE) which is the initial interaction of a compound with DNA or its molecular machinery. A series of Key Events (KEs) and Key Event Relationships (KERs) describe the progression of the biological response, where the KEs are directly or indirectly measurable, and the KERs can optimally be quantitatively described by dose-response relationships if data is available. The key concepts of the AOP approach are visualised in **Figure** *1.15* below.
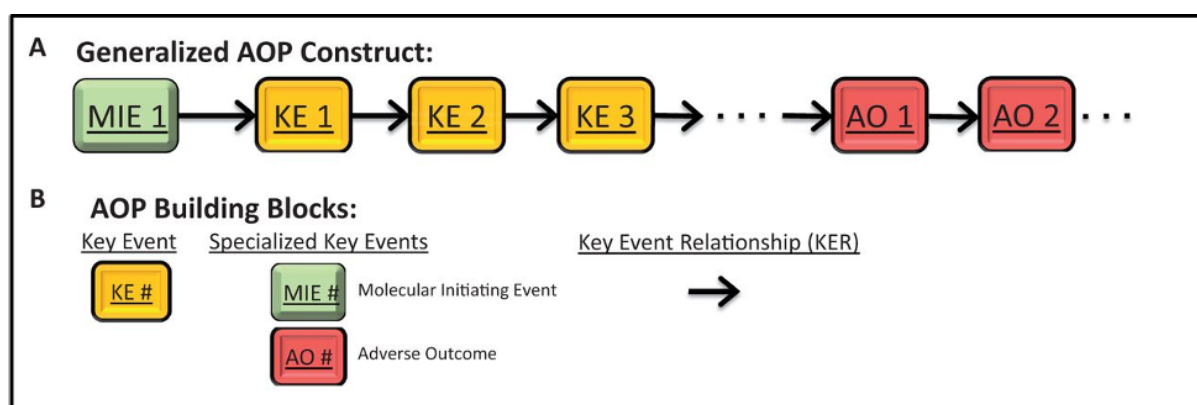
Figure 1.15. Schematic showing a generalisation of an AOP network that begins an interaction of a stressor with the genetic material or machinery, known as the molecular initiating event (MIE). A series of cellular responses follow that are known as Key Events (KEs) in sequence of biologically plausible order. KEs are connected by Key Event Relationships (KERs) that describe the relationship between the KEs. The KEs on the cellular level promote the eventual adverse outcome (AO), which in the case of genotoxicity could be aneuploidy, or mutagenicity and chromosome breaks depending on the mode of action of the exposed chemical. Figure adapted from Villeneuve et al. (2014).

Strong proponents of the AOP approach imagine a situation where the AOP provides a framework for incorporation of new assays that supports KEs to gain regulatory approval (Yauk et al., 2015). Opponents voice opinions that the AOPs are a mere rearrangement of already understood biological principles (personal communication, Rosalie Elespuru, US FDA). However, without a simple holistic approach such as AOPs that are endorsed by major organisations such as the OECD, it is difficult to imagine regulators changing practice from risk averse hazard characterisation. From a more practical perspective, AOPs could provide

the basis for realising key data and test gaps, and potentially facilitate grouping and read across.

1.5.3 Modern Test Systems

To be able to move away from the hazard assessment assay types already briefly introduced in section 1.2, Dearfield et al. (2017) idealises that more modern test systems that specifically address the underlying biological MOA are required. The assays should be selected based on the rational biological argument and knowledge base that has been collected from QSAR and AOP databases, including data gaps and uncertainties. Only then shall a biological assay be selected. For example, a bacterial mutagenicity test would be inapt to assess the humanly relevant risk of exposure to a compound which has a structural alert for a tubulin binding property.

1.6 Thesis Aims

This thesis investigates the application of the BMD approach to novel *in vitro* genotoxicity test systems with the aim of realising some of the main themes postulated by Dearfield et al. (2017) in the 'next generation testing strategy for assessment of genomic damage'. Prior to the initiation of the research presented in this thesis, BMD modelling had only been applied to datasets from 'gold standard' genotoxicity endpoints such as *in vitro* micronuclei, where limited mechanistic information can be derived. Contrastingly, high quality datasets with large numbers of tested concentrations of compounds accompanied the advent of high throughput multiplexed genotoxicity assays such as MultiFlow and ToxTracker. One of the key themes of the next generation genotoxicity assessment strategy is to encourage a paradigm shift from hazard assessment to a risk assessment approach that encompasses quantitative dose-response understanding of genotoxic potential coupled with mode of action information.

The two key questions asked in this research are "can BMD modelling be applied to multiplexed genotoxicity biomarker quantitative data? And, if so, what additional mode of action information can be derived from BMD modelling of the quantitative dose-response biomarker data?". In testing the overarching hypotheses, two key concepts developed by previously cited authors are emulated in this thesis: namely, 1) the use of the BMD combined covariate approach to rank potencies and perform comparative correlations between test systems developed by scientists at RIVM (Hernandez et al., 2010; 2011a; 2011b, Soeteman-

Hernández et al., 2015; 2016); and 2) the use of the BMD combined covariate approach and compound structural information for potency determinations previously explored by Wills et al. (2016a).

This thesis is subdivided into 4 results chapters (chapters 3 to 6) that is an all-encompassing report of the progressive hypotheses that were developed over the course of the research period. The layout of the thesis is summarised in **Figure 1.16**.

Chapter 3

Is MultiFlow data suitable for BMD modelling? And how does it compare with results from a gold standard assay?

Data Source:

Litron provided MultiFlow (γH2AX, pH3, p53) and MicroFlow *in vitro* micronucleus raw data

Methods Used:

Database curation, BMD modelling, BMD correlations

Chapter 4

Is the use of BMD point estimates statistically robust for comparative potency analyses of *in vitro* genotoxicity biomarker datasets?

Data Source:

Litron provided MultiFlow (γH2AX, pH3, p53, RNC) raw data for 10 clastogens in the presence and absence of S9 exposure

Methods Used:

Database curation, Pubchem searches, BMD modelling, BMD correlations, potency rank validation, literature search

Chapter 5

Can *in vitro* genotoxicity biomarker BMD potency ranks be derived for the Topoisomerase II Poison compounds?

Data Source:

Litron provided MultiFlow (γH2AX, pH3, p53) raw data; Toxys provided ToxTracker raw data; literature derived fluorescence anisotropy assay summary data

Methods Used:

Database curation, BMD modelling, BMD correlations, potency rank validation, literature search, Derek Nexus structural alert interpretation

Chapter 6

Do *in vitro* genotoxicity biomarkers map to the Topoisomerase II Poison Adverse Outcome Pathway (AOP)? And can a QSAR Model be built using the same AOP?

Data Source:

Raw data structural alerts from Lhasa Ltd, molecular descriptor raw data from the ChEMBL database

Methods Used:

HESI GTTC AOP application, literature search, structural alert to KE network mapping, K-Nearest Neighbour Analysis

Figure 1.16. Layout of the results chapters in this thesis

Chapter 2. General Materials and Methods.

2.1 General Materials and Methods

This chapter provides an account of the general materials and methods and their theory that are central to this thesis.

2.2 Genotoxicity Assay Data Sources

All raw data was kindly provided by collaborating scientists throughout the research period. Specifically, Litron Laboratories provided data from the MicroFlow and MultiFlow assay flow cytometric datasets, Toxys provided data from the ToxTracker assay flow cytometric datasets. The following subsections describe the acquisition of the MicroFlow, MultiFlow, and ToxTracker assay raw data and the transformation of the raw data into numerical values suitable for quantitative dose-response analysis.

2.2.1 MicroFlow Datasets

MicroFlow data was utilised to test the hypothesis presented in results chapter 3. As briefly describe in the introduction section, the MicroFlow assay is a flow cytometric assay that involves the incubation of mammalian cell cultures with test compounds of interest. Next, necrotic, and mid/late-stage apoptotic cells with ethidium monoazide fluorescent dye, followed by washing and cell membrane lysis and incubation with RNase and CYTOX green nucleic acid green to provide a suspension of free nuclei and micronuclei that are differentially stained (EMA-/CYTOX+) relative to chromatin associated with dead/dying cells (EMA+/CYTOX+). Additionally, latex particles are included to serve as counting beads, whereby flow cytometric nuclei to bead ratio measurements are an indirect measure of cytotoxicity/cell survival (Avlasevich et al., 2006; 2011). To generate *in vitro* micronuclei data for testing under the hypotheses presented in this research project, all test compounds were exposed to human lymphoblastoid TK6 cells for 24 hours and treated per the standard MicroFlow protocol described by Avlasevich et al. (2006). Importantly, the incidence of micronuclei was determined through the acquisition of at least 20,000 gated nuclei per culture.

Important aspects of the MicroFlow assay study design relative to quantitative dose response data analysis includes the acquisition of solvent control micronuclei frequencies,

and at least 4 different concentrations of test compound exposure – as necessitated for dose-response analysis by Slob and Setzer (2014). Raw data was provided by Litron Laboratories for a set of 34 reference genotoxicants (subsets of aneugens and clastogens) which included solvent controls for DMSO, and between 8 to 20 evenly spaced concentrations of test compounds. Relevant flow cytometric data was provided in Excel format which included the compound name, the concentration (μM), percent micronuclei, and percent relative nuclei count to serve as a measure of cytotoxicity.

2.2.2 MultiFlow Datasets

MultiFlow datasets were utilised to test the hypotheses presented in all results chapters (chapters 3,4,5, and 6). Applicable MultiFlow biomarkers included γH2AX for the quantification of DNA double strand breaks, p-H3 for the quantification of mitotic cells, and p53 as a pan-genotoxic and nuclei content biomarker. Like the MicroFlow assay datasets, the MultiFlow datasets included percent RNC derived from the addition of latex counting beads to inform treatment-related cytotoxic effects. To generate MultiFlow endpoint data for testing under the hypotheses presented in this research project, TK6 cells were exposed to all test compounds and treated per the standard MultiFlow protocol described by (Bryce et al., 2016; 2017). Flow cytometric analyses were performed at the 4- and 24-hour sampling times, except where otherwise indicated in the specific results chapters.

Raw data was provided by Litron Laboratories for the diverse sets of reference genotoxicants (subsets of aneugens clastogens, and Topoisomerase II Poisons indicated in the specific results chapters of this thesis) which included solvent controls, and multiple evenly spaced concentrations of test compounds. Relevant flow cytometric data was provided in Excel format which included the compound name, the concentration (μM), γH2AX and p53 median fluorescence intensity, p-H3 for both 4- and 24-hr sampling points, and RNC frequency for the 24-hr sampling point. As described in (Bernacki et al., 2016, Bryce et al., 2016; 2017) the γH2AX and p53 endpoint measurements are based on median fluorescence intensity; the p-H3 endpoint measurements are the proportion of p-H3 positive events that exhibit propidium iodide-associated fluorescence of 4n and greater DNA content relative to the number of total events with 2n and greater DNA content; and, RNC is based on the nuclei to counting bead ratios. The provision of median fluorescence intensity values (for the γH2AX and p53 endpoints), and cellular event frequency values (for the p-H3 and RNC

endpoints) allows the variation within the controls to be expressed and considered within dose-response analyses.

2.2.3 ToxTracker Datasets

The ToxTracker biomarkers loosely relate to the genotoxic modes of action that they describe – namely Bscl2-GFP and Rtkn-GFP for DNA damage, Btg2-GFP for p53 activation, Srxn1-GFP and Blvrb-GFP for oxidative stress, and Ddit3-GFP for protein damage. Only the DNA damage (Bscl2-GFP and Rtkn-GFP) and p53 responsive (Btg2-GFP) biomarkers were considered applicable to the hypotheses being tested in this thesis. Hence, Srxn1-GFP, Blvrb-GFP, and Ddit3-GFP biomarker responses were not evaluated in the results chapter of this thesis. ToxTracker dose-response datasets were used and reported in chapter 5 of this thesis.

Raw data was provided by Toxys for the Topoisomerase II Poisons indicated in chapter 5 of this thesis. Data included solvent controls, and 5 evenly spaced concentrations of test compounds. Briefly, mouse embryonic stem (mES) cells that were GFP tagged for the assays specific biomarkers were exposed to a test set of 6 Topo-II Poison compounds. Following a 24-hour period of incubation, the induction of GFP was measured by flow cytometry. The reporter activity was determined as the mean fluorescent intensity from 5,000 intact cells. The median fluorescence values were formatted into and Excel spreadsheet together with the corresponding compound and concentration in μg/mL.

2.3 Benchmark Dose (BMD) Modelling

The BMD approach includes fitting a family of mathematical models that best describe the dose-response. The mathematical models are central to the BMD approach applied to this thesis. The mathematical theory is described herein.

It has been postulated that no single dose-response model can fully describe a given dose-response according to the underlying biological mechanisms, and hence descriptive models are more appropriate (Slob, 2002). Since not all models are suitable for risk assessment purposes, Slob (2002) considers the following important points relevant in the development of the family of dose-response models to be applied to toxicological data: a) toxicological continuous measurements are non-negative, and hence polynomial functions $(y = a + bx + cx^2 + \cdots)$ shall be excluded; b) fundamentally, threshold models are hard to defend and shall not be considered plausible; 3) relative sensitivity between populations is

important and shall be expressed as one of the models parameters. This is relevant to toxicological studies since observations can be different between study groups, e.g., inter-species, inter-sex, inter-human, inter-cell line etc. Slob (2002) expands on this point by explaining that responses between groups can only be defined as equally sensitive when the background response is equal for all doses. This is often achievable when two populations have responses that do not differ in background for a given endpoint. However, take the example of body weight differences between males and females that usually differ. An equal response in both sexes is defined by Slob (2002) as:

$$\frac{maleBW(d)}{maleBW(0)} = \frac{femaleBW(d)}{femaleBW(0)} \text{, for every dose d,}$$

where Slob (2002) uses the same reasoning of population sensitivity and expresses it for the dose-response model in the form:

$$y = af(x) \text{ (Equation 1)}$$

Such equation (1) describes that two different populations which may differ in background value *a* (body weights in this example), can be characterized by the same dose-response function *f(x)*, whereby they are considered equally sensitive. Differences in sensitivity can be expressed in terms of a dose (potency) factor, and can be estimated by obeying:

$$y = af\left(\frac{x}{b}\right) \text{ (Equation 2)}$$

Slob (2002) states that the function *f(x)* in equation 2 should obey *f(0)* = 1, which exponential functions are capable of achieving. Thus, Slob (2002) proposes the following set of flexible exponential models which can be tested for goodness of fit (using Log likelihood ratios), and that level off at high doses whilst not changing much at low doses:

Model 1: $y = a$ with $a > 0$

Model 2: $y = a \exp\left(\frac{x}{b}\right)$ with $a > 0$

Model 3: $y = a \exp\left(\pm \left(\frac{x}{b}\right)^d\right)$ with $a > 0, \ b > 0, \ d \geq 1$

Model 4: $y = a \left[c - (c - 1) \exp\left(-\frac{x}{b}\right) \right] with \ a > 0, \ b > 0$

Model 5: $y = a \left[c - (c - 1) \exp \left(-\frac{x}{b} \right)^d \right]$ with $a > 0, b > 0, c > 0, d \geq 1$,

where the continuous endpoint is denoted by y, and the dose by x. The parameters a, b, c, and d describe the shape of the dose-response, with the 5 models including additional parameters as the model becomes more complex. Parameter a corresponds to the level of the endpoint at dose 0, synonymous to the response of the background. As previously introduced, parameter b reflects the sensitivity of the subject, which can also be interpreted as the potency/efficacy of the test chemical, and $b^{\frac{1}{d}}$ is directly proportional to potency in the exponential model (Slob and Setzer, 2014). Models 4 and 5 include the additional parameter c, with the models levelling off at high doses (value ac), and hence, parameter c represents the maximum fold change. Parameter d is used to constrain values equal to or larger than 1, thereby preventing the slop of the function at dose 0 being infinite and reflects the impact of dose on the rate of change in the response (Slob and Setzer, 2014).

The appropriate model is selected by performing the log-likelihood ratio test (Slob, 2002) which tests whether the additional parameters result in a statistically significant improvement of fit at $P = 0.05$ (Gollapudi et al., 2013b). Slob (2002) depicted the shapes of the different dose-response models and is shown in **Figure 2.1**.

$$y = a$$

$$b = \infty$$

$$y = a \exp(x/b)$$

$$c = 0$$

$$y = a\,[c - (c - 1)\exp(x/b)]$$

$$d = 1$$

$$d = 1$$

$$y = a \exp((x/b)^d)$$

$$c = 0$$

$$y = a\,[c - (c - 1)\exp((x/b)^d)]$$
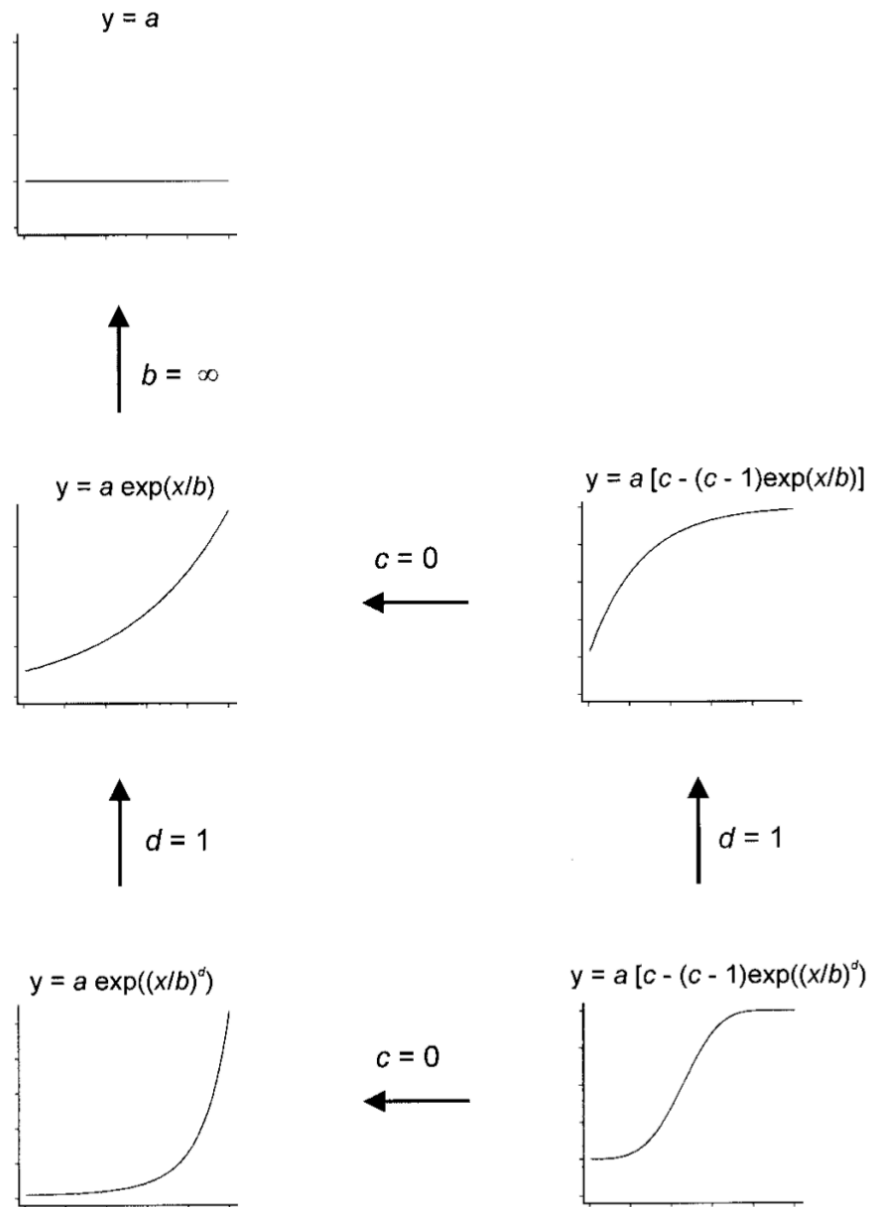
Figure 2.1. Illustration of the shapes of the dose responses corresponding to the various descriptive models. Parameters: a, background response; b, potency; c, maximum effect; d, rate of change (effect on slope/curvature). Figure adapted from Slob (2002).

EFSA (2009) describes that selection of the optimum model from the nested set of models best describes the dose response. EFSA (2009) explains a linear dose response fits a 2-paramter model, whereas a sigmoidal curve fits a minimum 3-parameter model.

EFSA (2009) introduced that Hill models are also suitable for describing the dose-response in addition to exponential models. Hill models built on the same 4 parameters (a, b, c, and d) and have the following formulae:

Hill model 1: $y = a$

Hill model 2: $y = a \left[1 - \frac{x}{(b+x)}\right]$

Hill model 3: $y = a \left[1 - \frac{x^d}{(b^d + x^d)}\right]$

Hill model 4: $y = a \left[1 + \frac{(c-1)x}{b+x}\right]$

Hill model 5: $y = a \left[1 + \frac{(c-1)x^d}{b^d + x^d}\right]$

Slob and Setzer (2014) analysed the goodness of fit of both the exponential and hill four-parameter models to continuous data resulting from the following study types: subchronic toxicity study endpoints (body weight, relative liver weight, and relative kidney weight); acetylcholinesterase activity; *in vivo* and *in vitro* micronucleus assays; local lymph node assay; and whole culture rat embryo development effects test. They demonstrated similar fits from fitting either exponential or hill models to the dose response datasets. And even if some differences occur in goodness of fit, the shapes of the dose response curves were visually difficult to distinguish (Slob and Setzer, 2014). Despite (EFSA, 2017) continuing to recommend fitting both exponential and hill models to dose response datasets, based on the findings of Slob and Setzer (2014), only the exponential models were used in the BMD modelling presented in this thesis.

2.4 Working in PROAST

The BMD modelling for the research presented this thesis was performed using PROAST which is an open-source statistical software package developed by the Dutch National Institute for Public Health and the Environment (RIVM) written for operation in the R statistical programme (Slob, 2018). A recent web-based version is available, however lacks much of the more complex functions needed for testing the hypotheses presented in this thesis. Due to this, versions of PROAST were operated in R.

The versions 61.2 and 69.1 of PROAST were used over the course of the research presented in this thesis. Specific version instances are reported in the methods section of each results chapter. The different software versions are progressive iterations where software

functionalities were improved, however still function by using the same mathematical functions and under the same principles.

Data needed to be formatted for import into PROAST. The database curation and specifics related to the data formatting are specified in each results chapter. However, in general, dataset formatting must follow strict PROAST formatting requirements where special characters and spaces are not permitted. Heading rows need to be formatted to specify the response data types in each column (continuous, binary, ordinal, quantal), or if the data is not a response variable (compound name, test condition etc). Since these data formats are highly specific to the data type under evaluation, the specific formats are presented in each results chapter under the chapter materials and methods sections.

PROAST works by navigating a series of questions mainly including variables selection, choice of CES, choice of BMD model (exponential and/or hill), and various output options. PROAST was used for BMD modelling including the calculation of BMD confidence intervals (chapter 3, 4, and 5) and for correlating BMD confidence intervals between systems (chapter 3, and 5) depending on the hypothesis being tested. The BMD combined covariate approach was the preferred method used in each results chapter where BMD analysis was performed (chapter 3, 4, and 5). The following flow chart (**Figure 2.2**) shows the general process flow using PROAST for BMD analysis:

1. Dose response data fitting	A four-parameter exponential model was fitted to the combined datasets. Compound was selected as the combined covariate. The resulting exponential BMD model was used for further analysis.
2. Check dose-response fitting	The model's fitting was checked visually by plotting the data and fits on double Log scales for each individual subgroup (compound). Poorly fit models could be improved by removing top concentrations where relatively high cytotoxicity was observed.
3. Confidence interval determination	Confidence intervals (CIs) were then calculated and saved for each sub-group in 'bigtables' (PROAST notation)
5. Combine data from both systems	Matching CEDs and CIs for both systems were combined in a 'crosstable' (PROAST notation). This permitted the collation of BMD metrics from different test systems for correlation.
6. Plot data	'Crosstable' data was then plotted to permit system comparison. Additional information was then overlaid such as dashed horizontal lines encompassing the confidence intervals to study the order of magnitude of which they compare.

4. Repeat steps 1 to 3 per system

Figure 2.2. Flow chart showing the general BMD data analysis process flow using PROAST. Chapter 3 and 5 used options 5 and 6 to generate cross-correlations of BMD confidence intervals between systems.

2.4.1 PROAST Variables Selection

The following general variables were selected using the PROAST navigation questions. Any specific differences are indicated in the results chapters:

- Dependent variable – always 'concentration (μM)'
- Response for analysis by set of models – biomarker response (either 4-hour or 24-hour γH2AX, 4-hour or 24-hour p-H3, 4-hour or 24-hour p53, Rtkn-GFP, Bscl2-GFP, Btg2-GFP).
- Model selection – always exponential

Further settings were changed when there were large numbers of compounds to be tested. The number of compounds for combined covariate analysis were restricted to small groups to reduce the processing time required for analysis.

2.4.2 BMD Model Selection

One of the major differences between PROAST v61.2 and v69.1 is the choice of BMD model selection options. PROAST v61.2 provides the option to select the model with the appropriate covariate dependent parameters, whereas PROAST v69.1 permits model 3 or model 5 selection options. To this end, the data presented in chapters 3 and 5 utilised PROAST v61.2 where the maximum four-parameter exponential model (model 5) was always selected. The data presented in chapter 4 utilised PROAST v69.1 where the BMD model varied between exponential model 3 and 5. The exponential model 3 and 5 formulae are as follows:

$$\text{Model 3: } y = a \exp\left(\pm\left(\frac{x}{b}\right)^{d}\right) \text{ with } a > 0,\ b > 0,\ d \geq 1$$

$$\text{Exponential Model 5: } y = a\left[c - (c-1)\exp\left(-\frac{x}{b}\right)^{d}\right] with\ a > 0, b > 0, c > 0, d \geq 1,$$

where y denotes the continuous endpoint, and the dose by x. Parameter a = the level of the endpoint at dose 0; b = potency/efficacy of the test compound on the shape of the response; c = maximum fold-change in response; and d = rate of change in the response constrained to values equal to or larger than 1.

2.5 Machine Learning Algorithms

Advanced statistical analysis of some data presented in this thesis was performed using machine learning (ML) algorithm techniques. The ML methods include unsupervised (hierarchical clustering) or supervised (k-nearest neighbours) techniques described in the following 2 subsections.

2.5.1 Unsupervised Hierarchical Clustering

Unsupervised hierarchical clustering (also called hierarchical cluster analysis) is a method of cluster analysis which looks for patterns in datasets and arranges them based on similarities or differences. The methods are unsupervised in the sense that the data does not require domain knowledge or labels provided by experts (Hirano et al., 2004), and algorithms work intuitively to determine the best groups to describe the trends in the datasets. Hierarchical clustering has been widely used in the biological sciences to determine groupings in complex datasets such a gene expression patterns (Herrero et al., 2001) and cancer cell histopathological markers (Au et al., 2004). The data outputs can be regarded as unbiased interpretations of the underlying datasets that does not rely solely on human interpretation.

In quantitative datasets, clustering is usually based on analysis of the distances between data points (Ward and Hook, 1963). Hirano et al. (2004) describes that the main type of algorithm for hierarchical clustering is agglomerative hierarchical clustering in which vectors initially receive an independent cluster. The algorithm proceeds to seek the most similar pair of clusters and merges them into one cluster. The process is repeated until all initial clusters are merged into a single cluster. Agglomerative hierarchical clustering contains several options for determining how to merge clusters, including the simplest single linkage which groups similarity based on the closest pair, complete linkage which groups based on the furthest pair, or Ward's method which is based on the sum of squares within a cluster (Hirano et al., 2004).

Unsupervised hierarchical cluster analysis was used for data analysis presented in chapter 4 of this thesis. Analysis was performed using JMP software's unsupervised clustering platform (JMP, v12.0.1) using distances based on Ward's method. Ward's minimum variance method is a method of agglomerative hierarchical clustering that expresses the distance between two clusters as the ANOVA sum of squares between the two

clusters summed over all the variables (Ward and Hook, 1963). Firstly, a Euclidean space is determined based on the squared Euclidean distance between data points (vectors) which can be expressed by the following sum of squares formula from (Spencer, 2013):

$$d^2(p, q) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2.$$

where d represents distance and p, q are data points in the Euclidean space.

Secondly, distances are arranged into clusters based on the distance centroids, and then iteratively arranged based on Ward's method which can be expressed as the following formula joining two clusters K and L from (JMP, 2021):

$$D_{KL} = \frac{\|\bar{x}_K - \bar{x}_L\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

where 1) N_K is the number of observations in C_K and C_K is the Kth cluster, subset of {1, 2, …, n) of which n is the number of observations; 2) $\bar{x}$ is the sample mean vector and $\bar{x}_K$ is the mean vector for cluster C_K; and 3) $\|x\|$ is the square root of the sum of the squares of the element of x (the Euclidean length of vector x). The result is a 2-dimensional dendrogram where the distance between clusters is based on the error sum of squares. A benefit of using Ward's method is that the algorithm takes into consideration the variation in the dataset (Murtagh and Legendre, 2014), which is advantageous in clustering since regions of high density and regions of high dissimilarity can be accurately represented in the clustering dendrogram. The reason for the selection of Ward's method in the JMP clustering method was due to its application as a preferred clustering algorithm.

The data for unsupervised hierarchical clustering included BMD metrics (BMDL, BMD, and BMDU) for the MultiFlow biomarker responses from clastogens exposed to TK6 cells for up to 24-hours in the presence and absence of low dose rat liver S9, as presented in chapter 4.

2.5.2 K-Nearest Neighbour Analysis

K-Nearest Neighbour analysis is a form of supervised ML that trains itself to recognise patterns in datasets (Kramer, 2013). (KNN) analysis was performed on data presented in chapter 6 of this thesis. Regression KNN analysis is a non-parametric

classification method developed by Fix (1985) and later expanded by Altman (1992). Put simply, a scatter plot is the simplest format to organise data and introduce bivariate relationships as mentioned by Altman (1992). In practical terms using statistical modelling software, the scatter plot is referred to as a metric space where the distances between the members are defined (Hackeling, 2017).

SPSS version 26 was used for the KNN analysis reported in this thesis. The distances in the metric space were calculated by SPSS based on the Euclidean distance which refers to a straight line between points. Gonçalves et al. (2014) describes that the Euclidean distance is the square root of the sum of the squared differences between the origin (i) and the destination (j) values for all variables, and can be represented by the following formula:

$$d(i,j) = \sqrt{\sum_{v=1}^{p} (X_{iv} - X_{jv})^2}$$

where X_{iv} represents the value of vector i; X_{jv} represents the value of vector j; p is the number of portions in the sample; and v is the number of individuals in the sample.

The Euclidean distances are used in the KNN classification algorithm. Distances are compared based on the pre-determined variable 'k'. The hyperparameter 'k' represents how many neighbours can be used in the estimation. The k variable was determined as 3 by rounding up $\sqrt{n}$ where maximum n was 8 (maximum number of target compounds studied in a single metric space). KNN neighbours were visually compared using the SPSS graphical outputs, or by using the Euclidean distance metrics computed by SPSS.

The data for KNN analysis included BMD metrics (BMDL, BMD and BMDU), and molecular descriptor variables (molecular weight, LogP, and polar surface area) for 9 Topoisomerase-II Poison compounds. More detail on the compounds and molecular descriptors is presented in chapter 6.

Chapter 3. Benchmark Dose Modelling of *In Vitro* MultiFlow Endpoints for Compound Potency Ranking and Comparison with the Current *In Vitro* Micronucleus Assay.

3.1 Introduction

The following chapter determines the amenability of MultiFlow biomarker dose-response information to BMD analysis, and to ultimately provide genotoxic potency ranking of tested clastogens and aneugens compared with the gold-standard micronucleus assay[1]. Despite expert consensus that the benchmark dose (BMD) approach is the most appropriate method to perform for dose-response analysis of genotoxicity datasets (Gollapudi et al., 2013b, Macgregor et al., 2015), previous work had only evaluated the utility of BMD modelling to standard genotoxicity endpoints, with no published work existing in which the BMD methodology was applied to novel genotoxicity biomarker response screening tests. Additionally, the *in vitro* MultiFlow assay had been developed as a follow-up to *in vitro* micronucleus positive events, which maintained the paradigm of dichotomising genotoxicity into pass/fail criteria by evaluating the fold-increase in biomarker responses relative to control.

3.1.1. The *In Vitro* MultiFlow Assay's Dichotomisation of Genotoxicity

The *in vitro* MultiFlow assay includes several biomarkers that efficiently quantifies DNA damage by the flow cytometric analysis of phosphorylation of H2AX (γH2AX) for DNA double strand breaks, phosphorylation of histone H3 (p-H3) for mitotic cells, and nuclear p53 content for global DNA damage and treatment-related cytotoxicity. The following graphic summarises the processing steps involved in the acquisition of DNA damage response data using the MultiFlow assay (**Figure 3.1**).

---

[1] The work presented in this chapter was published in "Predictions of Genotoxic Potential, Mode of Action, Molecular Targets, and Potency via a Tiered MultiFlow® Assay Data Analysis Strategy" (Dertinger et al., 2019), with a copy of the research article available in Appendix 1.

MultiFlow™ Processing
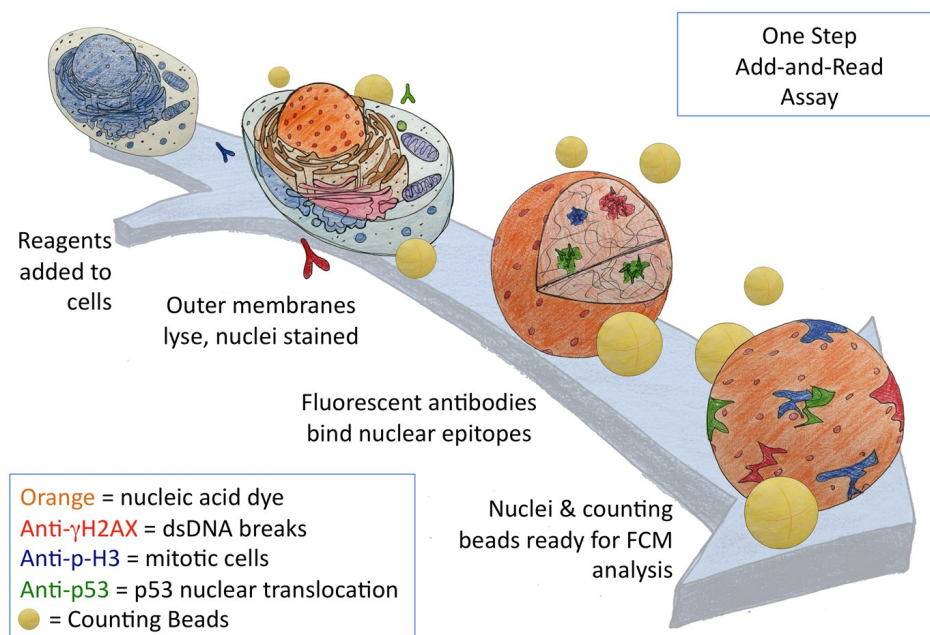
Figure 3.1. Graphic representing the MultiFlow processing steps involved in data acquisition from the assay's DNA damage response biomarkers. Relevant mammalian cells are prepared by cell membrane lysis to expose nuclear content. Fluorescent antibodies specific to the DNA damage response markers (γH2AX foci, nuclear p53, p-H3) bind to their respective nuclear epitopes. The cells and their suspensions are prepared for flow cytometric (FCM) analysis. Latex microspheres act as counting beads to monitor cell-to bead ratios for the evaluation of relative nuclei counts or relative population doublings. Figure adapted with permission from Litron Ltd. Copyright 2016.

The assay performs by measuring the biomarker responses at the 4- and 24-hour timepoints in multi-well plates via flow cytometry. The biomarker responses are prepared for statistical analysis by converting the reported median fluorescence (for γH2AX and p53) or frequency (p-H3) values into fold-change values relative to plate specific solvent controls (Bernacki et al., 2016; Bryce et al., 2016). The assay developers introduced a novel data analysis strategy that involves supervised machine learning, specifically multinomial logistic regression analysis for the categorisation of chemicals into one of three modes of action (MOA) – clastogens, aneugens, or non-genotoxicants. Briefly, algorithms were developed that provide probability scores indicative of a compound's mode of action from intra- and inter-laboratory validation studies using *a priori* mode of action information for between 67

and 103 selected training set test compounds (Bryce et al., 2016; 2017; 2018). The logistic regression output determines whether a test compound is classed as either clastogenic, aneugenic, or non-genotoxic based upon a models' probability score.

In the interlaboratory validation, Litron used an analogy to the Global Evaluation Factor (GEF) criteria employed in the Mouse Lymphoma Assay (MLA) to determine whether a biomarker's response was significant compared to the distribution of the solvent control data from the seven proficient laboratories involved in the validation (Pfizer, Servier Group, Orion Pharma, Sanofi-Aventis Deutschland, Sanofi, Bayer AG, and Roche Pharma). Aggregate data from the replicated studies performed at the seven laboratories showed that assay sensitivity and specificity values were 92.1% and 96.2%, respectively, when using the GEF approach (Bryce et al., 2017). The assay has since been evaluated further by using real-world test compounds provided by Merck & Co. (Dertinger et al., 2019), and by performing unsupervised hierarchical clustering machine learning tools to determine if patterns are present which may inform further information on genotoxic molecular targets (Dertinger et al., 2019). For example, **Figure 3.2** shows dose-response Area Under the Curve (AUC) unsupervised clustering results for a series of select aneugenic compounds. Distinct clusters exist for kinase inhibitors and tubulin binders based on *a priori* compound mechanistic information.
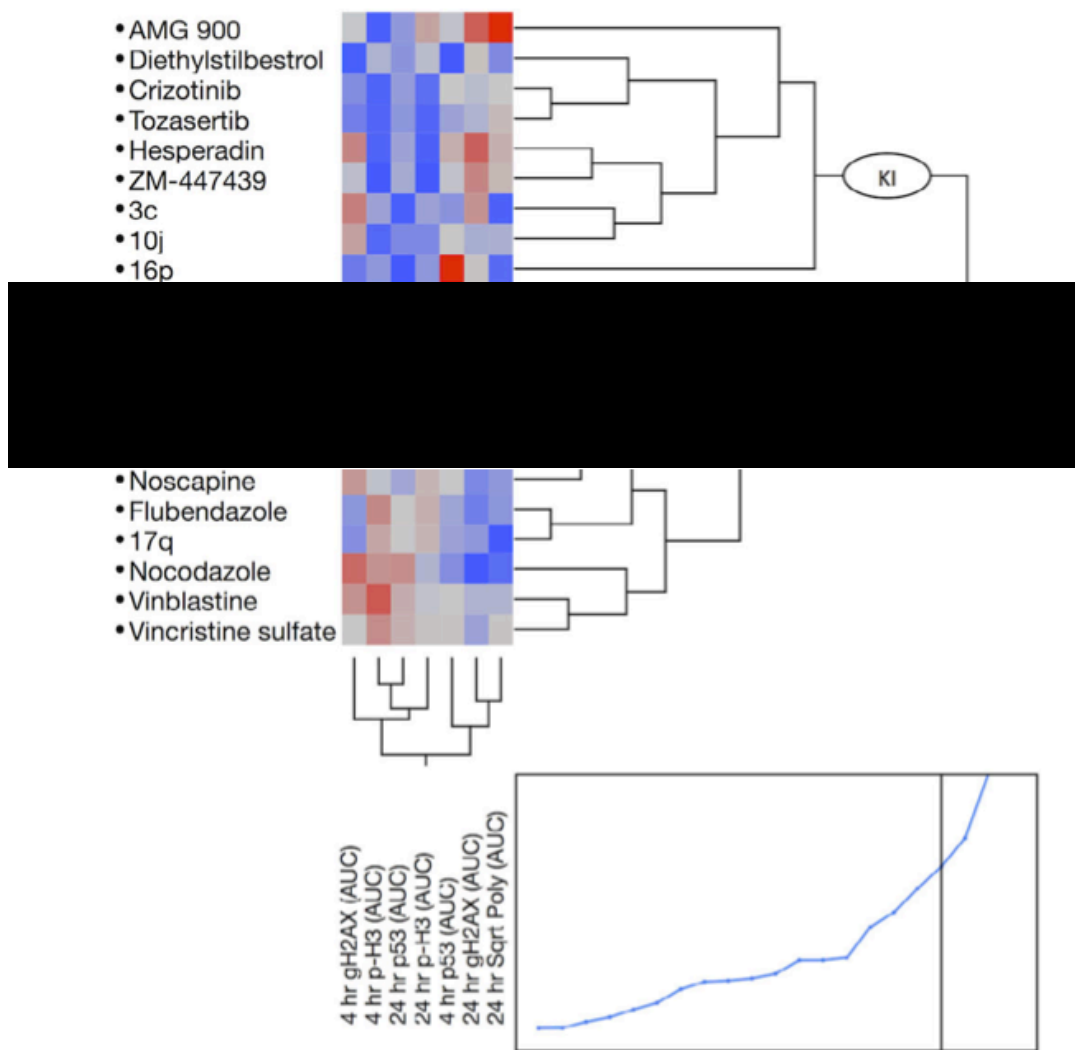
Figure 3.2. JMP software derived unsupervised clustering results shown on a 2-dimensional dendrogram for 21 aneugens based on the dose-response for each biomarker which was converted into Area Under the Curve (AUC) for the analysis. The biomarkers are plotted on the x-axis with the test compounds displayed on the y-axis. Two distinct clades are present and denoted by Kinase Inhibitors (KI) and Tubulin Binders (TB) based on the known activities of these compounds. The bottommost graph represents the horizontal distances between clusters. Figure taken directly from Dertinger et al. (2019).

The novel ability for the MultiFlow assay to discriminate between different genotoxic modes of action offers a clear advantage over the traditional *in vitro* genotoxicity assays which preceded its inception. However, the published literature presents the assays results in a dichotomous call of genotoxic potential which maintains the hazard characterisation paradigm. Whilst the assay developers use the dose-response information in a useful manner to determine concentration-dependent biomarker response fold-increases for genotoxic mode

of action calls, and for unsupervised machine learning through AUC measurements, the assay endpoints had not been reported on the utility of BMD modelling.

3.1.1 The *In Vitro* MultiFlow Assay Endpoints as Candidates for Continuous BMD Modelling

The different data types (quantal, continuous) for BMD modelling have already been summarised in the introduction chapter (section 1.3.4.1) based on the discussion from Slob (2002). The *in vitro* MultiFlow assay produces continuous data. The criteria for datasets to be suitable for continuous BMD modelling includes at least 3 evenly spaced dose/concentration groups plus a solvent control group (Haber et al., 2018); and, either individually reported responses for each repeat of dose/concentration level, or summary statistics including group means and standard deviations (Slob, 2002). Thus, the *in vitro* MultiFlow assay biomarker responses were identified as suitable candidates for continuous BMD modelling since individual median fluorescence intensity values are reported for each concentration level, and multiple concentration levels are tested in addition to a solvent control. The use of median fluorescence values in the concentration and solvent control groups, versus the use of fold-increase values allows the response in the controls to be considered which is a fundamental principle in BMD modelling of continuous endpoints (Slob, 2002).

3.1.2 BMD Modelling to Provide Test Article Potency Ranks

Prominent researchers have utilised the output of the BMD modelling approach to rank test article genotoxicity in order of potency for the test system evaluated. Graphical interpretation of test article potency is achieved by plotting BMD confidence intervals on a Log scale with each respective test article presented on the y-axis. The efficient single combined analysis of dose-responses from multiple test articles is permitted when the 'test article' is selected as a covariate in the BMD analysis. The resulting confidence intervals have been used to rank test article potency yielded from dose-response studied in multiple *in vitro* and *in vivo* test systems. For example: Bemis et al. (2016) presented BMD confidence intervals in rank order for 7 genotoxicants tested in both the *in vitro* and *in vivo* micronucleus assays **Figure 3.3**; and Wills et al. (2016a) displayed the genotoxic potency of different low-dose ionising radiation regimes exposed to cultured human lymphocytes **Figure 3.4**.
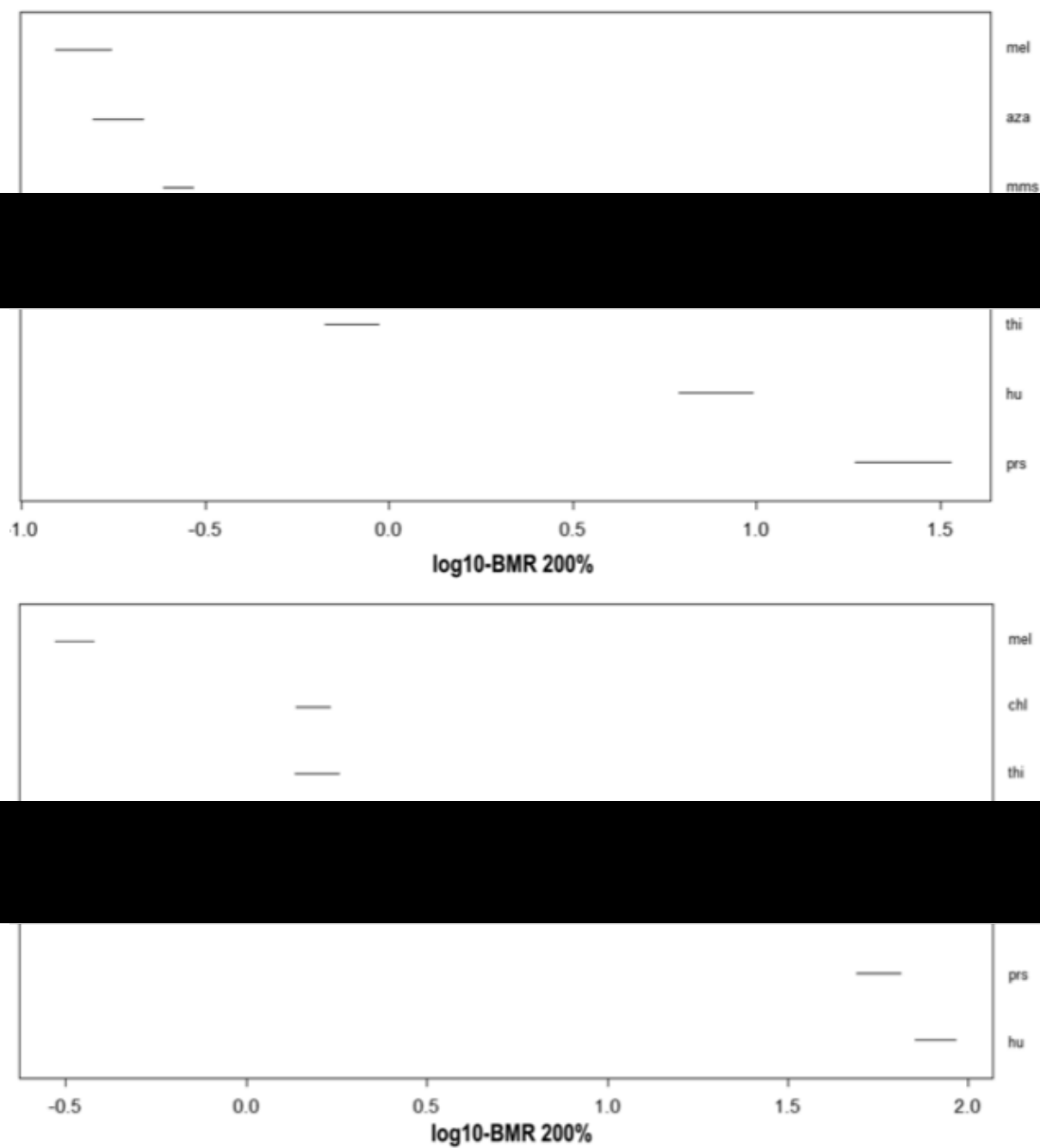
Figure 3.3. Plotted BMD confidence intervals from the *in vitro* (upper panel) and *in vivo* (lower panel) micronucleus assay for 7 genotoxicants. BMD confidence intervals plotted from most potent (mel) to least potent (hu) compound rank order. The x-axis represents Log10 BMD for micronuclei analysed with a BMR (synonymous to critical effect size, CES) of 200%. Abbreviations mel = melphalan, chl = chlorambucil, thi = thiotepa, aza = azathioprine, mms = methyl methanesulfonate, prs = 1,3-propane sultone, hu = hydroxyurea. Figure taken directly from Bemis et al. (2016).
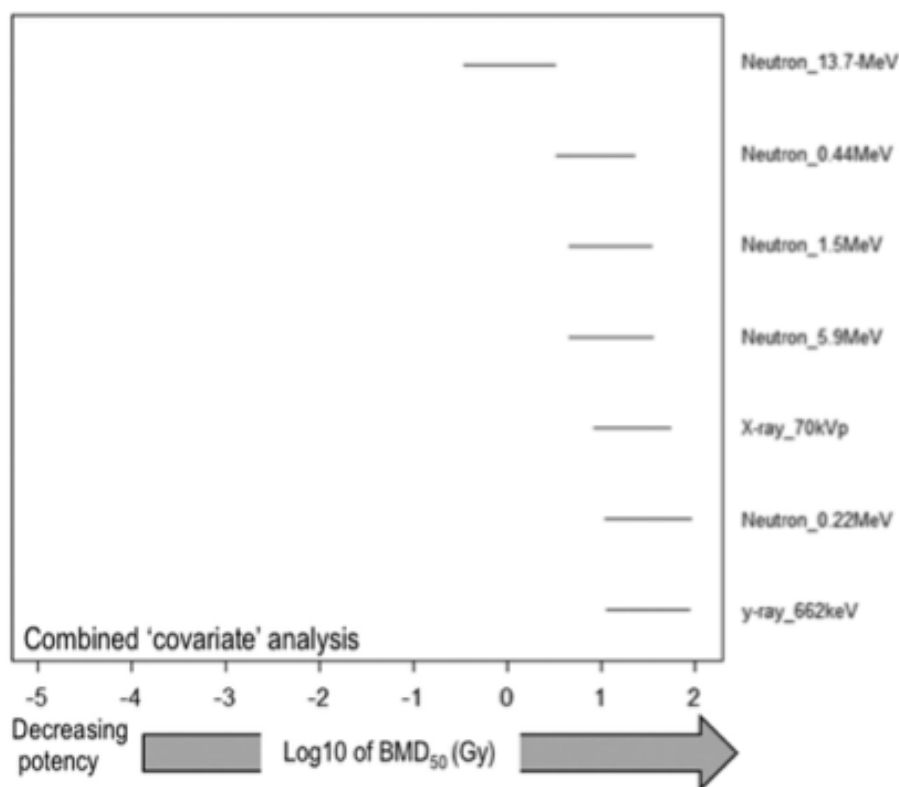
Figure 3.4. Plotted BMD confidence intervals calculated from micronuclei frequency dose-response datasets for cultured human lymphocytes exposed to different forms of ionising radiation conditions. The x-axis represents Log10 BMD for micronuclei analysed with a critical effect size (CES) of 50%. BMD confidence intervals plotted from most potent (Neutron_13.7-MeV) to least potent (y-ray_662keV) ionising radiation condition rank order. Potency is depicted on the x-axis with a shaded arrow in the direction of decreasing potency from left to right. Figure adapted from Wills et al. (2016a).

The outputs from the aforementioned studies can be interpreted as an efficient representation of the rank order potencies for the test articles that were studied. However, the datasets used in the referenced articles were from well characterised test systems, with nothing published on how novel genotoxicity test systems can inform compound potency using the same methodology.

3.1.3 Correlation of BMD Confidence Intervals Across Test Systems

Whilst BMD confidence intervals can be plotted and interpreted to represent the relative potency of a test article, several researchers have postulated that the potency of a compound in one test system may be comparable in a different test system. Thus, a method

for correlating BMD confidence intervals has been widely reported on by the developers of the PROAST BMD modelling software at the Dutch National Institute of Public Health and the Environment (RIVM). The method is simple to execute as demonstrated in the following graphical depiction (**Figure 3.5**).
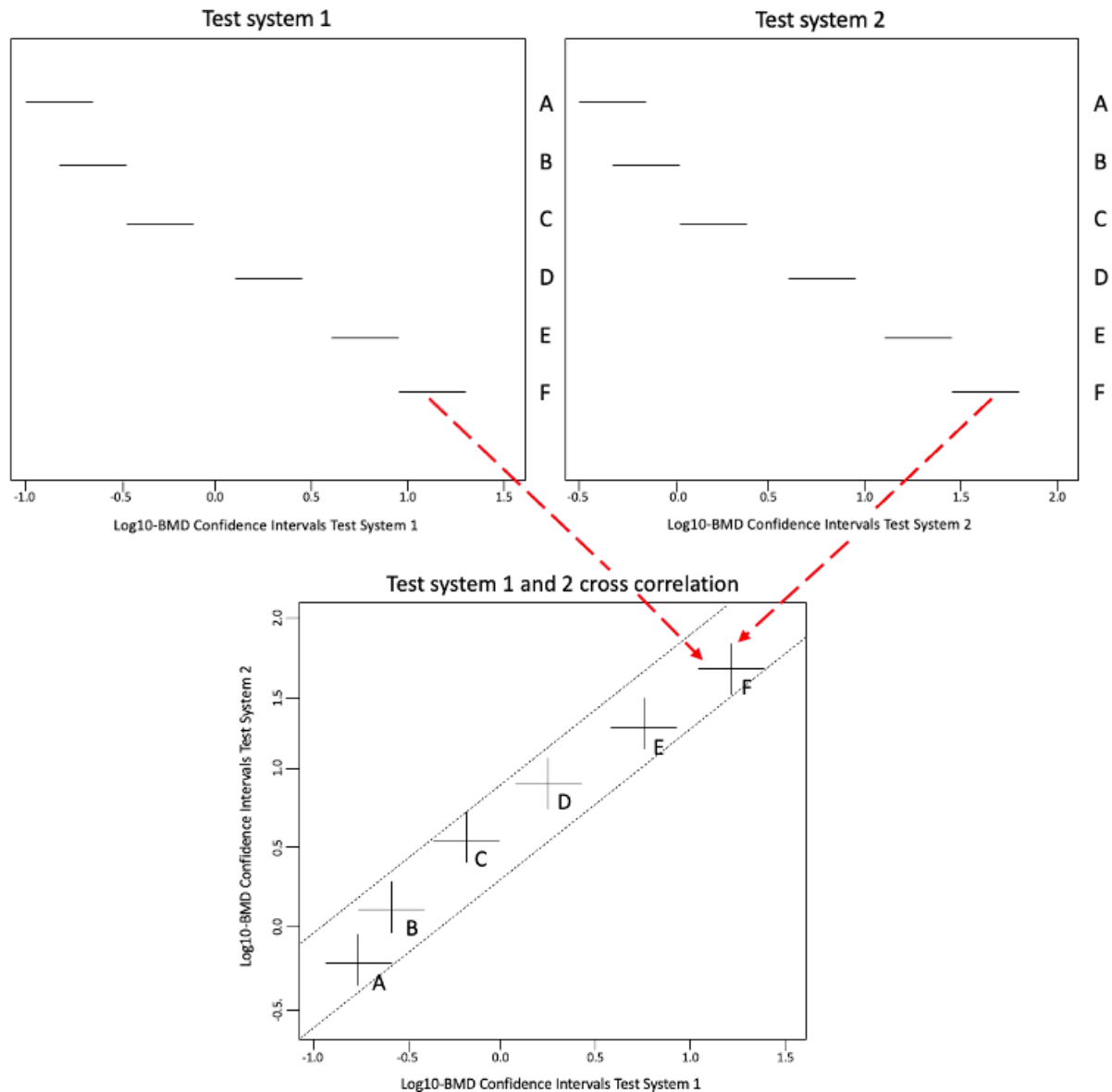
Figure 3.5. The simplicity of the BMD confidence interval correlation method is shown in this figure. BMD confidence intervals for 6 hypothetical test-articles (A to F) are shown plotted in potency rank order for the two test systems (uppermost 2 panels). The software 'PROAST' produces a cross-correlation plot where the BMD confidence intervals are cross-correlated on a double Log plot (bottommost panel). A dashed black line encompasses most BMD confidence intervals to depict the strength of the correlation within approximately 0.5-

Log. Test-article 'F' is highlighted here with 'red arrows' pointing from the respective BMD confidence interval origin from the test systems shown in the uppermost panels.

The BMD confidence interval correlation method has been applied by several researchers to demonstrate the correlation between compound potency evaluated by *in vitro* and *in vivo* micronuclei test systems. Correlation of *in vitro* and *in vivo* micronuclei from various test compounds by Soeteman-Hernández et al. (2015) (**Figure 3.6**) led the researchers to discuss a generalisation that *in vivo* confidence intervals tend to be smaller than those derived from *in vitro* test systems, and that observed scatter in the *in vitro* test system may be due to the fact that the *in vitro* test system does not account for toxicokinetic properties in the same manner that an *in vivo* system does. However, the researchers mention an important point that the *in vitro* datasets are limited by a poor statistical power from the low number of test concentrations and replicates.
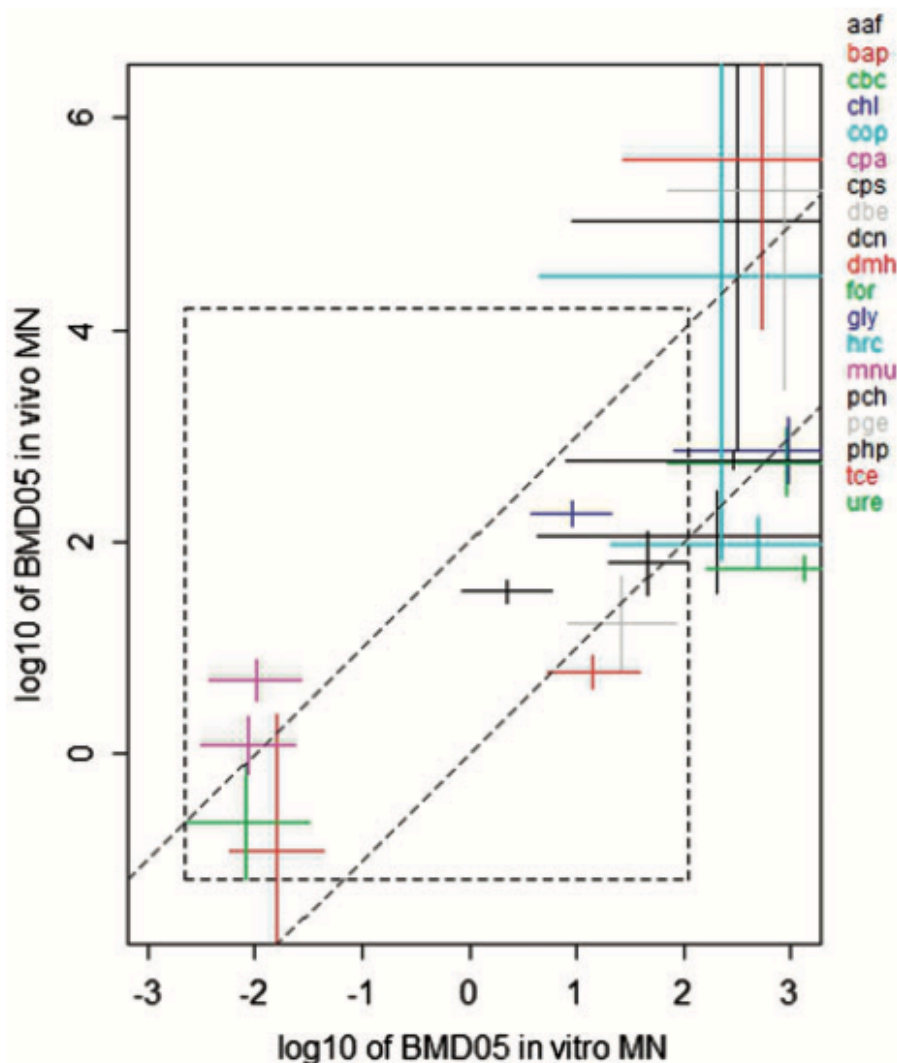
Figure 3.6. BMD confidence interval cross-correlation plot for *in vitro* versus *in vivo* micronucleus frequencies for a series of select test compounds. Figure adapted from Soeteman-Hernández et al. (2015).

Another example includes a study by Bemis et al. (2016) who applied the same BMD confidence interval correlation method for a series of clastogens applied to *in vitro* and *in vivo* micronucleus test systems. In this example, there was considerably less scatter, and the width of the confidence intervals were comparably small between the two test systems (**Figure 3.7**). Notably, the micronuclei frequencies obtained for BMD analysis in the Bemis et al. (2016) study were from flow cytometric analysis in comparison to the manual scoring techniques employed by Soeteman-Hernández et al. (2015), which may account for the higher precision obtained in the Bemis et al. (2016) study since flow cytometric analyses can score large number of cells for a large range of test article concentrations.
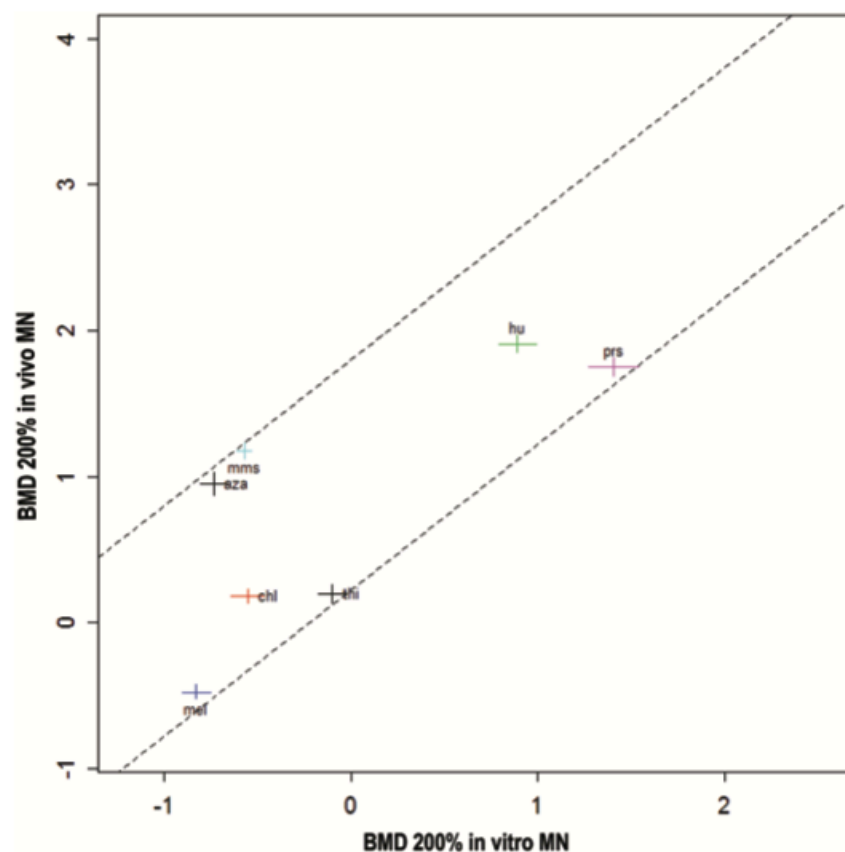
Figure 3.7. BMD confidence interval cross-correlation plot for *in vitro* versus *in vivo* micronucleus frequencies for several tested clastogenic compounds. Small width confidence intervals (high precision) were obtained from BMD analysis from both systems. Figure adapted from Bemis et al. (2016).

This method provides an effective measure of the potency differences observed from compounds studied in two comparative systems and can be applied to test several hypotheses.

3.1.4 Estimation of Test Article Potency from BMD Confidence Interval Correlation

Soeteman-Hernández et al. (2015) used BMD confidence interval cross-correlation plots to estimate the carcinogenic potency of a compound from the *in vitro* micronucleus assay **Figure 3.8**. One may challenge the applicability of the extrapolation, given the acknowledged toxicokinetic differences between *in vitro* and *in vivo* test systems; however, the proof-of-concept approach demonstrates how the BMD confidence interval cross-correlation plots can be used as an efficient graphical predictive tool for cross-system potency comparisons.
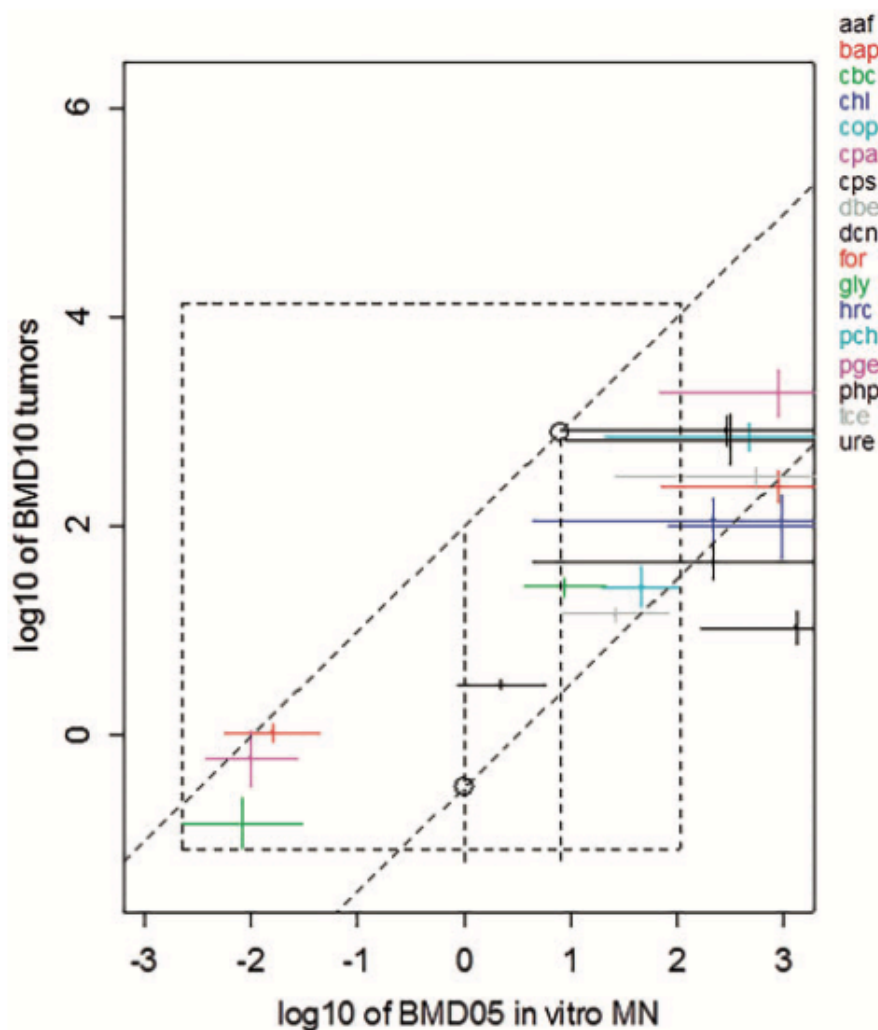
Figure 3.8. BMD confidence interval cross-correlation plot for *in vitro* micronucleus versus *in vivo* tumour incidence for several tested compounds. Dashed diagonal lines are drawn to encompass most of the BMD confidence intervals. Round black circles are drawn around select BMDL intersects which represent read-across values for predictive BMDL values in the comparator test system. Figure adapted from Soeteman-Hernández et al. (2015).

3.1.5 *In Vitro* Micronucleus Scoring *via* Flow Cytometry with the Litron MicroFlow Assay

Litron's MicroFlow assay is one of the flow-based approaches which is the preferred method for characterising micronuclei *in vitro* (Johnson et al., 2014b, Verma et al., 2017). In 2010, Bryce et al. (2010) used the same assay and exposed human lymphoblastoid TK6 cells to 22 closely spaced concentrations of 8 different genotoxicants. After data acquisition, Bryce et al. (2010) demonstrated that the MicroFlow data is amenable to dose-response relationship

modelling. However, the authors used the bi-linear hockey stick model which was discussed in detail in the introduction chapter and in the publication by Lutz and Lutz (2009).

The suitability of BMD modelling applied to the MicroFlow assay data has been demonstrated by other authors. The earlier presented example in **Figure 3.7** where Bemis et al. (2016) correlated *in vitro* and *in vivo* micronuclei included data from the *in vitro* MicroFlow assay. From an industry perspective, Allemang et al. (2018) used the *in vitro* MicroFlow assay to assess micronuclei after exposure of 15 pyrrolizidine alkaloids to *in vitro* cultured HepaRG cells. Amongst some of their findings was that the MicroFlow assay is amendable to BMD analysis, and that the resulting BMD confidence intervals adequately characterise the potency differences observed amongst the 15 pyrrolizidine alkaloids that were tested in the *in vitro* liver cell line that the researchers selected.

Based on these limited publications in which BMD modelling was applied to MicroFlow datasets, and the fact that the MicroFlow assay has received endorsement as a preferred flow cytometric genotoxicity assay (Johnson et al., 2014b, Verma et al., 2017) with an inter-laboratory validation (Bryce et al., 2008) and demonstrated to be comparable to an OECD endorsed method (OECD 487, *in vitro* mammalian cell micronucleus test) in terms of scoring proficiency (Bryce et al., 2008), it is plausible to regard data from the MultiFlow assay as gold-standard for drawing comparisons against. In fact, many of the tests using the same assay in industry for regulatory submission purposes will be performed according to Good Laboratory Practices (US Code of Federal Regulations Title 21 Part 58 or OECD equivalent) which ensures characterisation of the test article, proficient laboratory analysts, calibrated equipment, and quality data reporting.

3.1.6 Are MultiFlow Biomarker Responses Capable of Determining Usable Genotoxic Potency Ranks?

Several questions were raised upon collaboration with Litron Laboratories. Primarily, although useful to industry, the notion that the MultiFlow assay was designed as a follow up to positive *in vitro* micronucleus responses was challenged since this approach maintains the hazard characterisation paradigm that the genetic toxicologists are advocating against. In the aim to achieve a paradigm shift, Litron Laboratories supported the assessment of BMD modelling to their datasets to test the hypothesis of whether the MultiFlow datasets were amendable to BMD analysis. If said hypothesis was proven correct, in the context of

MultiFlow biomarker responses, it would be interesting to know if distinct potency rank orders for tested compounds are obtainable and how the potencies compare to those from the micronucleus endpoint. To this end, this chapter aims to describe the BMD modelling that was performed on the MutliFlow biomarker response datasets, and how potency information compares with that from the micronucleus endpoint to further inform the utility of the MultiFlow assay beyond a hazard assessment approach, and towards a quantitative risk assessment paradigm.

3.2 Materials and Methods
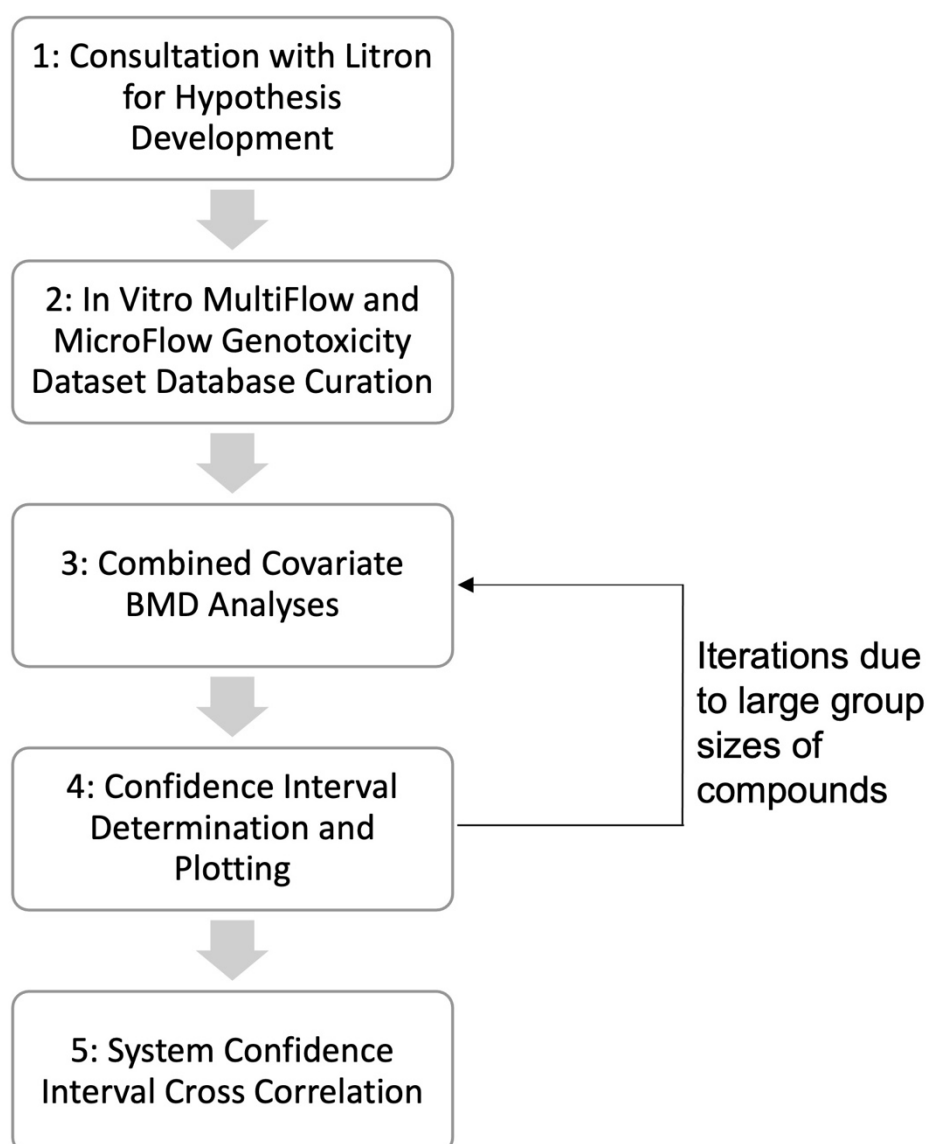
Figure 3.9. Materials and Method Flow Chart

3.2.1 Datasets

Raw data for analysis was provided by the collaborating scientists at Litron Laboratories. Raw data in this chapter is defined as tabulated median fluorescence, or median relative frequency data outputs from flow cytometric analysis performed at Litron Laboratories.

All raw datasets required database curation including formatting, database maintenance and data handling prior to further analyses. Specific information is included in the following subsections.

MultiFlow assay raw data took the format of an Excel spreadsheet with headings including compound name, a priori mode of action information (clastogen, aneugen), concentration (μM), median fluorescence data for each compound and concentration level for the biomarkers γH2AX and p53, median relative frequency values for p-H3, polyploidy, RNC (Relative Nuclei Count), and RPD (Relative Population Doubling). All measurements were performed at the 24-hour sampling timepoint.

MicroFlow assay raw data took the format of an Excel spreadsheet with headings including compound name, a priori mode of action information (clastogen, aneugen), concentration (μM), median relative percent scores for micronuclei, EMA, G1 phase cells, S phase cells, G2M cells, polyploidy cells, and RNC. An additional column of calculated cytotoxicity was provided in the Excel spreadsheet. All measurements were performed at the 24-hour timepoint.

3.2.1.1 MultiFlow Assay Specific Information Pertaining to the Dataset

Litron Laboratories performed the benchtop experiments to acquire the MultiFlow dataset for analysis. The assay has been extensively discussed in the published literature (Bryce et al., 2014; 2016; 2017; 2018). Briefly, human lymphoblastoid TK6 cells were prepared using the reagents and instructions specified in the MultiFlow DNA damage kit – p53, γH2AX, Phosphohistone-H3. Cell suspensions were exposed to a total of 44 training set compounds including an array of clastogens and aneugens (**Table *3.1.***) and processed for flow cytometric analysis at the 4- and 24-hour timepoints. Literature information was used to determine if an exogenous source of metabolic activation (S9) would be required. Only data from the 24-hour timepoint was included in this analysis.

Table 3.1. Compounds Studied in the MultiFlow Assay

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
AMG-900	amg	945595-80-2	Genotoxic; Aneugen
Carbendazim	car	10605-21-7	Genotoxic; Aneugen
Colchicine	col	64-86-8	Genotoxic; Aneugen
Diethylstilbestrol	des	56-53-1	Genotoxic; Aneugen
17β-Estradiol	est	50-28-2	Genotoxic; Aneugen
Fisetin	fis	528-48-2	Genotoxic; Aneugen
Flubendazole	flu	31430-15-6	Genotoxic; Aneugen
Griseofulvin	gli	126-07-8	Genotoxic; Aneugen
Mebendazole	meb	31431-39-7	Genotoxic; Aneugen
Nocodazole	noc	31430-18-9	Genotoxic; Aneugen
Noscapine	nos	128-62-1	Genotoxic; Aneugen
Paclitaxel	pac	33069-62-4	Genotoxic; Aneugen
Vinblastine sulfate	vin	143-67-9	Genotoxic; Aneugen
Vincristine sulfate	vis	2068-78-2	Genotoxic; Aneugen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
2-Acetylaminofluorine	aaf	53-96-3	Genotoxic; Clastogen
Aphidicolin	aph	38966-21-1	Genotoxic; Clastogen
Azathioprine	aza	446-86-6	Genotoxic; Clastogen
Azidothymidine	azt	30516-87-1	Genotoxic; Clastogen
Benzo[a]pyrene	bap	50-32-8	Genotoxic; Clastogen
Bleomycin sulphate	bls	9041-93-4	Genotoxic; Clastogen
Camptothecin	cam	7689-03-4	Genotoxic; Clastogen
Chlorambucil	chl	305-03-3	Genotoxic; Clastogen
Cisplatin	cis	15663-27-1	Genotoxic; Clastogen
Cyclophosphamide monohydrate	cpm	6055-19-2	Genotoxic; Clastogen
Cytosine arabinoside	cya	147-94-4	Genotoxic; Clastogen
7,12-Dimethylbenzanthracene	dba	57-97-6	Genotoxic; Clastogen
Ethyl methanesulfonate	ems	62-50-0	Genotoxic; Clastogen
Etoposide	etp	33419-42-0	Genotoxic; Clastogen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
5-Fluorouracil	ful	51-21-8	Genotoxic; Clastogen
Glycidamide	gly	5694-00-8	Genotoxic; Clastogen
Hydralazine HCl	hyd	304-20-1	Genotoxic; Clastogen
Hydrogen peroxide	hyp	7722-84-1	Genotoxic; Clastogen
Hydroxyurea	hyu	127-07-1	Genotoxic; Clastogen
Melphalan	mel	142-82-3	Genotoxic; Clastogen
Menadione	men	58-27-5	Genotoxic; Clastogen
Mitomycin C	mmc	50-07-7	Genotoxic; Clastogen
Methyl methanesulfonate	mms	66-27-3	Genotoxic; Clastogen
Methotrexate	mtx	59-05-2	Genotoxic; Clastogen
4-Nitroquinoline 1-oxide	nqo	56-57-5	Genotoxic; Clastogen
Olaparib	ola	763113-22-0	Genotoxic; Clastogen
1,3-Propane sultone	psu	1120-71-4	Genotoxic; Clastogen
Thiotepa	thi	52-24-4	Genotoxic; Clastogen
Temozolomide	tmz	85622-93-1	Genotoxic; Clastogen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
Topotecan	top	123948-87-8	Genotoxic; Clastogen

3.2.1.2 MicroFlow Assay Specific Information Pertaining to the Dataset

Litron Laboratories performed the benchtop experiments to acquire the MicroFlow dataset for analysis. The assay has been extensively discussed in the published literature (Avlasevich et al., 2006; Bryce et al., 2008; 2010; Avlasevich et al., 2011). Briefly, 38 compounds were evaluated for *in vitro* micronuclei formation using human lymphoblastoid TK6 cells from the same treated cultures as the MultiFlow assay. The MicroFlow analyses were conducted at the 24-hour timepoint and accomplished *via* flow cytometric analysis. For the micronucleus endpoint, compound concentrations were limited to those that resulted in $\leq$ 55% reduction to RNCs consistent with the approach of Avlasevich et al. (2006). **Table *3.2*** summarises the subset (n=38) of clastogens and aneugens that were evaluated for micronuclei in this analysis.

Table 3.2. Compounds Studied in the MicroFlow Assay

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
Aphidicolin	adc	38966-21-1	Genotoxic; Aneugen
Azathioprine	azp	446-86-6	Genotoxic; Aneugen
Carbendazim	car	10605-21-7	Genotoxic; Aneugen
Colchicine	col	64-86-8	Genotoxic; Aneugen
Diethylstilbestrol	des	56-53-1	Genotoxic; Aneugen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
17β-Estradiol	est	50-28-2	Genotoxic; Aneugen
Flubendazole	flu	31430-15-6	Genotoxic; Aneugen
Griseofulvin	gli	126-07-8	Genotoxic; Aneugen
Mebendazole	meb	31431-39-7	Genotoxic; Aneugen
Nocodazole	noc	31430-18-9	Genotoxic; Aneugen
Noscapine	nos	128-62-1	Genotoxic; Aneugen
Paclitaxel	pac	33069-62-4	Genotoxic; Aneugen
Vinblastine sulfate	vin	143-67-9	Genotoxic; Aneugen
Vincristine sulfate	vis	2068-78-2	Genotoxic; Aneugen
Azidothymidine	azt	30516-87-1	Genotoxic; Clastogen
Benzo[a]pyrene	bap	50-32-8	Genotoxic; Clastogen
Bleomycin sulphate	bls	9041-93-4	Genotoxic; Clastogen
Camptothecin	cam	7689-03-4	Genotoxic; Clastogen
Chlorambucil	chl	305-03-3	Genotoxic; Clastogen
Cisplatin	cis	15663-27-1	Genotoxic; Clastogen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
Cytosine arabinoside	cya	147-94-4	Genotoxic; Clastogen
Ethyl methanesulfonate	ems	62-50-0	Genotoxic; Clastogen
N-Ethyl-N-Nitrosourea	enu	759-73-9	Genotoxic; Clastogen
Etoposide	etp	33419-42-0	Genotoxic; Clastogen
5-Fluorouracil	ful	51-21-8	Genotoxic; Clastogen
Glycidamide	gly	5694-00-8	Genotoxic; Clastogen
Hydralazine HCl	hyd	304-20-1	Genotoxic; Clastogen
Hydrogen peroxide	hyp	7722-84-1	Genotoxic; Clastogen
Hydroxyurea	hyu	127-07-1	Genotoxic; Clastogen
Menadione	men	58-27-5	Genotoxic; Clastogen
Mitomycin c	mmc	50-07-7	Genotoxic; Clastogen
Methyl methanesulfonate	mms	66-27-3	Genotoxic; Clastogen
4-Nitroquinoline 1-oxide	nqo	56-57-5	Genotoxic; Clastogen
Olaparib	ola	763113-22-0	Genotoxic; Clastogen
1,3-Propane sultone	psu	1120-71-4	Genotoxic; Clastogen

Compound	Compound Abbreviation	CAS No.	*a priori* Mammalian Cell Genotoxicity & MoA Classifications
Thiotepa	thi	52-24-4	Genotoxic; Clastogen
Temozolomide	tmz	85622-93-1	Genotoxic; Clastogen
Topotecan	top	123948-87-8	Genotoxic; Clastogen

3.2.2 Database Curation

The PROAST software requires data to be formatted for analysis including the following specific strict rules:

- All decimal points in point format not comma format
- Empty cells are not allowed
- Spaces between text or units are not allowed
- The raw data takes the format of the data matrix. In this case, the data matrix was the compound names, concentration values, a priori mode of action information, median fluorescence, median relative frequency, or percent values.
- Each column must have a header which describes the data in the data matrix
- Four rows must exist above the data matrix and complete according to the following rules:
  - First line: a one-word title that describes the data. In this case the MicroFlow data had the heading "MicroFlow", and the MultiFlow data had the heading "MultiFlow".
  - Second line: the number of columns in the data matrix
  - Third line: includes a numerical code for the data type in the column (see below)
  - Fourth line: a one-word title for the column. The column title choices for the MultiFlow and MicroFlow datasets are presented in **Table 3.3** and **Table 3.4**.
  - Fifth and further lines: the data matrix.
- Data types (third line)

- o 0 – for all columns not representing a response variable
- o 1 – continuous data. In our case, all MultiFlow biomarker response variables (γH2AX, p-H3, p53), and all MicroFlow median relative percent micronuclei scores are continuous data.
- o 2 – binary data (e.g., yes/no response in individual animals)
- o 3 – ordinal data (e.g., histopathology scores of severities)
- o 4 – quantal data (e.g., number of responding animals per number of animals in the group)
- o 5 – continuous data that are nested (e.g., fetal weights within litters)
- o 6 – quantal data that are nested (e.g., number of fetuses affected within litters)
- o 10 – mean (continuous) response in a dose group

Data types 2 to 10 are out of scope of the MultiFlow and MicroFlow assay result variables. Hence the third line for the MicroFlow and MultiFlow databases all included either a value of 0 or 1.

Since not all variables are relevant to the hypotheses that were being tested in this study, some columns were excluded from analyses. For example, although the MultiFlow assay provides information relative to polyploidy through the evaluation of propidium iodide gating logic – said column was excluded from the database. Percent RNC and RPD columns were maintained since the data are indicative of treatment related cytotoxicity effects which could prove important in interpreting the BMD analysis outputs. Additionally, the MicroFlow data includes information on cell cycle effects which were not in scope of the hypotheses tested in this study and so the columns were removed. However, the percent RNC column was maintained as well as the cytotoxicity column. Cytotoxicity is calculated at 100% minus the %RNC at 24 hours. An extra column was added to each of the MicroFlow and MultiFlow databases to serve as a three-letter compound code. The three-letter compound code abbreviation information was added purposefully to reduce clutter in the graphical outputs that would be caused due to lengthy compound names. Some examples include 'est' for '17β-Estradiol' and 'mmc' for 'mitomycin c'. Refer to sections 3.2.1.1 and 3.2.1.2 for the full list of compounds and their abbreviations of which were tested in each assay. **Table 3.3** and **Table 3.4** summarise the final selected database columns, rows, and headers formats for the MultiFlow and MicroFlow datasets, respectively.

Table 3.3. Replication of the MultiFlow database columns, rows, and headers formatted per PROAST strict data formatting requirements

Row Number									
1	MultiFlow								
2	9								
3	0	0	0	1	1	1	1	1	1
4	Compound	Compound.abbreviation	a.priori.MOA	Conc. μM	g.H2AX[2]	pH3[2]	p53[2]	RNC.perc[2]	Cytotoxicity.perc[3]
5 and more…	Data Matrix…								

---

[2] All MultiFlow biomarker data was acquired at the 24-hour sampling time
[3] Cytotoxicity was calculated at 100% minus the %RNC at 24-hours.

Table 3.4. Replication of the MicroFlow database columns, rows, and headers formatted per PROAST strict data formatting requirements

Row Number						
1	MicroFlow					
2	6					
3	0	0	0	1	1	1
4	Compound	Compound.abbreviation	a.priori.MOA	Conc. µM	MN.perc[4]	RNC.perc[4]
5 and more…	Data Matrix…					

---

[4] All MicroFlow data were aquired at the 24-hour sampling time

Once the Excel databases were formatted according to the strict rules prescribed by PROAST, the Excel sheets were transformed into tab delimited files (.txt) and stored in a working repository.

3.2.3 Working in PROAST

PROAST v61.2 operating in the R statistical suite was used for the BMD analyses presented in this chapter. The 'MENU' version of PROAST was used to navigate through the data and variables selection process in the BMD analyses. Firstly, the .txt files were imported into PROAST and analysed independently. The 'continuous' data type was selected from the MENU option, and 'concentration' selected as the 'independent variable'. Settings were changed to analyse the clastogens separately from the aneugens. Upon being prompted to select whether a 'covariate' is required, 'compound' was selected as the covariate. Additionally, based on multiple trials of prior feasibility testing of compound covariate BMD analysis, further subdivision of large groups of compounds was determined to be necessary since PROAST cannot handle the analysis of large groups of compounds when the modelling software is operating on a standard desktop computer. Since there were 24 clastogens for analysis with micronuclei frequency data, and 30 clastogens for analysis with MultiFlow biomarker data, further subdivision was necessary into 2 groups of clastogens for each assay. Groups of 15 compounds for analysis was a feasible number based on the prior trials. Therefore, the clastogens were subdivided into 2 groups in alphabetical order according to the following summary:

- MultiFlow assay biomarker response data – Clastogens group A (clastogensa): aaf to etp; Clastogens Group B (clastogensb): ful to top.
- MicroFlow assay micronuclei frequency data – Clastogens Group A (clastogena): adc to gly; Clastogens Group B (clastogensb): hyd to top.

Since there was a manageable group of aneugens for each test system (n=14), no further subdivision was deemed necessary and all aneugens were analysed together in the 'compound covariate' selection subgroup. Further successful navigation through the PROAST MENU interface was based on in-depth knowledge of the BMD modelling approach, and the steps are described in the following sections.

3.2.4 BMD Analysis

BMD analysis of the MultiFlow DNA damage response markers was performed according to the appropriate biomarker to assess the *a priori* mode of action. (Bryce et al., 2016; 2017; 2018) previously described that the clastogenic mode of action is best predicted by increases in the MultiFlow assay's γH2AX and p53 biomarker responses, whereas increases in the MultiFlow assay's phosphohistone-H3 (pH3) and p53 responses are associated with an aneugenic mode of action. Therefore, the 24-hour γH2AX and p53 responses were analysed for clastogens, and the pH3 and p53 responses were analysed for aneugens in the compound subgroups described in section 3.2.3.

The automatic selection of the maximal exponential model was selected from the PROAST menu which defaulted to model 5. The exponential model 5 equation is presented below for the analysis of all endpoints (24-hour γH2AX, p53, p-H3 and micronuclei percent). As described in the Materials and Methods chapter of this thesis (Chapter 2.3), model 5 represents the 4-parameter model where the following information is considered: a) the response within the controls; b) potency/efficacy of the compound on the shape of the response; c) the maximum fold-change in the response; and d) the rate of change in the response. A Hill model was not included in the analyses. The maximal exponential model was selected based on it containing the greatest number of parameters for describing the dose response. The choice to exclude a Hill model from the analyses was based on the work of (Slob and Setzer, 2014) where it was demonstrated that the shape and steepness of dose-response curves do not differ between the results of either exponential or Hill models.

Exponential Model 5: $y = a\left[c - (c-1)\exp\left(-\frac{x}{b}\right)^d\right] \ with \ a > 0, b > 0, c > 0, d \geq 1,$

where y denotes the continuous endpoint, and the dose by x. Parameter a = the level of the endpoint at dose 0; b = potency/efficacy of the test compound on the shape of the response; c = maximum fold-change in response; and d = rate of change in the response constrained to values equal to or larger than 1.

3.2.4.1 Critical Effect Size

Since BMD modelling had not previously been performed on MultiFlow biomarker response data, there was obviously no prior experience of the critical effect size (CES) to select to attempt to model the response data. On the other hand, one published example of

BMD modelling of micronucleus frequency data had examined the impact of selecting different CESs on the resulting BMD metrics. Bemis et al. (2016) had applied several CESs to the *in vitro* MicroFlow micronucleus frequency data used in their analysis, with CESs of 5, 50, 100, 200, and 400%. The authors showed that the BMD confidence intervals are similar when calculated using different CESs, and that the BMD confidence interval precision (ratio between the BMDU/BMDL) hardly changes with differing CES values. For this reason, an arbitrary CES value of 50% was originally selected for the BMD analyses of MicroFlow micronuclei frequency data presented in this chapter. The same original CES of 50% was applied consistently to the MultiFlow biomarker response datasets since for comparison purposes. However, three clastogens (mitomycin c, 4-Nitroquinoline 1-oxide, and topotecan) as a further covariate were iteratively re-analysed using a CES of 100% for their micronuclei responses only based on poor performance with a CES of 50% (more detail in the results section of this chapter). BMD model curves were formulated for each of the compound covariate groups and for each level (compound) of the covariate for compound specific dose-response scrutiny.

3.2.4.2 Confidence Interval Determination

The PROAST standard 90% confidence intervals were calculated for each level (compound) of the covariate and save in an R data file. The confidence intervals were automatically plotted by PROAST in a Log plot. Note that PROAST v61.2 standard output did not organise each level of the covariate (compound) in order of potency rank order.

3.2.4.3 System Cross-Correlation

To perform system cross-correlations, the previously determined 90% confidence intervals for each covariate of the MicroFlow and MultiFlow assay biomarker responses were re-imported into PROAST and the 'f.bigtable' function was used to collate a tabulated format for the confidence interval values. In doing so, irrelevant columns were removed from the file by following the steps in the PROAST menu options. A .txt file was generated by PROAST that contained the BMD confidence bounds and the relevant labels for each of the associated factors that were originally present in the other columns of the dataset. The file was saved in the working directory as a 'bigtable'. A 'bigtable' was created for each endpoint's BMD analysis, i.e., the micronuclei percent, 24-hour γH2AX, p53 median fluorescence and p-H3 frequency data. Finally, a new working directory was created and the relevant 'bigtable' .txt

files were copied into the new working directory and the 'f.cor.table' PROAST function was used to begin the process of iteratively creating a BMD correlation plot. A 'crosstable' was created, where the confidence intervals from both systems are present for each dataset according to the chosen covariate. The 'crosstable' was used upon entering the 'f.cor.plot' command with 2 different endpoints of choice labelled on the x and y axes, respectively. Since the goal was to correlate the confidence intervals from compounds assessed for micronuclei frequency with relevant MultiFlow biomarkers, the following correlations were created based on the compound mode of action under study:

- Clastogens: Micronucleus frequency versus γH2AX and p53 biomarkers
- Aneugens: Micronucleus frequency versus p-H3 and p53 biomarkers

3.3 Results

3.3.1 MultiFlow Assay Biomarker BMD Response Modelling

BMD modelling of the MultiFlow datasets was performed with compound as covariate per endpoint combination as described in the materials and methods section of this chapter. The results of the BMD modelling are presented in the next subsections of this chapter, thus informing the responsiveness of the biomarker data to BMD modelling.

3.3.1.1 Clastogens

Clastogens were subdivided into two groups for covariate analysis as described in section 3.2.3: Clastogens group A (clastogensa): aaf to etp; Clastogens Group B (clastogensb): ful to top. Refer to **Table 3.1.** for the full list of compound abbreviations. BMD model analysis outputs for the clastogen compounds are presented in the subsections herein.

3.3.1.1.1 γH2AX
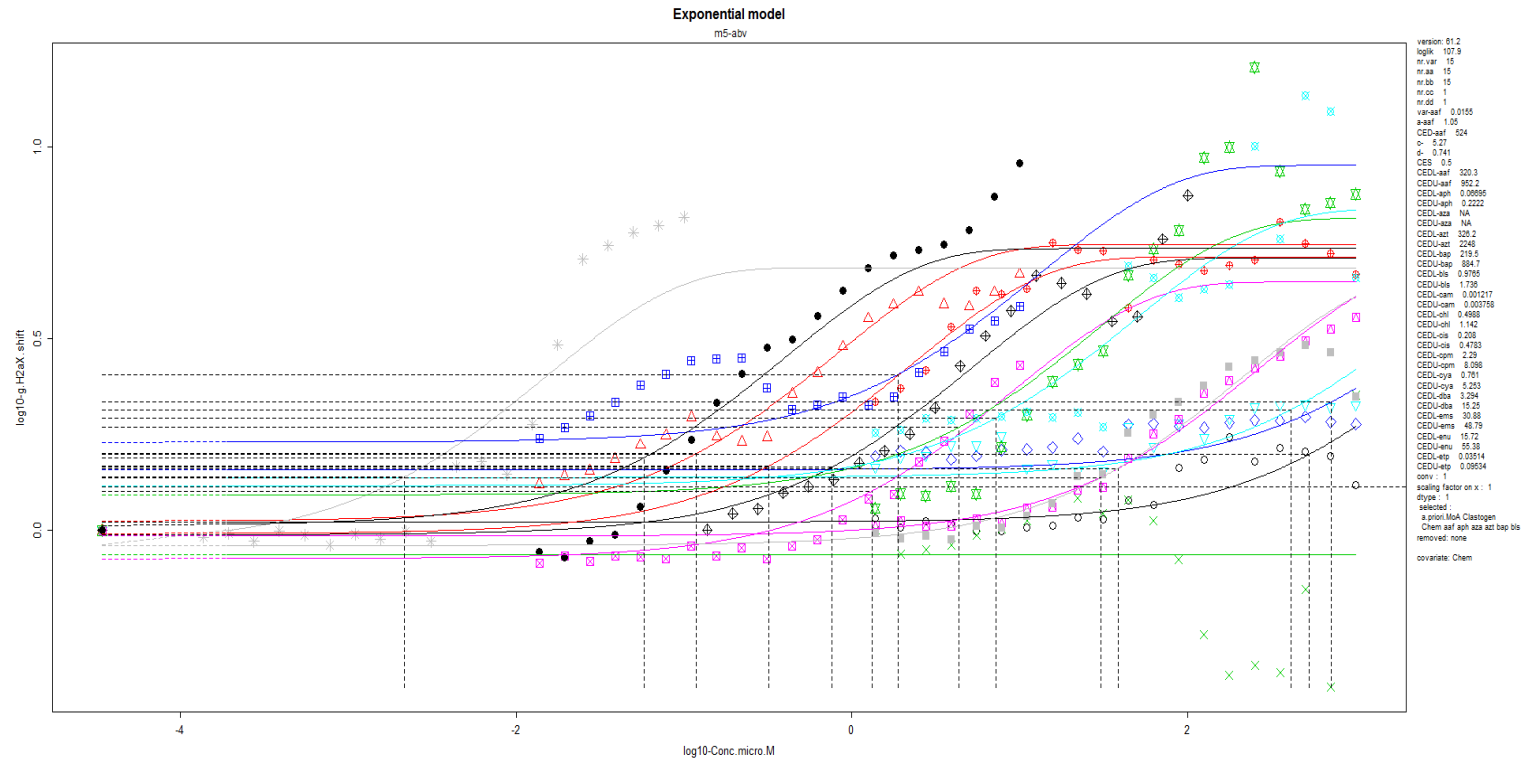
Figure 3.10. BMD modelling curves for the 'clastogensa' subgroup analysed with compound as the covariate level for γH2AX dose-responses. Double-log plot with concentration in Log10-μM on the *x*-axis and Log-10 γH2AX median fluorescence response on the *y*-axis. Coloured points and curves correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compound except for 'aza'.
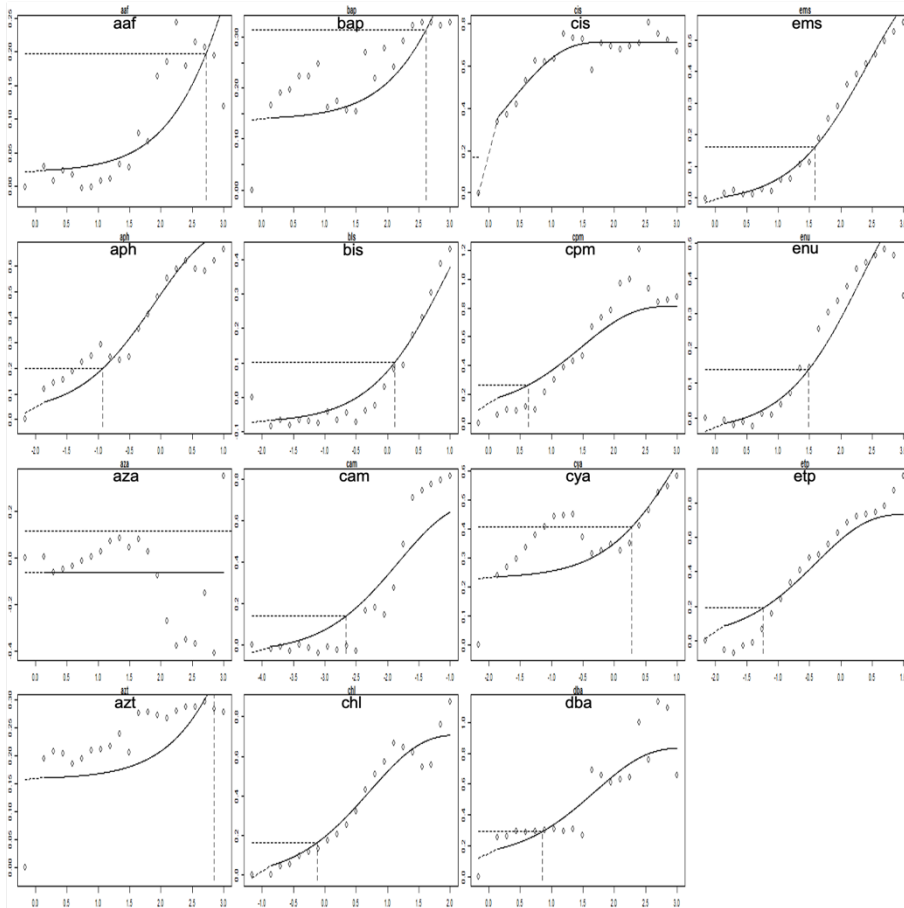
Figure 3.11. PROAST generated individual dose-response curves for each level of the 'clastogensa' covariate γH2AX biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 γH2AX median fluorescence response on the *y*-axis. Further scrutiny of 'aza' dose response shows a weak dose-response with a reduction of γH2AX response in the top 10 concentrations evaluated.
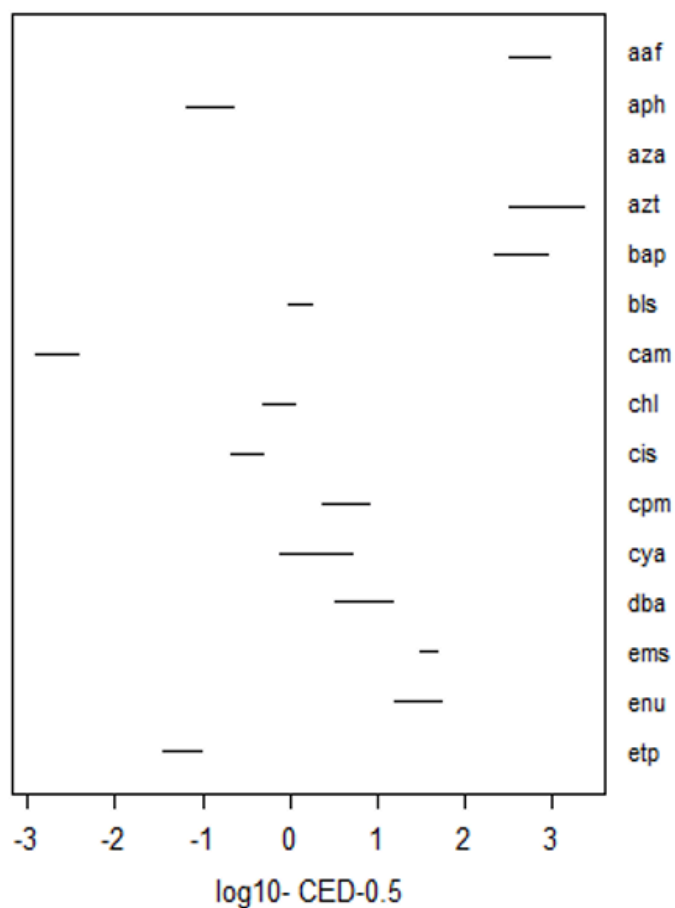
Figure 3.12. BMD confidence intervals plotted for the 'clastogensa' covariate subgroup γH2AX biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'aza' which did not calculate a confidence interval since there was no dose-response for said compound. The bound confidence intervals spanned a maximum distance of 1-Log unit.
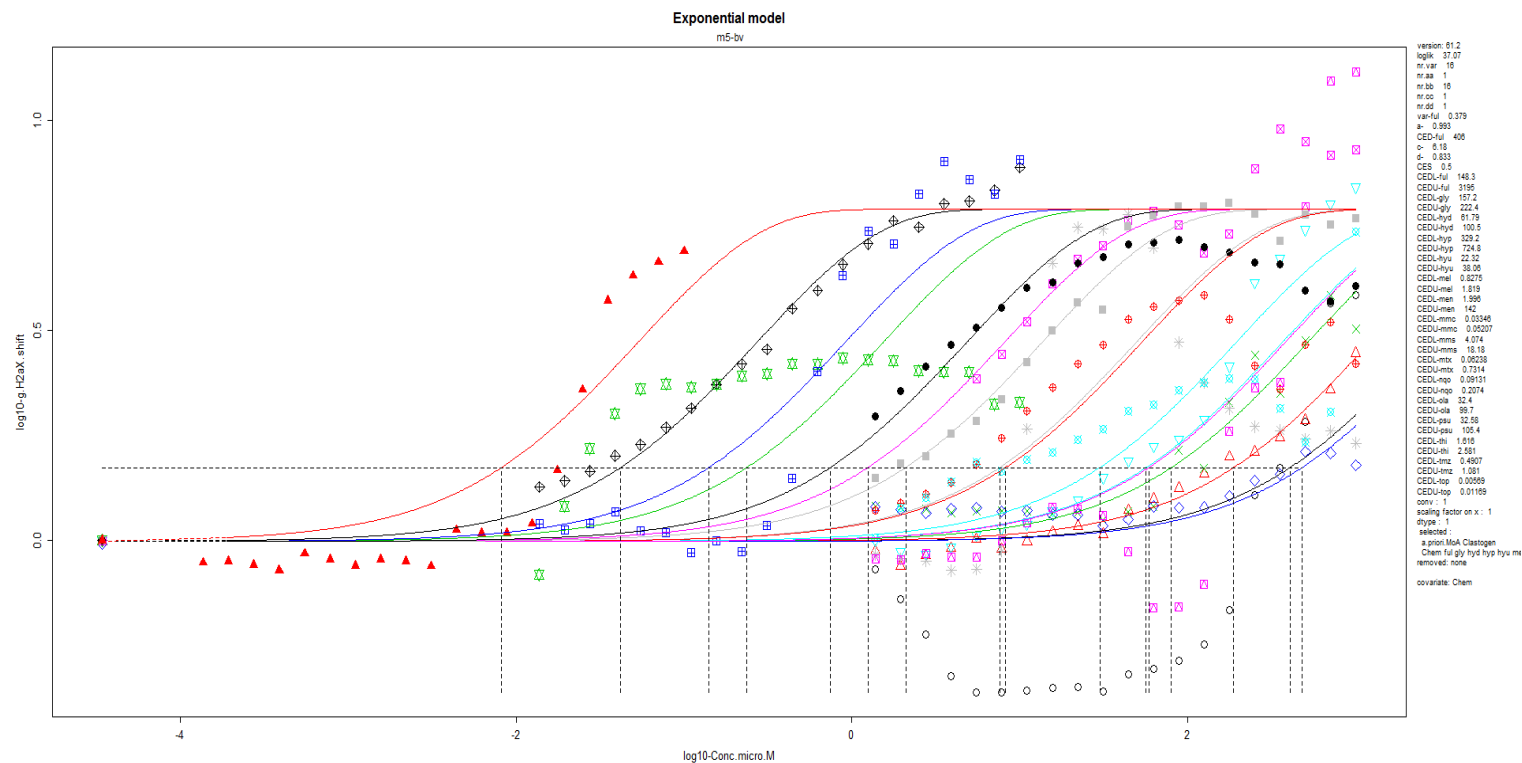
Figure 3.13. BMD modelling curves for the 'clastogensb' subgroup analysed with compound as the covariate level for γH2AX biomarker dose-responses. Double-log plot with concentration in Log10-µM on the *x*-axis and Log-10 γH2AX median fluorescence response on the *y*-axis. Coloured points and curves correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds.
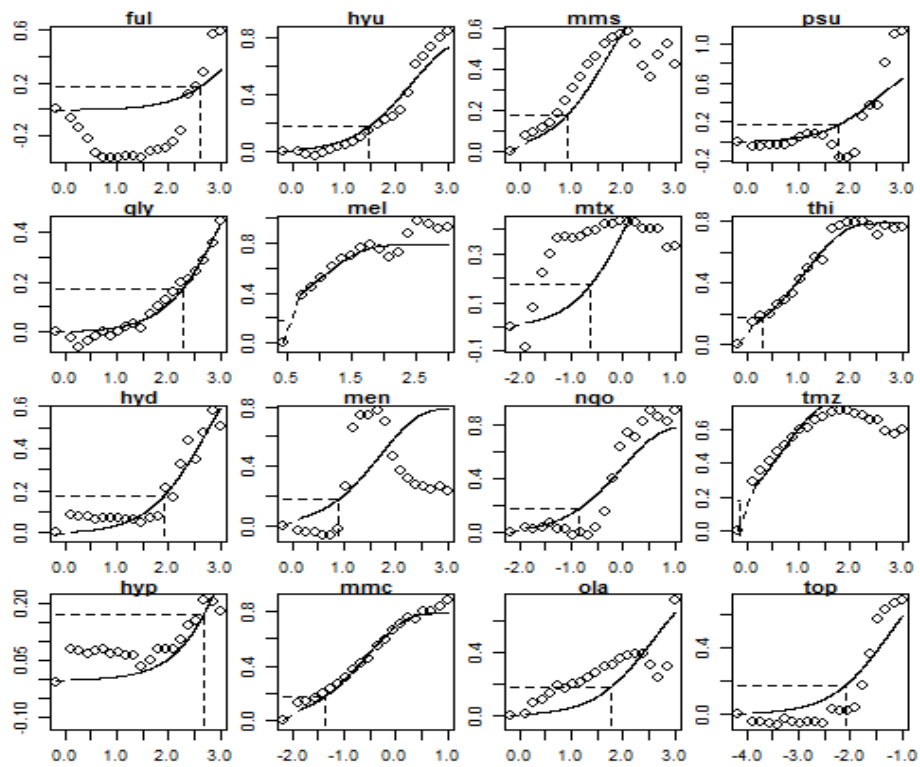
Figure 3.14. PROAST generated individual dose-response curves for each level of the 'clastogensb' covariate γH2AX biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 γH2AX median fluorescence response on the *y*-axis. Fitted dose-response curves for all compounds. The 'ful' and 'men' compounds exhibit a different dose-response curve shape compared to the other compounds.
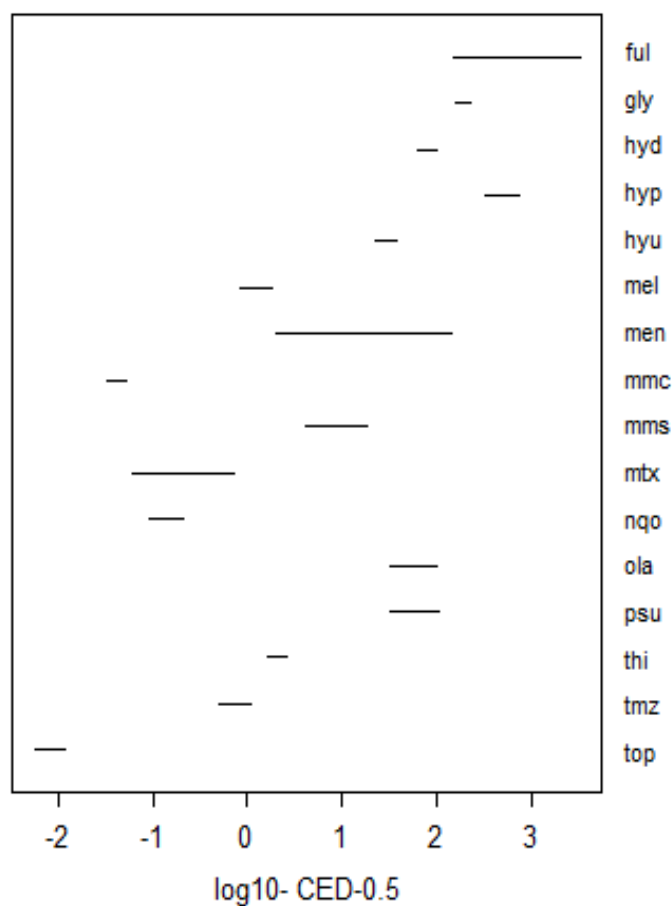
Figure 3.15. BMD confidence intervals plotted for the 'clastogensb' covariate subgroup γH2AX biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals. The confidence intervals spanned a maximum distance of 1-Log unit except for the 'ful' and 'men' compounds whose confidence intervals spanned between approximately 1.5- and 2.5-Log units, respectively.

3.3.1.1.2 p53
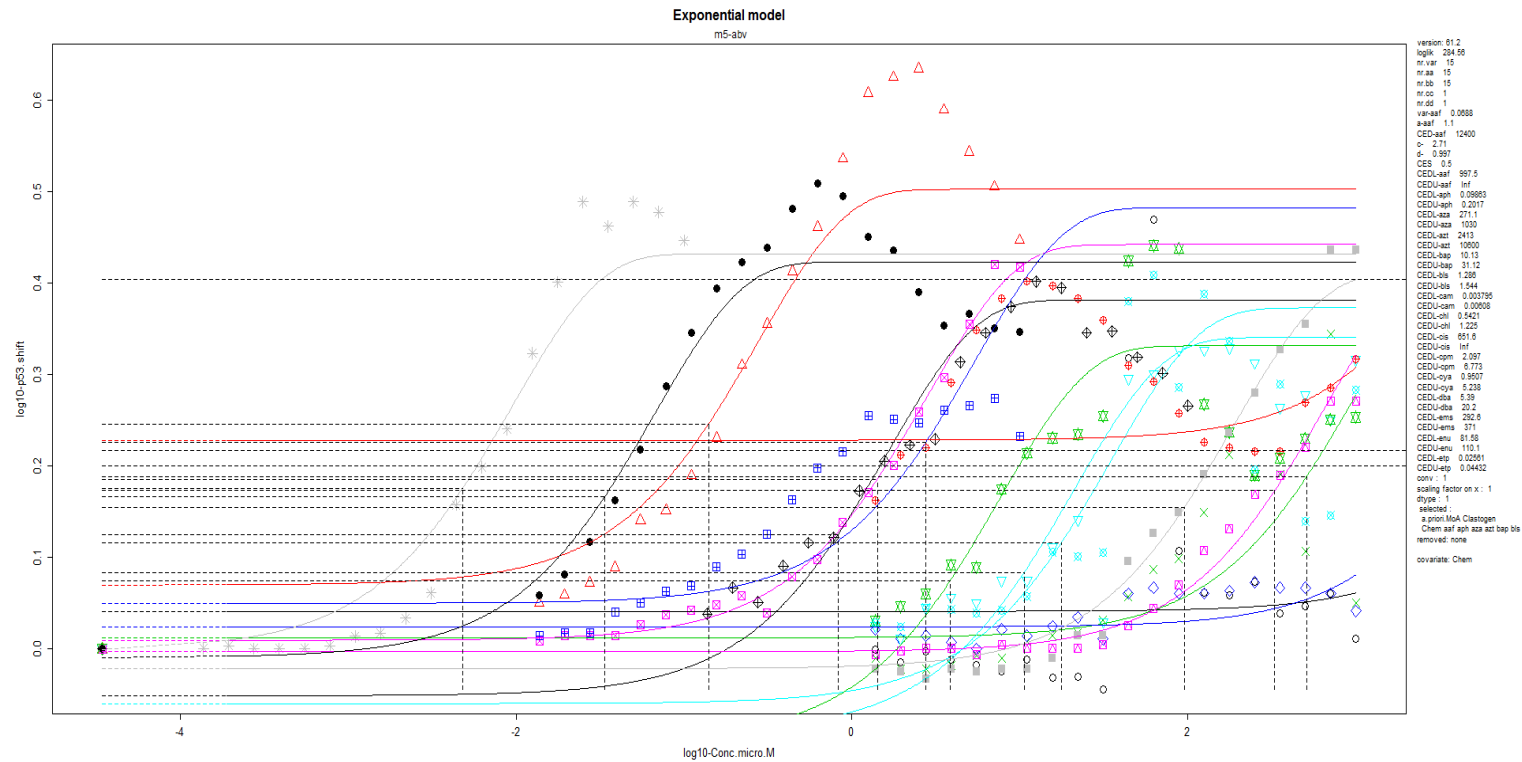
Figure 3.16. BMD modelling curves for the 'clastogensa' subgroup analysed with compound as the covariate level for p53 biomarker dose-responses. Double-log plot with concentration in Log10-μM on the *x*-axis and Log-10 p53 median fluorescence response on the *y*-axis. Coloured points and curves correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'aaf' and 'cis' retuned infinite CEDUs.
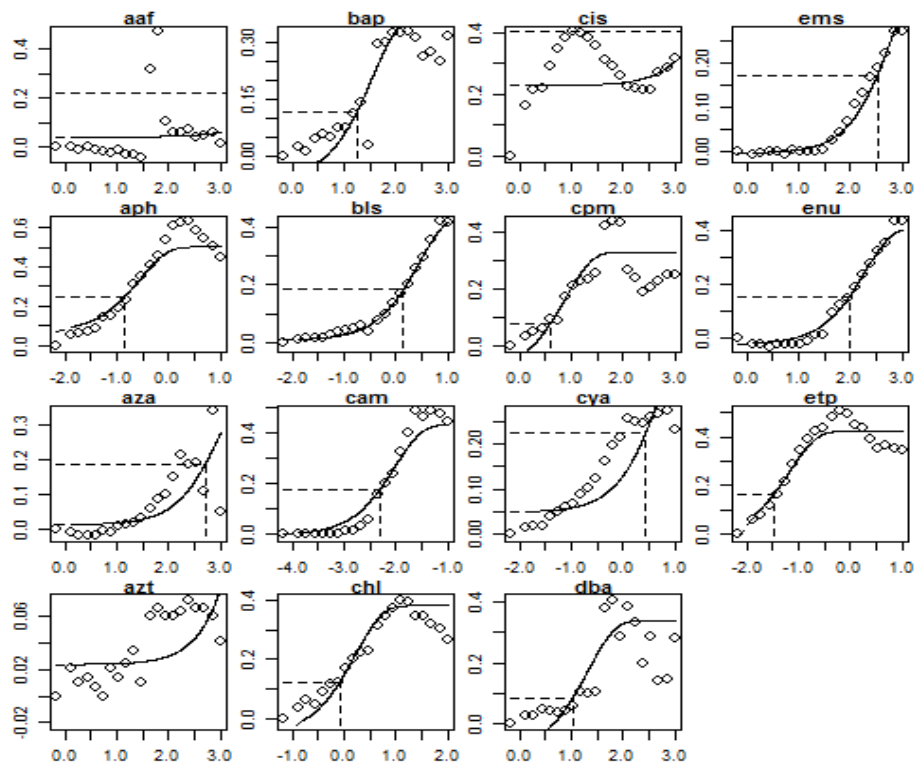
Figure 3.17. PROAST generated individual dose-response curves for each level of the 'clastogensa' covariate p53 biomarker dose-responses. Double-log plot with concentration in Log10 μM on the x-axis and the Log10 p53 median fluorescence response on the y-axis. Fitted dose-response curves for all compounds. The 'aaf' compound p53 responses did not exhibit a dose-response curve with concentration-response scatter observed in the compound specific plot. The 'cis' compound exhibited a different dose-response curve shape compared to the other compounds.
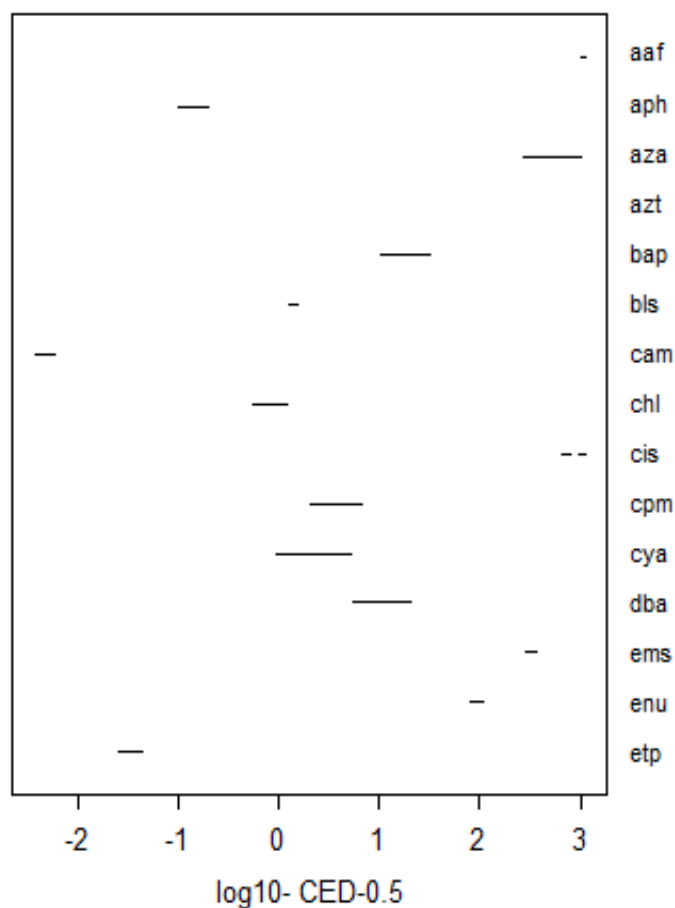
Figure 3.18. BMD confidence intervals plotted for the 'clastogensa' covariate subgroup p53 biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'aaf' and 'cis' whose dose-response confidence intervals returned an infinite upper bound evident with the dashed drawn confidence interval in the case of 'cis'. The bound confidence intervals spanned a maximum distance of 1-Log unit. There are some very small precise BMD confidence intervals, notably 'bls', 'ems' and 'enu'.
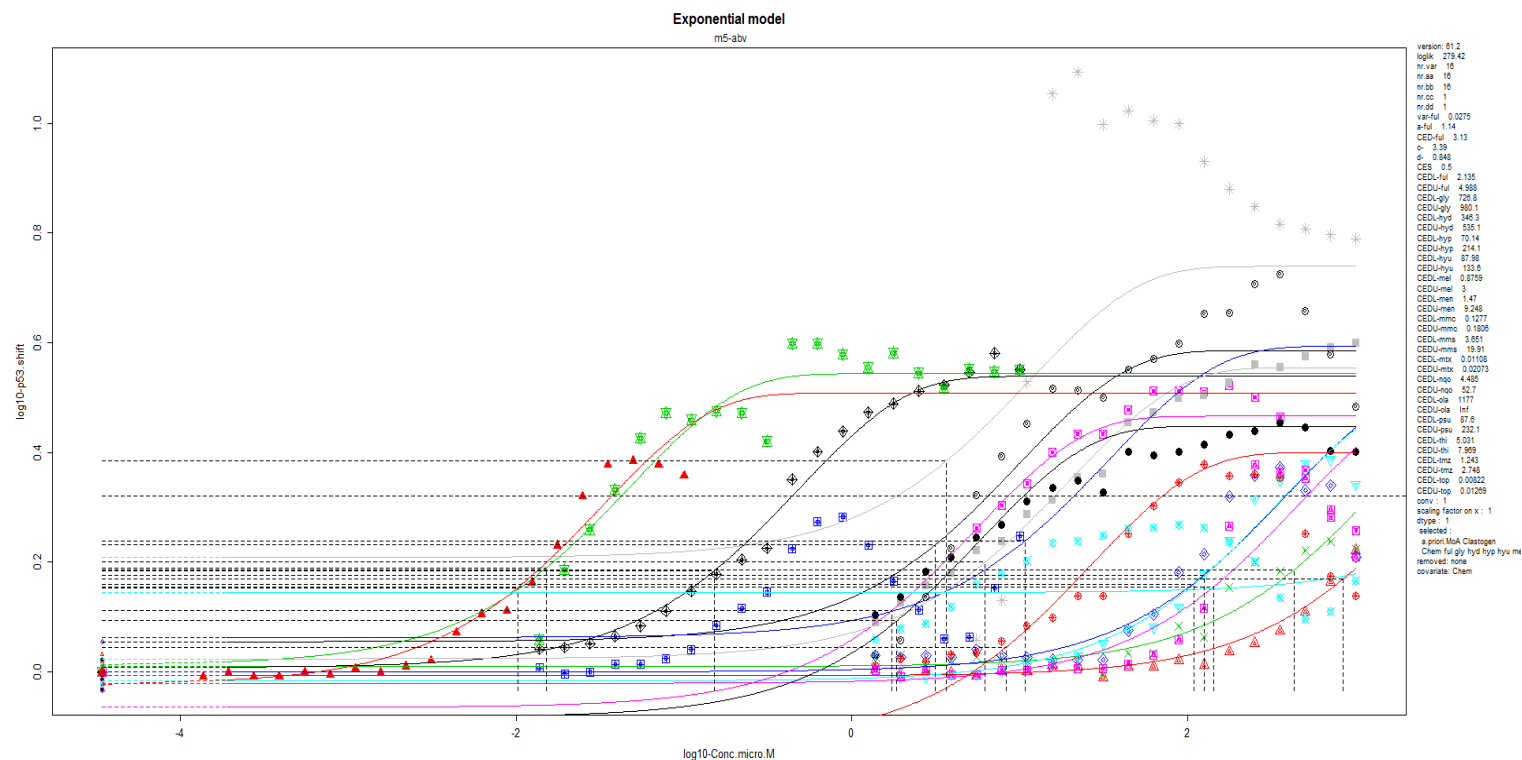
Figure 3.19. BMD modelling curves for the 'clastogensb' subgroup analysed with compound as the covariate level for p53 biomarker dose-responses. Double-log plot with concentration in Log10-µM on the *x*-axis and Log-10 p53 median fluorescence response on the *y*-axis. Coloured points and curves correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'ola' retuned an infinite CEDU.
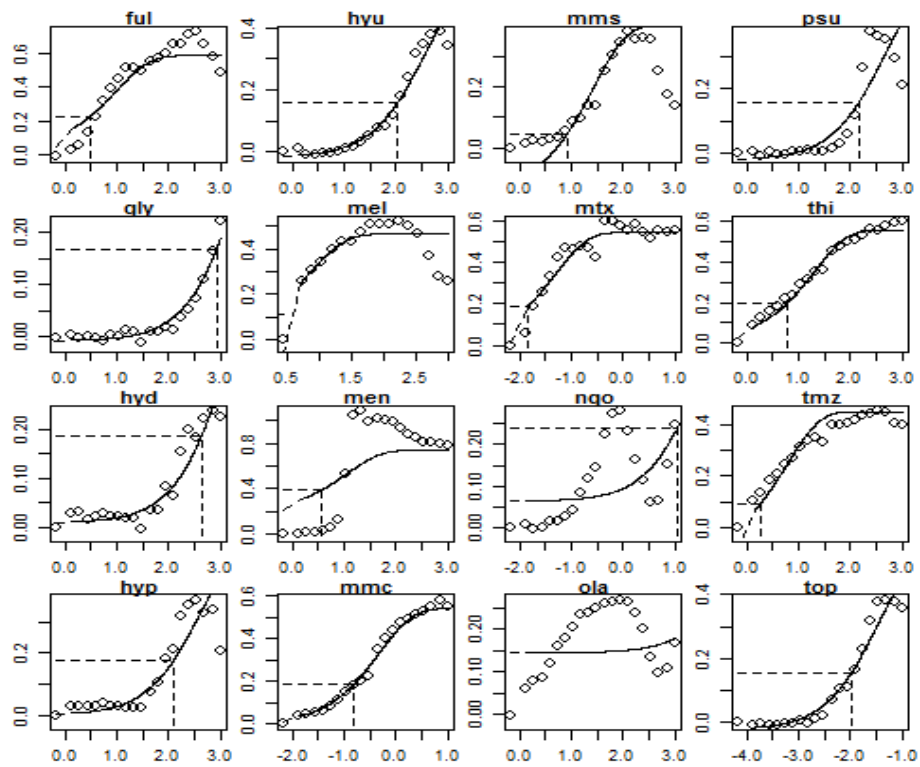
Figure 3.20. PROAST generated individual dose-response curves for each level of the 'clastogensb' covariate p53 biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 p53 median fluorescence response on the *y*-axis. Fitted dose-response curves for all compounds. The 'nqo' and 'ola' compounds p53 responses exhibited a scatter amongst the dose-response curves, particularly at higher concentrations.
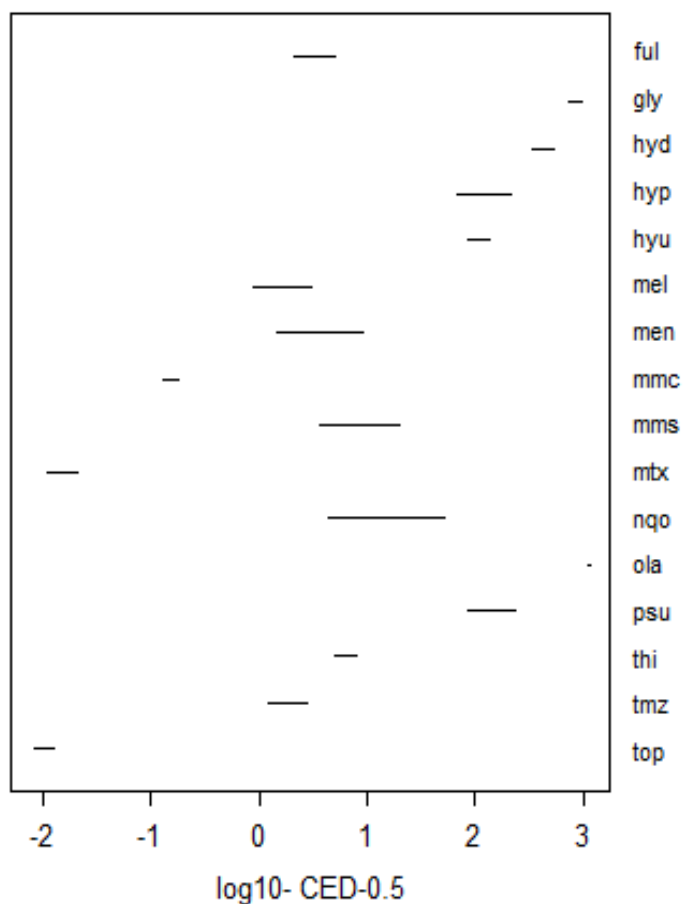
Figure 3.21. BMD confidence intervals plotted for the 'clastogensb' covariate subgroup p53 biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'ola' whose dose-response confidence interval returned an infinite upper bound. The bound confidence intervals spanned a maximum distance of approximately 1.5-Log unit. There are several small precise BMD confidence intervals.

3.3.1.2 Aneugens

No subdivision of the aneugen compounds was necessary as described in section 3.2.3. Refer to **Table 3.1.** for the full list of compound abbreviations. BMD model analysis outputs for the aneugen compounds are presented in the subsections herein.

3.3.1.2.1 p-H3
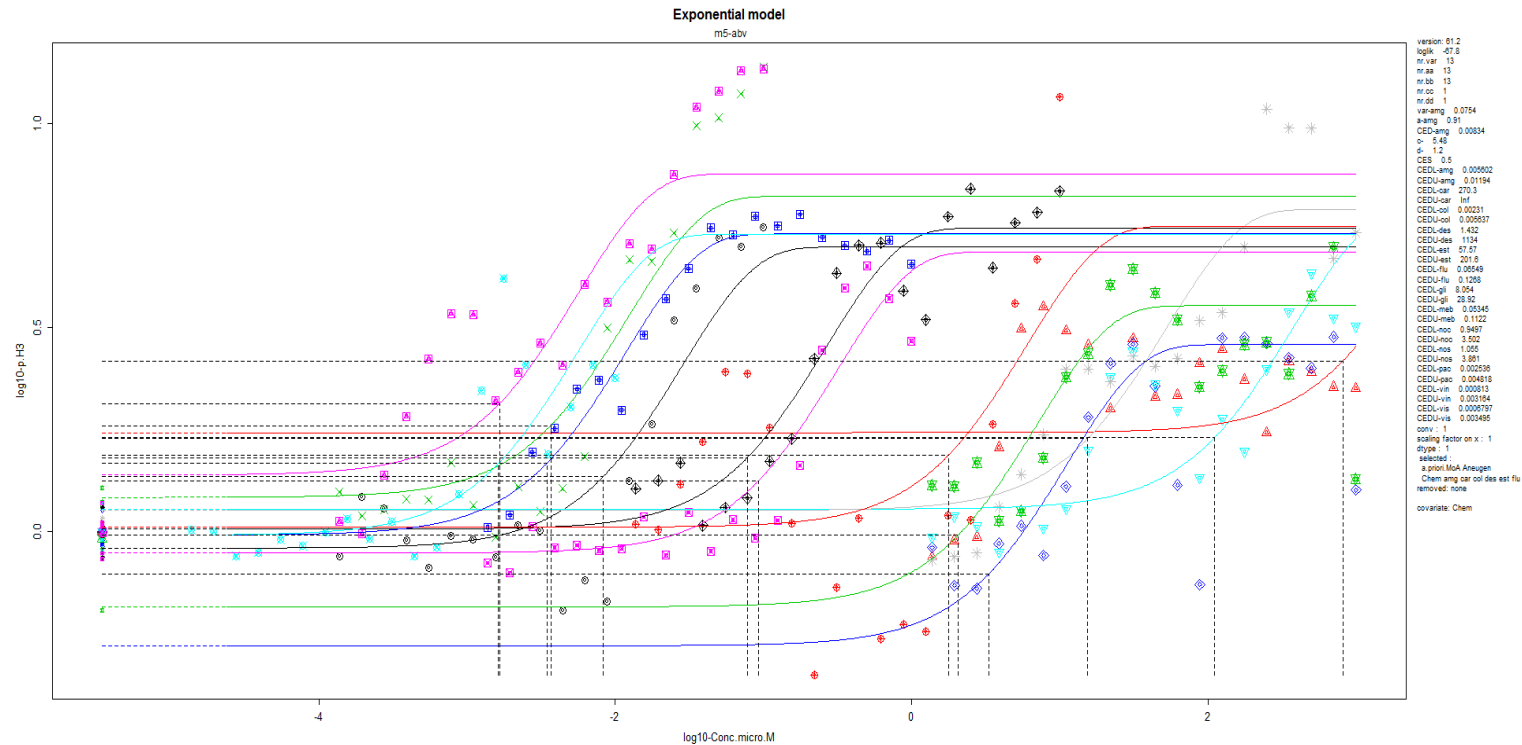
Figure 3.22. BMD modelling curves for the 'aneugen' subgroup analysed with compound as the covariate level for p-H3 biomarker dose-responses. Double-log plot with concentration in Log10-µM on the *x*-axis and Log-10 p-H3 frequency response on the *y*-axis. Coloured points and curves correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'car' retuned an infinite CEDU.
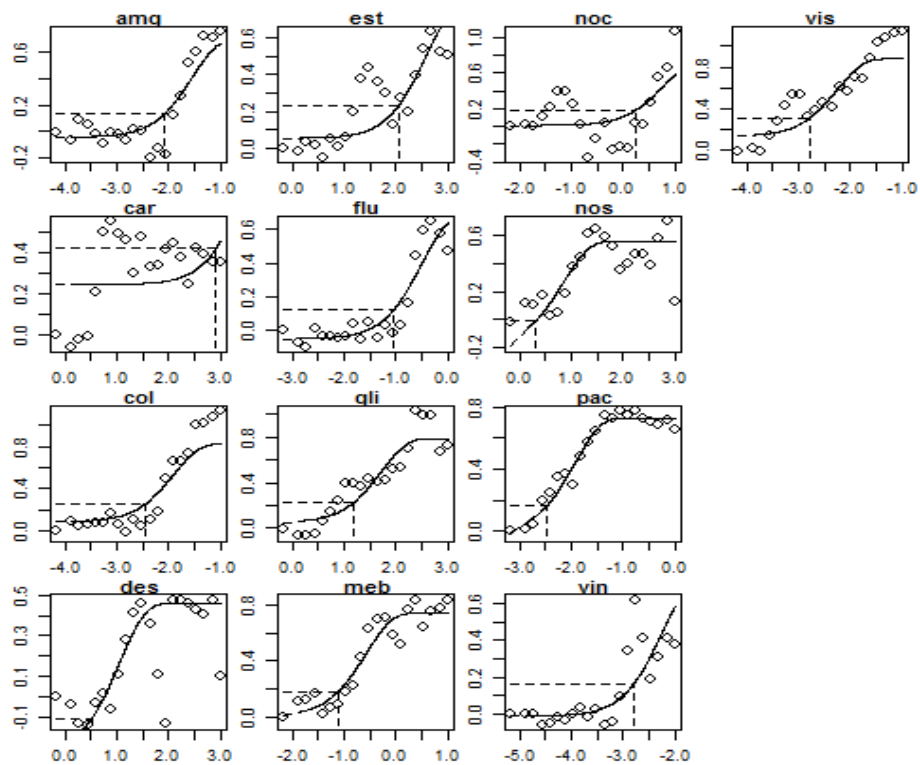
Figure 3.23. PROAST generated individual dose-response curves for each level of the 'aneugen' covariate p-H3 biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 p-H3 frequency response on the *y*-axis. Fitted dose-response curves for all compounds. The 'car' 'des' and 'est' compounds exhibited some scatter.
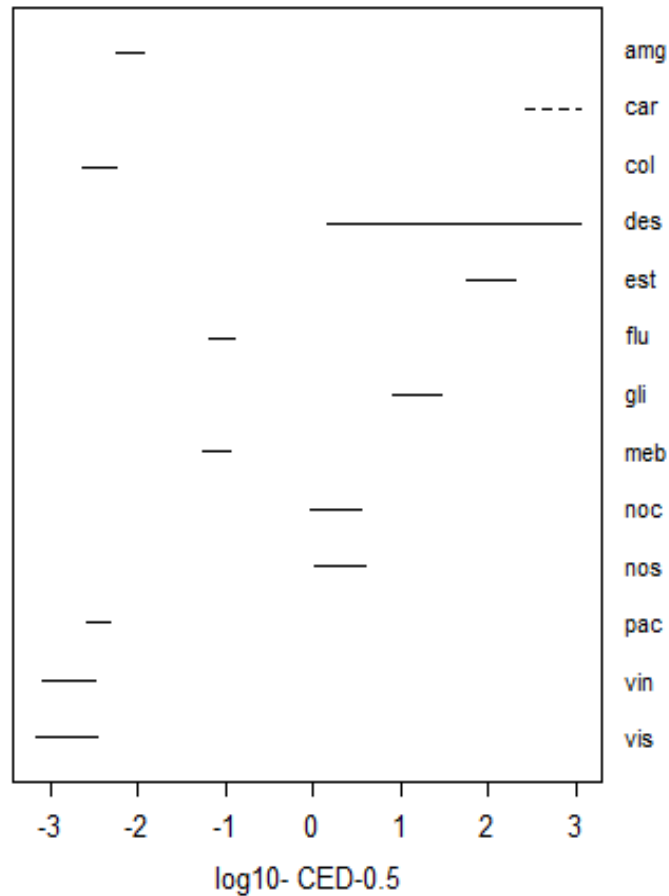
Figure 3.24. BMD confidence intervals plotted for the 'aneugens' covariate subgroup p-H3 biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'car' whose dose-response confidence interval returned an infinite upper bound evident with the dashed drawn confidence interval. The bound confidence intervals spanned a maximum distance of approximately 1-Log unit except for 'des' whose confidence interval spanned over 3-Log units. There are several small precise BMD confidence intervals.

3.3.1.2.2 p53
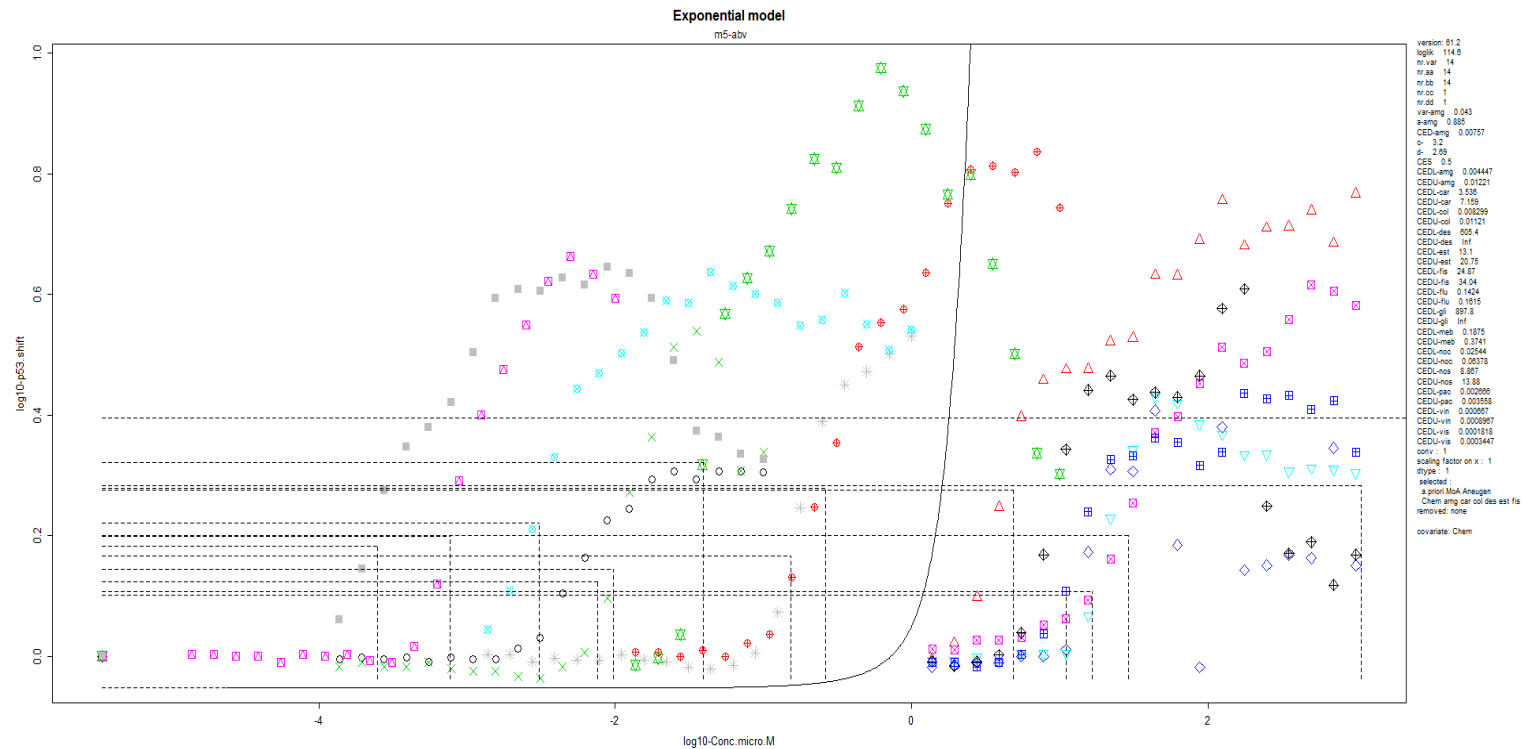
Figure 3.25. BMD modelling curves for the 'aneugen' subgroup analysed with compound as the covariate level for p53 biomarker dose-responses. Double-log plot with concentration in Log10-μM on the *x*-axis and Log-10 p53 median fluorescence response on the *y*-axis. Coloured points correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'des' and 'gli' retuned infinite CEDUs.
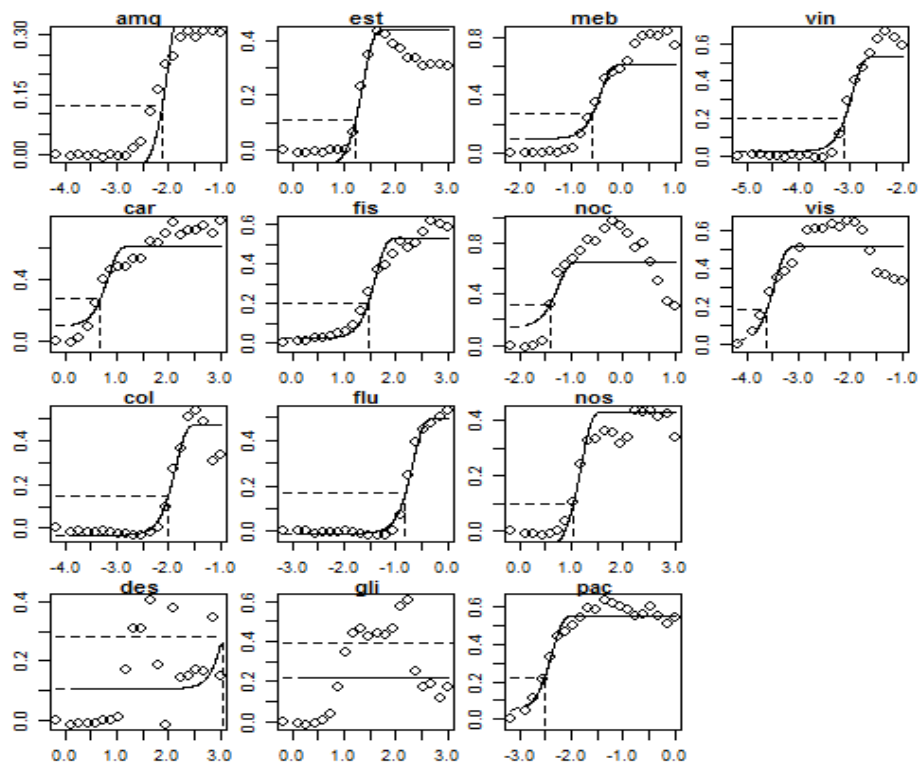
Figure 3.26. PROAST generated individual dose-response curves for each level of the 'aneugen' covariate p53 biomarker dose-responses. Double-log plot with concentration in Log10 µM on the *x*-axis and the Log10 p53 median fluorescence response on the *y*-axis. Fitted dose-response curves for all compounds. The 'des' and 'gli' compounds exhibited some scatter, with a weak dose-response visible in the 'gli' individual curve.
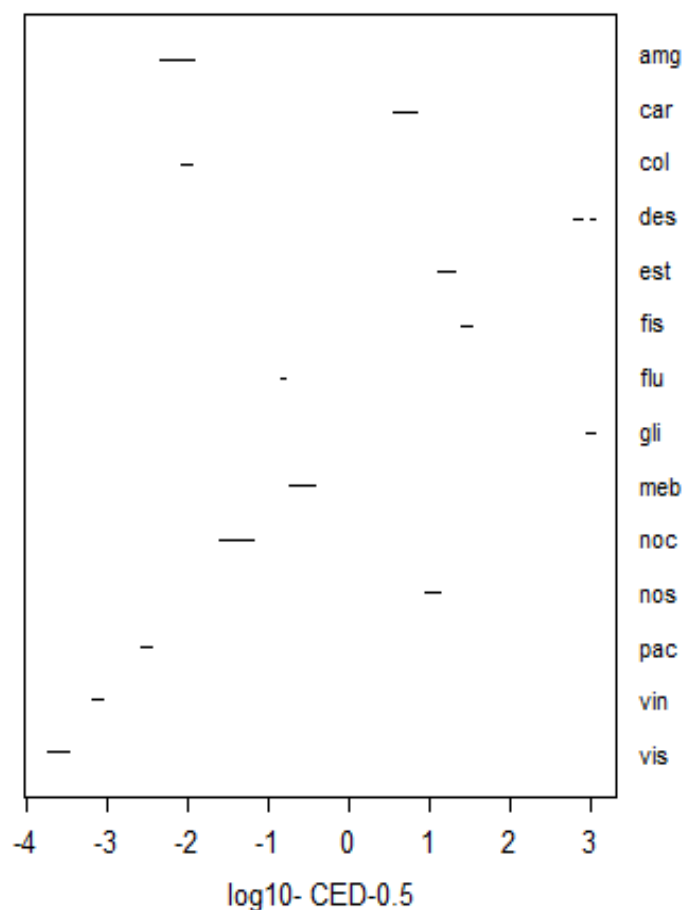
Figure 3.27. BMD confidence intervals plotted for the 'aneugens' covariate subgroup p53 biomarker dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'des' 'gli' whose dose-response confidence interval returned an infinite upper bound evident with the dashed drawn confidence interval in the case of 'des'. The bound confidence intervals spanned a maximum distance of approximately 0.75-Log units. There are several small precise BMD confidence intervals.

3.3.2 MicroFlow Assay Micronucleus Frequency BMD Response Modelling

With the goal of comparing potency metrics obtained from the MultiFlow biomarker responses with the micronucleus endpoint, one needed to obtain similar dose-response potency metrics from the micronucleus datasets. Thus, BMD modelling of the MicroFlow datasets was performed with compound as covariate as described in the materials and methods section of this chapter. The results of the BMD modelling are presented in the next subsections of this chapter.

3.3.2.1 Clastogens

Figure 3.28. BMD modelling curves for the 'clastogena' subgroup analysed with compound as the covariate level for micronucleus percent dose-responses. Double-log plot with concentration in Log10-μM on the *x*-axis and Log-10 micronucleus percent response on the *y*-axis. Coloured points and curves and correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'cis' retuned an infinite CEDU.

Figure 3.29. PROAST generated individual dose-response curves for each level of the 'clastogena' covariate micronucleus biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 micronucleus percent response on the *y*-axis. Fitted dose-response curves for all compounds. The 'cis' compounds exhibited some scatter with a weak dose-response visible.
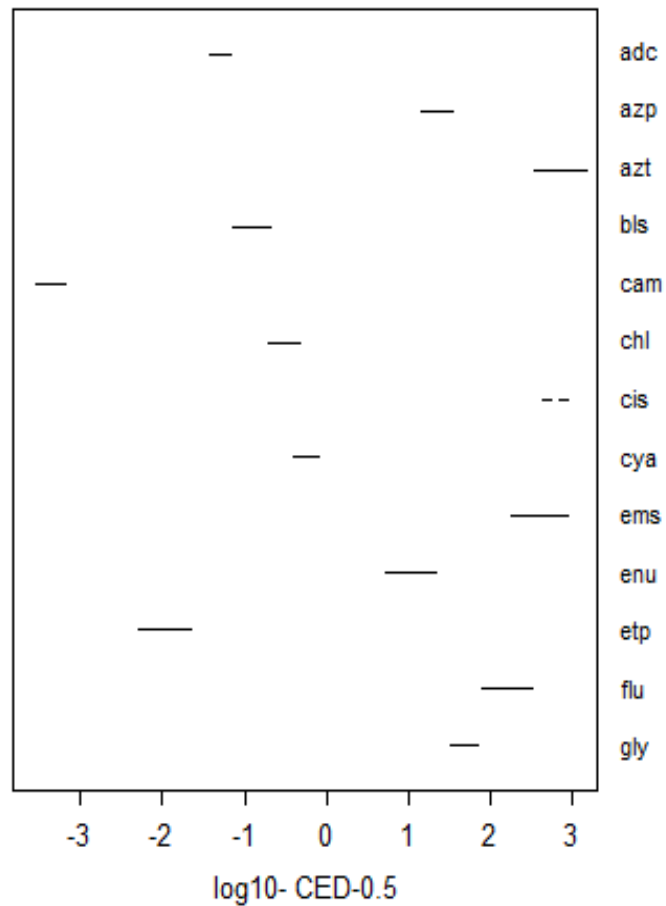
Figure 3.30. BMD confidence intervals plotted for the 'clastogensa' covariate subgroup micronucleus percent response dose-responses. All compounds returned upper and lower bound finite confidence intervals except for 'cis' whose dose-response confidence interval returned an infinite upper bound evident with the dashed line drawn confidence interval. The bound confidence intervals spanned a maximum distance of approximately 1-Log unit. There are several small precise BMD confidence intervals.
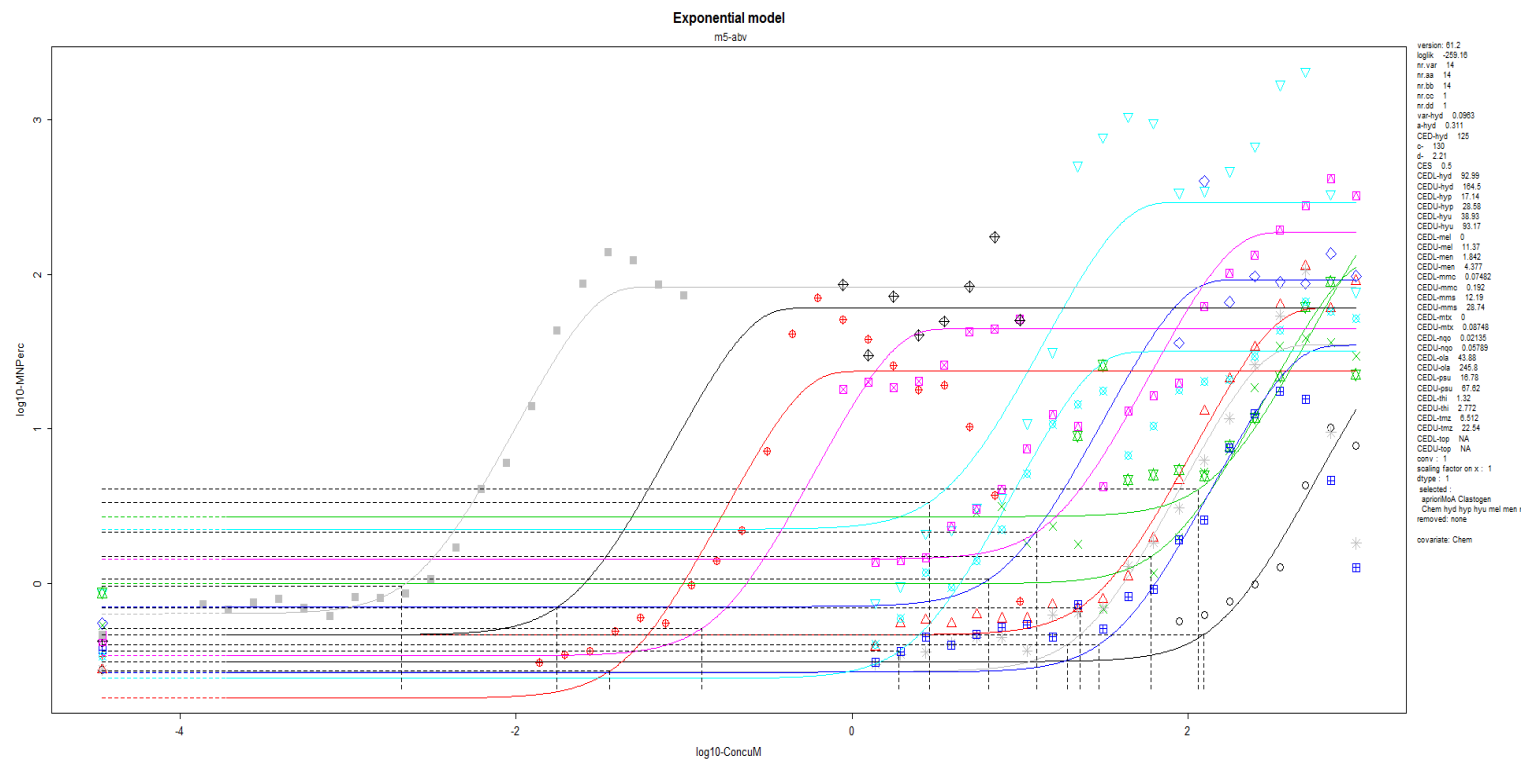
Figure 3.31. BMD modelling curves for the 'clastogenb' subgroup analysed with compound as the covariate level for micronucleus percent dose-responses. Double-log plot with concentration in Log10-µM on the *x*-axis and Log-10 micronucleus percent response on the *y*-axis. Coloured points and curves and correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds. Notably, 'mel' and 'mtx' retuned a CEDL value that was not distinguishable from 0.
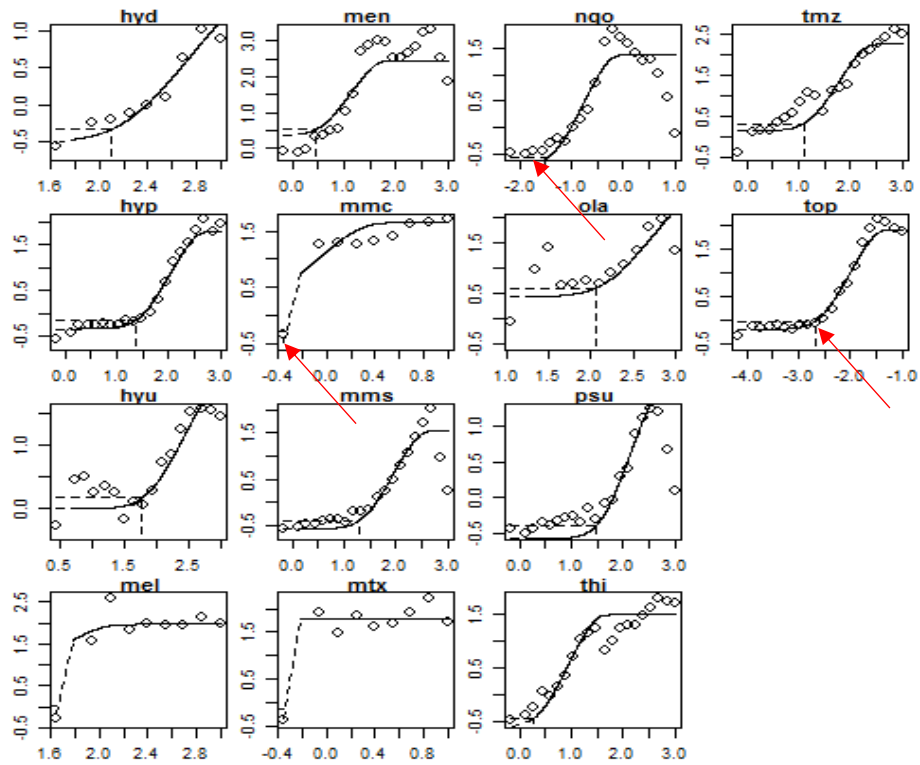
Figure 3.32. PROAST generated individual dose-response curves for each level of the 'clastogenb' covariate micronucleus biomarker dose-responses. Double-log plot with concentration in Log10 µM on the *x*-axis and the Log10 micronucleus percent response on the *y*-axis. Fitted dose-response curves for all compounds. The 'mel' and 'mtx' compounds exhibited show a steep dashed line from intersect which corresponds to the CEDL not being distinguishable from 0 and no evidence of a dose-response in these instances. The 'mmc', 'nqo', and 'top' compounds analysed with CES of 50% result in BMD metrics in the low highly variable regions of the dose response curves, indicated with the red arrows.
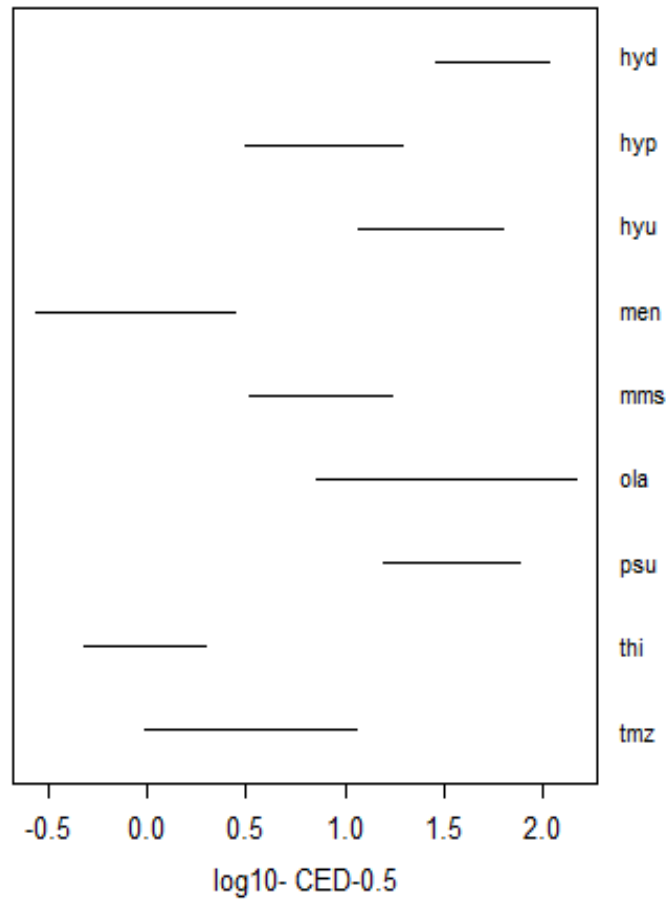
Figure 3.33. BMD confidence intervals plotted for the 'clastogensb' covariate subgroup micronucleus percent response dose-responses. All compounds returned upper and lower bound finite confidence intervals. The bound confidence intervals spanned a maximum distance of approximately 1.5-Log units. Note that 'mel' and 'mtx' are not included since there was no evidence of a dose-response and hence no determinable BMD confidence intervals.
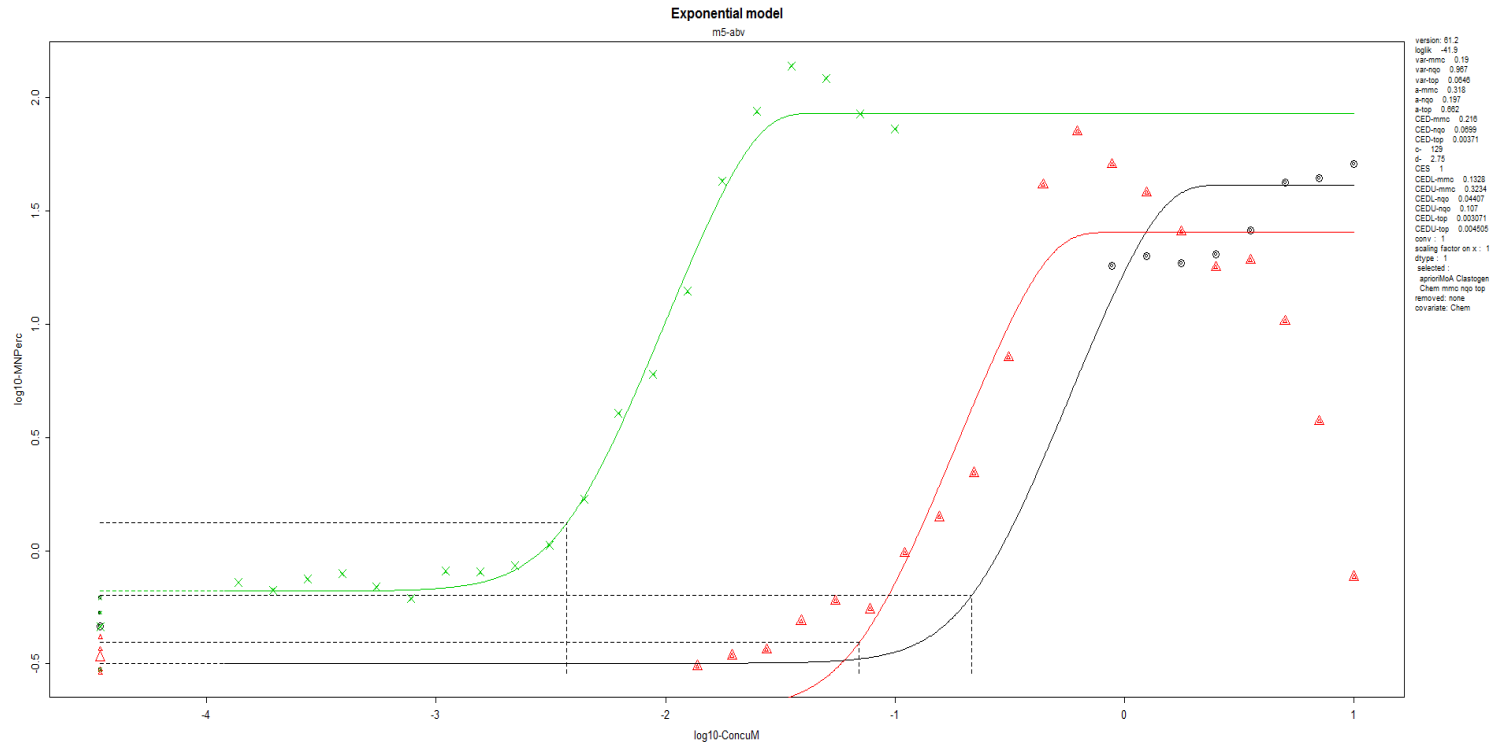
Figure 3.34. BMD modelling curves for the 'mmc', 'nqo', and 'top' concatenated subgroup analysed with compound as the covariate level for micronucleus percent dose-responses. Double-log plot with concentration in Log10-µM on the *x*-axis and Log-10 micronucleus percent response on the *y*-axis. Coloured points and curves and correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES was increased to 1.0 (100%) for this analysis with the aim to improve BMD metric outcomes. CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds with calculable CEDLs and CEDUs.
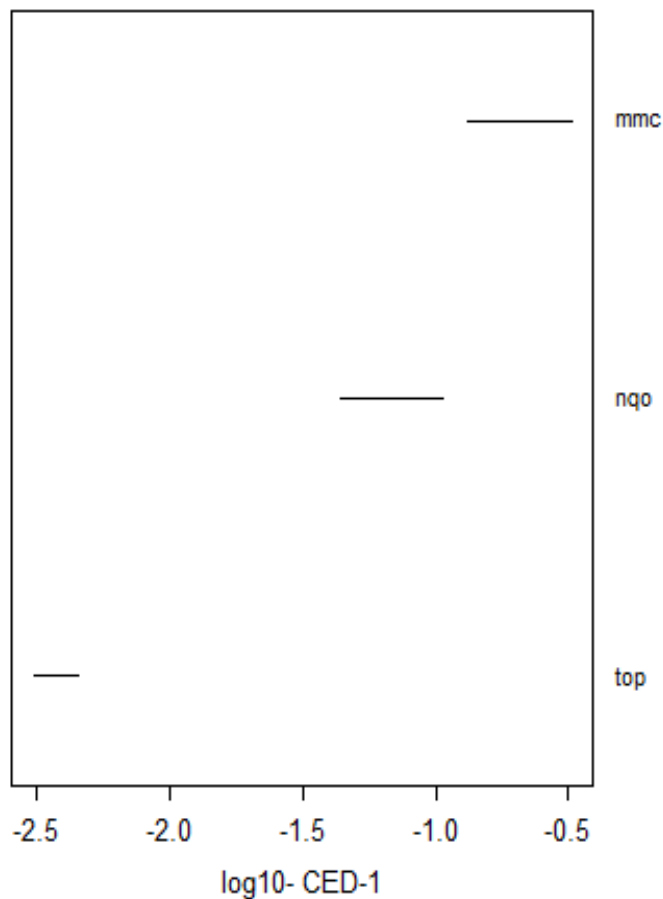
Figure 3.35. BMD confidence intervals plotted for the 'mmc', 'nqo', and 'top' concatenated subgroup micronucleus percent response dose-responses with a CES of 0.1 (100%). All compounds returned upper and lower bound finite confidence intervals. The bound confidence intervals spanned a maximum distance of approximately 0.5-Log units.

3.3.2.2 Aneugens
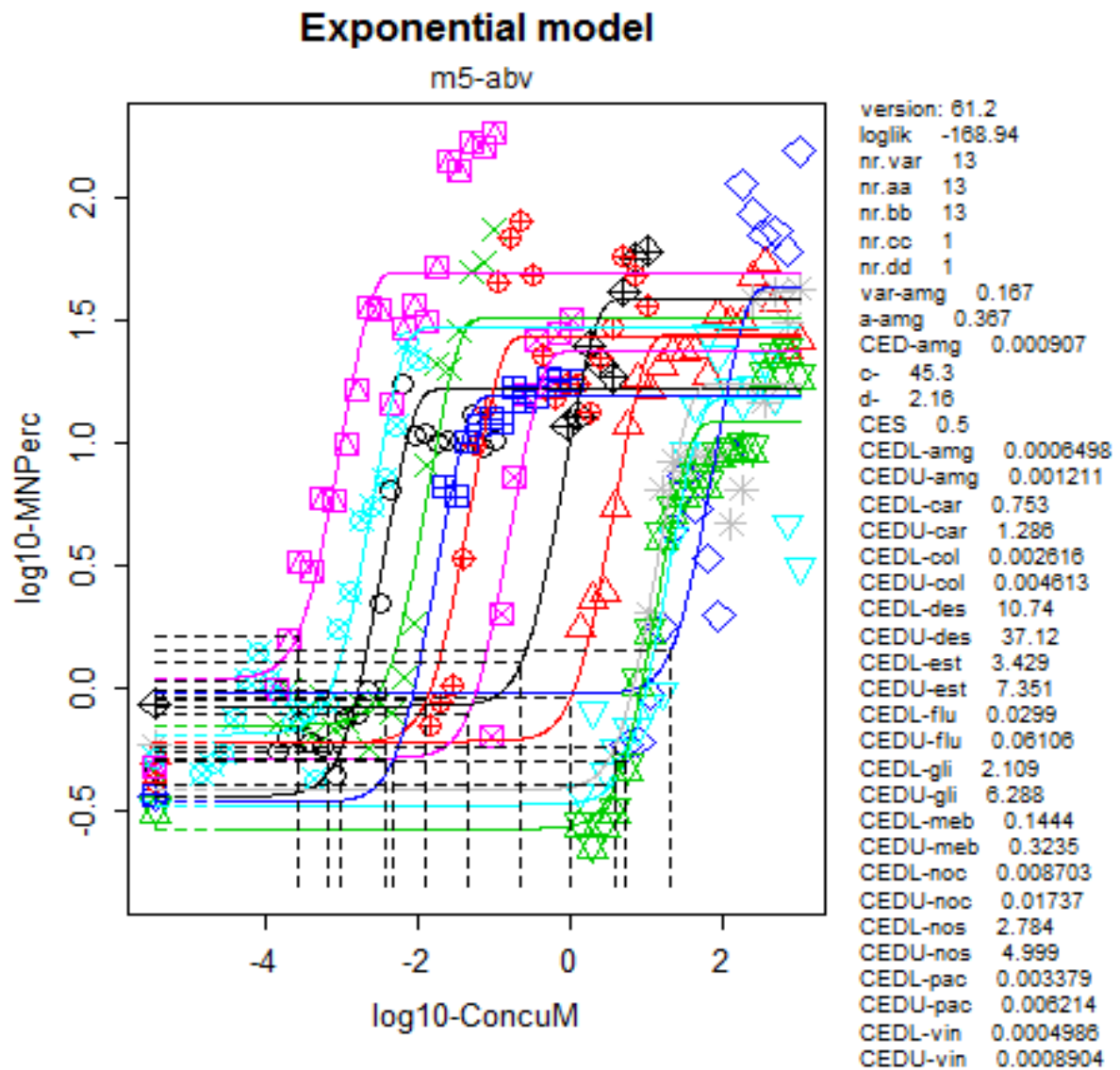
Figure 3.36. BMD modelling curves for the aneugens subgroup analysed with compound as the covariate level for micronucleus percent dose-responses. Double-log plot with concentration in Log10-μM on the *x*-axis and Log-10 micronucleus percent response on the *y*-axis. Coloured points and curves and correspond to the dose-response curve for each respective compound. Summary statistics are presented on the right hand side including the model parameters (*a,c,d*). CES 0.5 (50%). CEDL and CEDU are synonymous with BMDL and BMDU. Exponential Model 5 returned dose-responses for all compounds with calculable CEDLs and CEDUs.
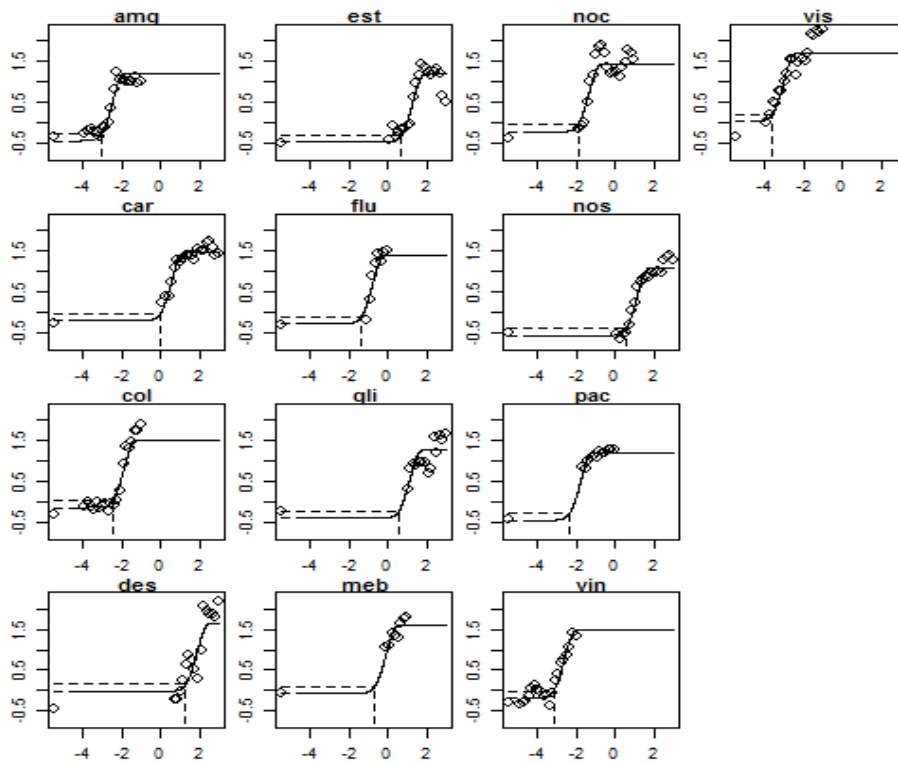
Figure 3.37. PROAST generated individual dose-response curves for each level of the 'aneugen' covariate micronucleus biomarker dose-responses. Double-log plot with concentration in Log10 μM on the *x*-axis and the Log10 micronucleus percent response on the *y*-axis. Fitted dose-response curves for all compounds. Clear dose-response for all compounds.
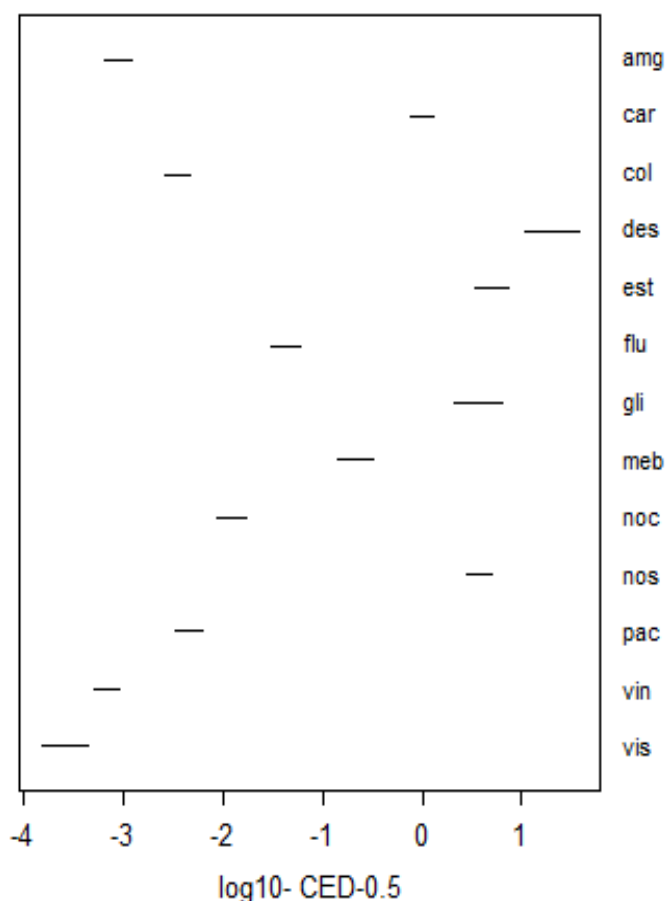
Figure 3.38. BMD confidence intervals plotted for the 'aneugens' covariate subgroup micronucleus percent response dose-responses. All compounds returned upper and lower bound finite confidence intervals. The bound confidence intervals spanned a maximum distance of approximately 1-Log unit.

3.3.3 Compound MultiFlow Biomarker Response Potencies Correlated with *in vitro* Micronucleus Response Potencies

Following the feasibility studies and determination that the MultiFlow biomarker response datasets are amenable to BMD response-modelling, the BMD analysis results presented in the next subsections of this chapter serve as the basis to determine whether MultiFlow biomarker(s) could provide a reliable indication of a chemicals' genotoxic potency as measured by the *in vitro* micronucleus assay.

3.3.3.1 Clastogens
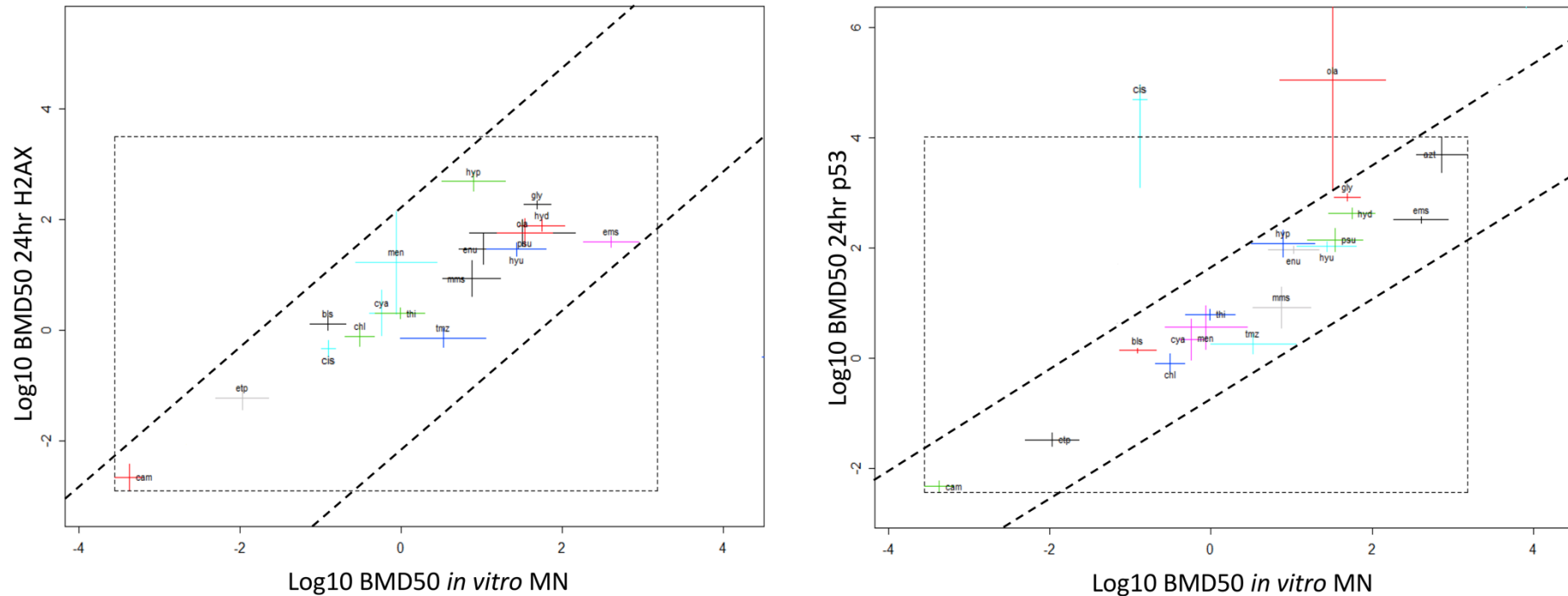
Figure 3.39. Cross-correlation plots for the tested clastogen BMD CIs. The *x* and *y* axes represent Log10 concentration of compounds in µM. CES of 50% for all BMD analyses. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs except outliers. Left Panel: *in vitro* MN versus γH2AX responses. Right Panel: *in vitro* MN versus p53 responses. Compound 'ola' lies outside the general observed trend, with an unbound upper CI in the p53 endpoint. Compound 'cis' displays an unbound upper CI in the p53 endpoint. Abbreviation: BMD = Benchmark Dose, CI = Confidence Interval, MN = micronucleus. Figure from Dertinger et al. (2019).
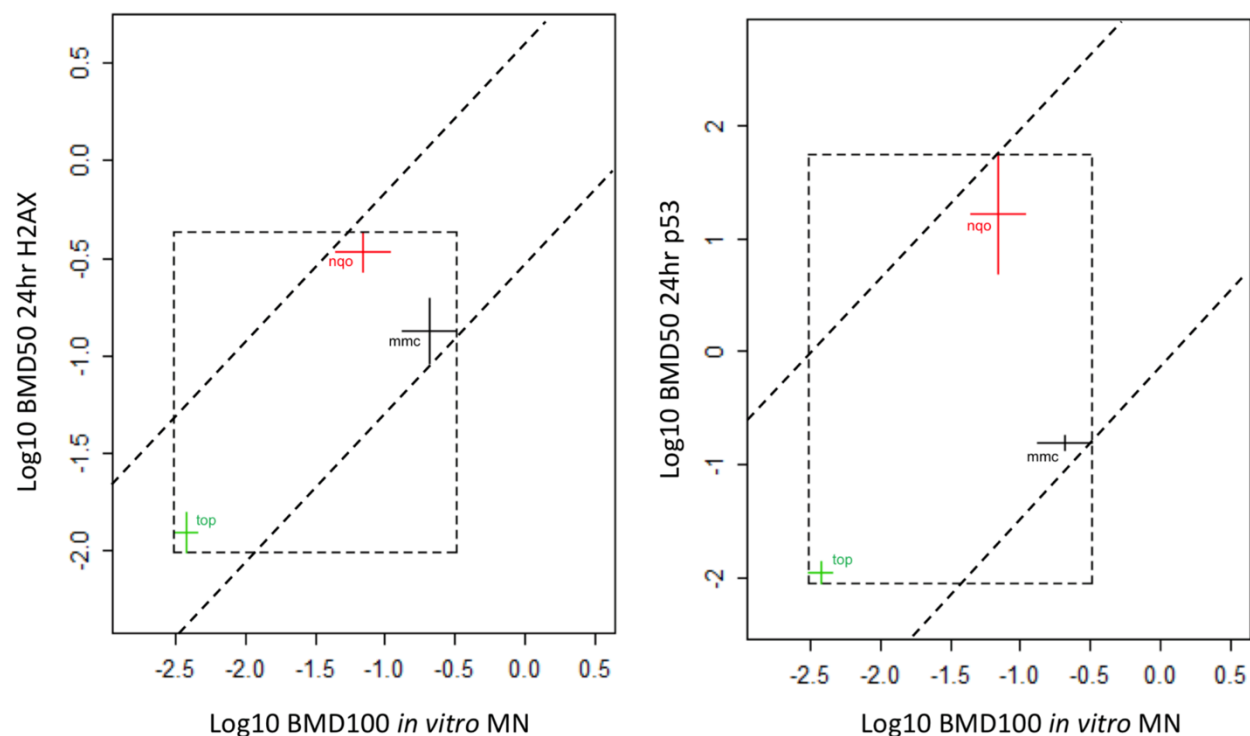
Figure 3.40. Cross-correlation plots for the tested clastogen BMD CIs that were re-analysed with CES 100% in the *in vitro* micronucleus endpoint, but remained constant at 50% in the γH2AX and p53 biomarkers. The *x* and *y* axes represent Log10 concentration of compounds in μM. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs. Left Panel: *in vitro* MN versus γH2AX responses. Right Panel: *in vitro* MN versus p53 responses. BMD = Benchmark Dose, CI = Confidence Interval, MN = micronucleus. Figure from Dertinger et al. (2019).

3.3.3.2 Aneugens
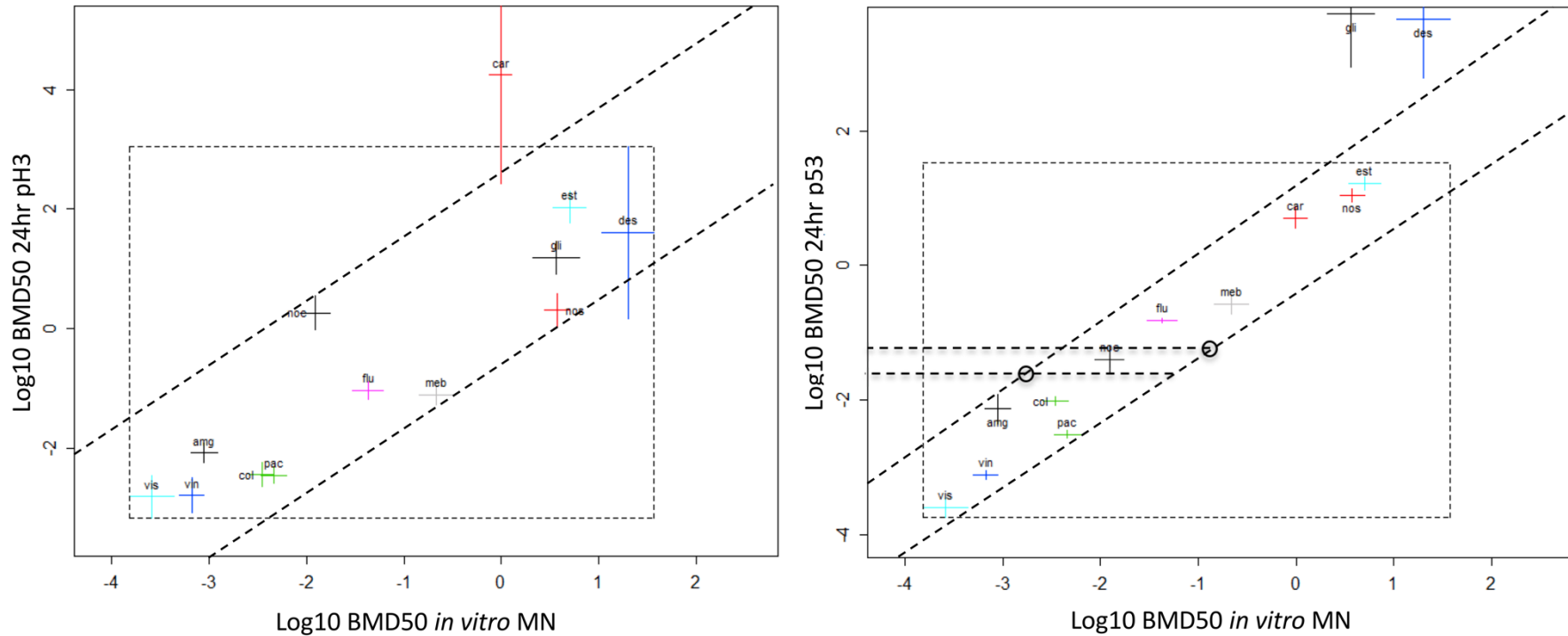
Figure 3.41. Cross-correlation plots for the tested aneugen BMD CIs. The *x* and *y* axes represent Log10 concentration of compounds in μM. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs except outliers. Left Panel: *in vitro* MN versus p-H3 responses. Right Panel: *in vitro* MN versus p53 responses. The compounds 'car', 'gli', and 'des' lie outside the general observed trends. Additional dashed lines with terminating circles were drawn on the right panel to show how the CI intersects from one axis can be used to predict the intersect with another axis with a margin of uncertainty. Abbreviation: BMD = Benchmark Dose, CI = Confidence Interval, MN = micronucleus. Figure from Dertinger et al. (2019).

3.4 Discussion

The aim of this present chapter is to determine the amenability of MultiFlow biomarker dose-response information to BMD analysis, and to ultimately provide genotoxic potency ranking of tested clastogens and aneugens compared with the gold-standard micronucleus assay. To test this, several training set clastogens and aneugens MultiFlow biomarker and micronucleus dose-response datasets were obtained and subjected to BMD dose-response modelling and confidence interval determination and correlation.

3.4.1 Dose-Responses for the MultiFlow Biomarker and MicroFlow Responses

3.4.1.1 General Observed Trends and Outliers

Overall, most tested compounds returned reliable dose-response metrics when assessed by the exponential BMD dose-response Model 5 used in these analyses. It was promising to see fitted dose-response curves with minimum scatter for the majority of compound and biomarker response combinations. Some outliers as already highlighted in the results sections included the following:

- Clastogens
    - γH2AX – 'aza' no dose-response; 'ful' and 'men' different dose-response shape
    - p53 – 'aaf', 'cis', and 'ola' infinite BMDUs; 'nqo' high degree of scatter
    - Micronucleus – 'cis' infinite BMDU; 'mmc', 'nqo', and 'top' CES50% BMDs in the low dose region.
- Aneugens
    - p-H3 – 'car' infinite BMDU; 'des' and 'est' with varying degrees scatter, with des exhibiting the most. The 'des' confidence has the maximum width compared to other confidence intervals.
    - P53 – 'des' and 'gli' infinite BMDUs.

3.4.1.2 Unbound Confidence Intervals

With respect to the instances where an unbound BMDU is observed, this is indicative of the model uncertainty which indicates that there may not be any reliable dose-response information available for these compounds. Wills et al. (2016a) describes that in this respect,

the fact that a lower BMD confidence bound can be calculated highlights the strength of the BMD approach. Applying the same reasoning to the MultiFlow biomarker response results provides conclusive information that a dose-related effect (above the CES of 50%) will most likely occur at concentrations higher than the established BMDL. This information may prove useful to the investigating laboratory that concentration selections could be refined to improve the reliability of the dose-response model.

3.4.1.3 Differences Among the Individual Underlying Dose-Response Shapes

Some dose-responses that exhibited underlying dose-response patterns that fluctuated in direction compared to others were highlighted as anomalies, specifically 'ful' and 'men' in the clastogen γH2AX biomarker dose-response models as re-shown in the following **Figure 3.42**.
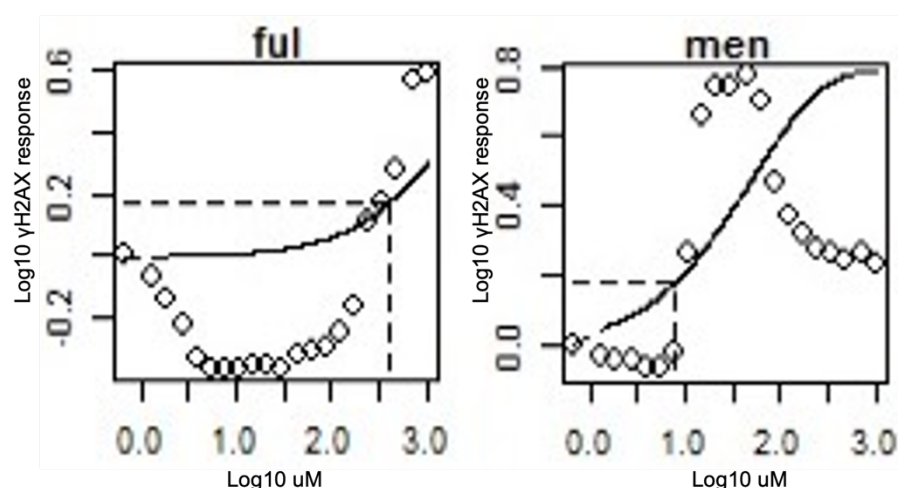
Figure 3.42. PROAST generated individual BMD dose-response curves for 'ful' and 'men' in the γH2AX biomarker dose-response models in which there was observed fluctuation in the dose-response directions.

There are some plausible explanations for said fluctuation. The most likely explanation is due to biological variability within the controls or test compounds fluorescence interference which could have been resolved with repeat benchtop experiments. Another plausible explanation is the phenomenon of biological hormesis, or J-shaped dose-response curves depicting low dose stimulation and high dose inhibition. Hormesis is manifested in toxicological endpoints such as glutathione-s-transferase (GST)-positive foci in hepatic carcinogenesis (Fukushima et al., 2005), and as an adaptive response to chemical and

physical mutagens (Calabrese and Blain, 2005). In a biological context, these low-dose effects can be attributed to overcompensation which can restore the homeostatic balance (Calabrese and Baldwin, 2001). Regardless of the cause of the observations in the 'ful' and 'men' γH2AX biomarker dose-responses, this is an interesting observation but a further in-depth discussion of biological hormesis is out of scope of the aims of this chapter, and there is insufficient data to warrant further investigation into the effects observed in these results. Specific to the aims of this results chapter, the BMD model still obtained BMD confidence intervals for these two compounds ('ful' and 'men'), however the confidence interval widths spanned the maximum distances of 1.5- and 2.5-Log units, respectively, which is double the distance observed from the other confidence intervals. The presence of wide confidence intervals in these examples directly relates to the high level of uncertainty in the underlying datasets.

The generation of PROAST individual dose-response curves for each level of the covariate (compound) for each dose response model followed by individual scrutiny of the dose-responses remains crucial in the interpretation of the BMD model outputs. Whilst it is a labour intensive and iterative process, some may advocate for the development of automated tools or macros to replicate the BMD model process. One cannot fully support this notion, as can be highlighted by taking a closer look at the 'nqo' and 'ola' compound dose-response curves for the p53 biomarker (reshown in **Figure 3.43** below). The general trend of both compounds is an increase in response as concentration increases until a decrease in response from the 7th from last concentration in 'nqo' and 8th from last concentration in 'ola'. When referring to the original raw data, the same concentrations correspond to high levels of treatment-related cytotoxicity (>70%). It is therefore imperative to include expert human judgment in the dose-response analysis interpretation to account for events such as this, and one contends that a fully automated model would miss these nuances. Whilst recalculation of the BMD model was not performed in this proof-of-concept study, an investigating scientist may wish to remove the uppermost concentrations that correspond to treatment-related cytotoxicity to improve the reliability of the BMD model. This would be especially important when using the data to define reference values for determining safety levels such as acceptable intakes, which is out of scope of this thesis and *in vitro* genotoxicity data in general.
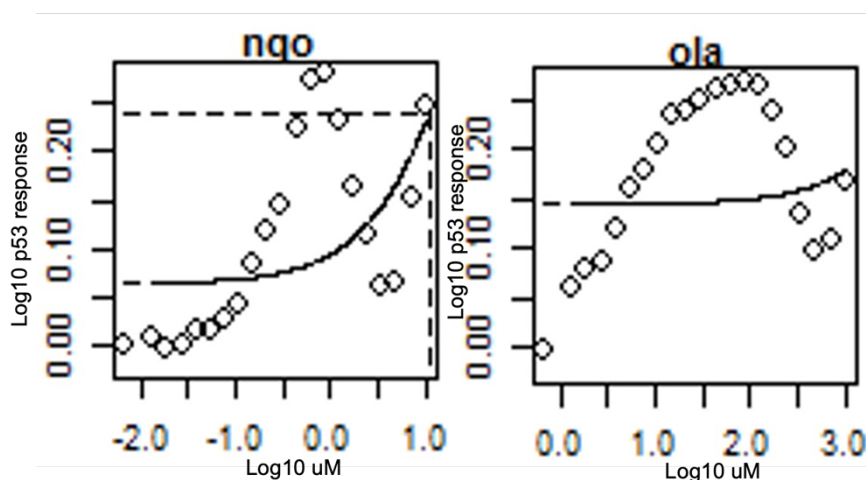
Figure 3.43. PROAST generated individual BMD dose-response curves for 'nqo' and 'ola' in the p53 biomarker dose-response models in which there were observed top dose treatment-related cytotoxicity effects.

3.4.1.4 Confidence Interval Precision

On the topic of the comparison of confidence interval widths, it was promising to see confidence intervals generally spanning in the 1 order of magnitude (1-Log) range. As originally theorised by Slob and Setzer (2014) and later experimentally demonstrated by Wills et al. (2016a), the ratio between the BMDU and BMDL relates directly to the precision of the BMD estimate, with smaller BMDU to BMDL ratios being the more precise estimate. There were many instances in the BMD dose-response outputs where very small confidence intervals (<0.5Log) existed in comparison to some other confidence intervals, which accounts for the high degree of precision in the BMD model of the compound dose-responses in those instances. Some examples include the aneugen p53 dose-responses. Contrastingly, some very wide confidence intervals were observed, for example the aneugen 'des' p-H3 BMD dose-response confidence interval with a confidence interval of 3 orders of magnitude, which coincides with the high degree of scatter observed in the individual dose-response curve for the same compound.

3.4.1.5 Critical Effect Size Determination

A CES of 0.5 (50%) appeared to be appropriate for most of the compound and biomarker response combinations analysed in this study. An in-depth assessment on the correct CES to select for MultiFlow biomarker dose-response modelling was not a hypothesis in scope of this results chapter; however, suffice to say that good quality dose-response metrics were

obtained for most analyses with the goal of determining potency ranking metrics for the tested compounds in the MultiFlow and MicroFlow endpoints. The findings are consistent with the published literature where a CES 50% appears to be the preferred metric for the *in vitro* micronucleus endpoint (Bemis et al., 2016, Zeller et al., 2016). However, it was deemed necessary to re-analyse the clastogens ('mmc', 'nqo', and 'top') in the micronucleus endpoint with a CES of 100%. As indicated in **Figure 3.32**, a BMD analysis with a CES of 50% yielded BMD estimates that reside in the lower doses of the dose-response curve which are associated with high levels of uncertainty. Endpoint specific critical effect sizes have been reviewed in the literature (Zeller et al., 2016; Slob, 2017), with some advocates for determining an appropriate CES based on the background and maximum response levels. Zeller et al. (2016) highlights the level of uncertainty in the low dose region of the curve. Other authors show that the choice of CES is unimportant when the scope and aim of BMD modelling is to determine compound response confidence intervals for potency ranking (Bemis et al., 2016). To this end, a CES of 100% was implemented for the dose-response analysis of these compounds which yielded bound confidence intervals suitable for comparative BMD potency analyses.

3.4.1.6 System Cross-Correlation

The BMD potency metrics served as a basis for determining whether MultiFlow biomarker(s) could provide a reliable indication of chemicals' genotoxic potency as measured by the *in vitro* micronucleus assay. To this end, the BMD confidence interval cross-correlation plots were formulated as presented in section 3.3.3.

As shown in **Figure 3.39**, **Figure 3.40**, and **Figure 3.41**, the BMDs in the MultiFlow biomarker responses were plotted against the micronucleus response BMDs on a double-log scale. As proposed by the developers of the cross-system correlation plots (Soeteman-Hernández et al., 2015; 2016), in opposition to representing correlation with a numerical coefficient value, a simpler approach is to show a linear relationship with intercept zero equalling a straight line in a double-log plot. Therefore, two lines with unity slope were drawn on each correlation plot in such a manner that most of the BMD confidence intervals were encompassed between the lines. The distribution of BMD positions within the two lines shows approximate linearity, differing by a proportionality constant. Furthermore, the vertical distance between the two lines translates into an uncertainty margin given by the estimation

of a BMD on the y axis based on a specified BMD on the x axis, and vice versa. The uncertainty margin is used as a measure of correlation between two endpoints.

For the aneugens, when comparing micronucleus induction to p53 responses, the cross-system plots show good correlation, with most of the compounds located between the two lines (**Figure 3.41**). Taking the microtubule binder nocodazole ('noc') as an example, the horizontal dashed line intersections with the sloped dashed lines may be considered as the respective upper and lower bounds of the uncertainty range for the *in vitro* MN endpoint. The intercepts of approximately -3 and -1 on the Log scale correspond to lower and upper bounds of $10^{-3} = 0.001$ and $10^{-1} = 0.1$ μM, respectively. Hence, the *in vitro* micronucleus BMD for nocodazole is estimated to lie between 0.001 and 0.1 μM considering an uncertainty margin of approximately 1-Log. In fact, the *in vitro* micronucleus potency for nocodazole in the dataset represented in **Figure 3.41** has both the BMDL and BMDU either side of -2-Log, and hence within the estimated potency of -3- and -1-Log estimated from the p53 response. The micronucleus vs. 24 hr p-H3 system plot also indicates the BMDs for most compounds in both systems are proportionally related (**Figure 3.41**), however the two lines are drawn further apart than the micronucleus vs. 24 hr p53 system (i.e., 2-Log vs. 1-Log). In both cases, micronucleus vs. p53 and micronucleus vs. p-H3, the data are randomly scattered with good correlation.

For the clastogens, good correlation is observed for micronucleus vs. γH2AX and micronucleus vs. p53, with BMDs randomly scattered between the two diagonal lines of the unity slopes, with distances of approximately 3- and 2-Log, respectively, for each system (**Figure 3.39**). The *in vitro* micronucleus BMD100 CI for compounds 'mmc', 'nqo', and 'top', plotted against BMD50 γH2AX and p53 endpoints show similar correlation with BMD CIs lying within approximately 2-Log for both systems (**Figure 3.40**).

The correlations observed here are consistent with those of other genotoxicity endpoints which have been compared using similar methodologies. Bemis et al. (2016) compared *in vitro* micronucleus responses against *in vivo* micronucleus responses for a group of seven clastogens. Similarly, Soeteman-Hernández et al. (2015) assessed the ability to predict *in vivo* micronucleus potency from *in vitro* micronucleus data. BMD confidence intervals spanned two orders of magnitude, with *in vivo* BMD confidence intervals generally showing smaller than those from *in vitro* studies.

3.5 Concluding Remarks: Quantitative Dose-Response Modelling of MultiFlow Biomarker Responses

The applicability of the BMD approach to the MultiFlow biomarker responses has been demonstrated in this results chapter. In fact, during the time of these research efforts, this was the first instance of applying BMD quantitative dose-response modelling to the MultiFlow biomarker responses.

The strong correlation of MultiFlow biomarkers to *bona fide* genomic damage in the form of micronuclei demonstrates the relevance of the MultiFlow biomarkers to detect genetic damage from cellular exposure to small molecules. Furthermore, the correlations suggest that potency determinations based on MultiFlow endpoints, at least on a rank-order basis, are likely comparable to those derived from the micronucleus assay. This bolsters the use case whereby the constellation of MultiFlow assay biomarkers serves as a reliable genotoxicity screening tool that is predictive of *in vitro* micronuclei formation, with the benefit of providing more mechanistic information. Finally, dose-response analyses of novel biomarkers such as those in the MultiFlow assay are worth pursuing further because they reflect the paradigm shift that has been transitioning genotoxicity away from a simple binary yes/no characteristic to a quantitative metric that has the potential to better inform risk assessments as margin of exposure and other toxicological principles can be considered (Pottenger and Gollapudi, 2010; Johnson et al., 2014b; Macgregor et al., 2015; Dearfield et al., 2017).

In the context of a practical industrial use case, the MultiFlow biomarker to *in vitro* micronucleus correlations presented in this chapter could have enormous value in a high-throughput screening laboratory. For example, a pharmaceutical company that has a program to develop a blockbuster drug with upwards of 1,000 candidate compounds to screen would have to determine genotoxic risk from exposure to the selected compound(s). The company would be bound to performing regulatory genotoxicity studies to meet the requirements of the ICH S2 (ICH, 2008) guideline which would include a bacterial reverse mutation assay, and an *in vitro* mammalian test for chromosome damage which could optionally be an *in vitro* micronucleus assay. With the aim of reducing the high risk of failure and reducing the time and cost investment of performing the *in vitro* micronucleus assay on high-potency compounds that might never reach the clinic, a laboratory could perform MultiFlow

screening of all candidate compounds and only select low-potency low-risk compounds to subject to GLP assay testing for regulatory submission purposes.

Additionally, the findings from this chapter led to further research questions on the broad topic of "what other uses do MultiFlow biomarker BMD metrics serve beyond quantitative compound potency ranking?". There was the strong belief that the MultiFlow assay could be used beyond the hazard-identification paradigm that it was designed to serve by broad mode of action classification and engaged discussions with the researchers at Litron influenced their support for further collaboration to explore BMD analysis in more use cases. Extra collaboration was pursued on a case-study level basis including evaluating the statistical robustness of BMD confidence interval comparison in a S9 dependent manner presented in the next chapter (Chapter 4), and by assessing the genotoxic response of the Topoisomerase II specific class of compounds presented in Chapter 5. Lessons learned from the data analyses presented in this results chapter were applied when testing future hypotheses. Extra care was taken during interpretation of dose-responses relative to hormesis and/or toxic doses responses at high concentrations.

Chapter 4. Comparative Potency Analysis. A Critical Reanalysis of Quantitative *in vitro* Genotoxicity Potency Comparison Data.

4.1 Introduction

As demonstrated in the previous results chapter, it was tremendously encouraging to be able to view the Litron MultiFlow biomarker response data in a quantitative potency sense as opposed to the assay developer's originally intended data analysis fold-increase metrics that dichotomises genotoxic hazard. The use of BMD confidence intervals permitted the genotoxic potency ranking of the MultiFlow biomarker response datasets for a set of training set clastogens and aneugens. The chapter also highlighted some caveats of the use of the BMD approach, including the accurate interpretation of the BMD metrics such as the BMDL and BMDU values. This results chapter includes an in-depth review of the accurate use of BMD confidence intervals associated with the BMDL/BMDU ratio when using the metrics for performing comparative potency analyses across subgroups included in the analysis[5]. This discussion is warranted since it became apparent during this research project that some disagreement exists on which BMD metrics should be used to perform comparative potency analysis. This results chapter highlights a published case example and presents a re-analysis of the same dataset to inform a more statistically robust approach to performing comparative BMD potency analyses. The more statistically robust method includes the development and use of a simple novel algorithm for comparing potencies between conditions using the entire confidence interval.

4.1.1 BMD Metrics

As briefly described in the introduction chapter of this thesis, the key concept of the BMD approach includes the estimation of the BMD value based on a biologically relevant response effect size using computational algorithms that permit the fitting of a response curve using a family of nested exponential (Slob, 2002) or hill models (EFSA, 2017). Inevitable

---

[5] The data presented in this results chapter was published in "The Use of Benchmark Dose Uncertainty Measurements for Robust Comparative Potency Analyses" (Wheeldon et al., 2021) and some passages in this chapter have been quoted verbatim from the same article. A copy of the research article is available in Appendix 2. The same research article won best paper published in the Environmental and Molecular Mutagenesis journal by a student researcher for the year 2021.

existing uncertainty in experimental datasets can be considered by calculating a confidence interval for the BMD. The range delineated between the upper (BMDU) and lower (BMDL) confidence bounds is associated with the BMD confidence uncertainty measurements and can be interpreted as the precision of the estimated BMD. **Figure 4.1** illustrates the BMD confidence interval concept.
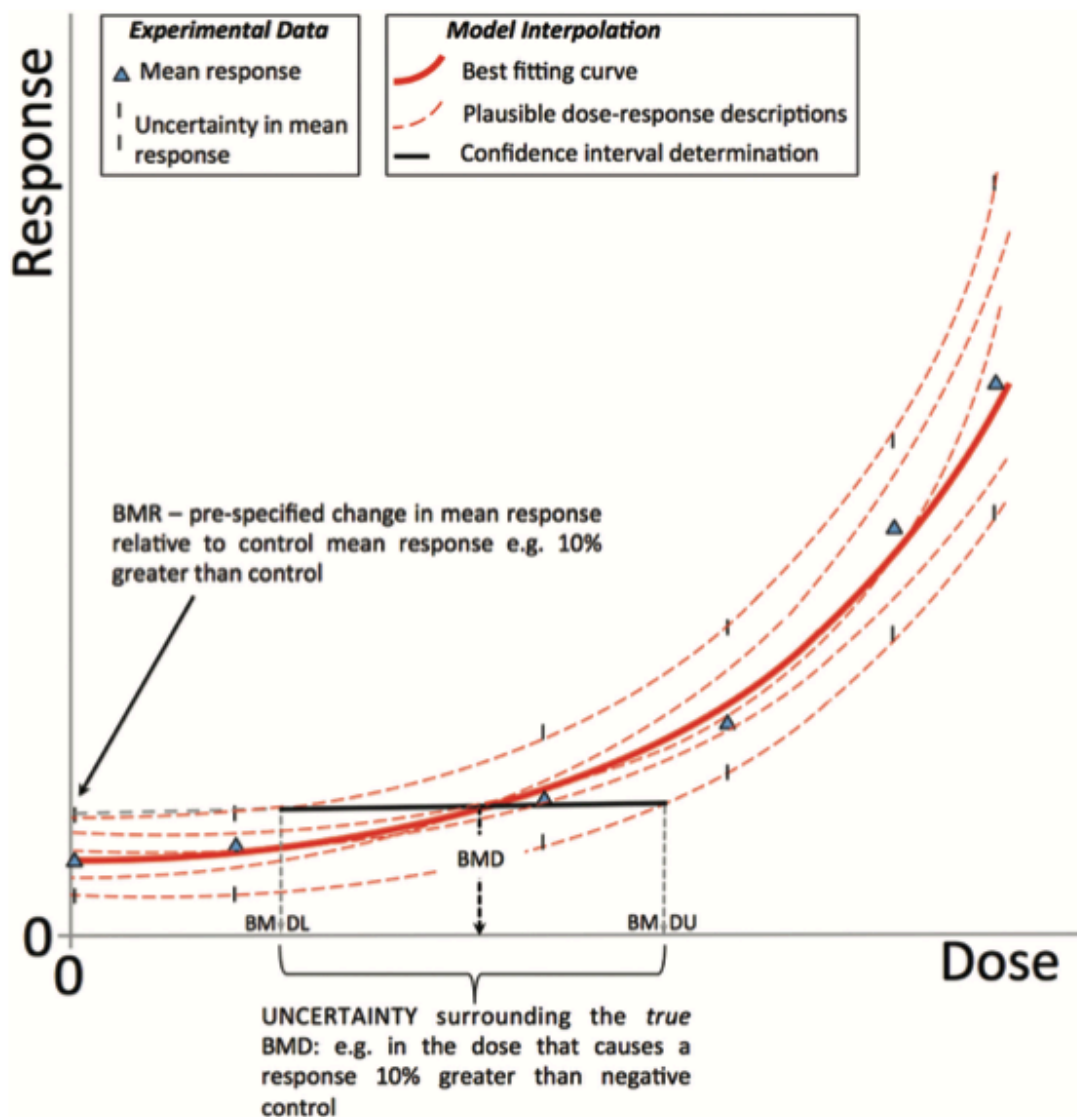
Figure 4.1. An illustrative example of the key concepts involved in the BMD analysis of dose-response data. Based on the predefined benchmark response (BMR, or Critical Effect Size CES in PROAST language), the BMD is estimated by the interpolation from the fitted dose-response curve. The calculated confidence intervals for the BMD consider the uncertainty in the underlying data. The dashed red curves are based on the parameters in the BMD model, and together they make up the confidence interval for the BMD. As an illustrative example, the black horizontal line that intersects the range of red dashed curves

results in the BMDL and BMDU, the lower and upper 90% confidence bounds of the BMD estimate, respectively, with the width of the confidence interval representing the precision of the measurements. Figure adapted from Wills et al. (2016a).

The definition of the BMD as the 'potency' estimate, and the BMDL to BMDU ratio as the likely potency range as defined by Wills et al. (2016a) shall be considered the central argument in this chapter.

4.1.2 Why use a BMD Confidence Interval?

As previously stated, BMD confidence intervals convey the underlying uncertainty in the dataset. The lower bound of the BMD 90% (or 95%) confidence interval (BMDL) derived from *in-vivo* toxicity studies has implications in reference dose-setting for human safety-based guidance values (EFSA, 2017; White et al., 2020). As a conservative approach, the BMDL is used as the dose-response curve PoD. Contrastingly, *in vitro* data remains restricted to potency ranking and mode of action determinations due to the limited toxicokinetic influence from the *in vitro* methods (Wills et al., 2016a). In scope of this thesis, and *in vitro* data in general, the use of BMD confidence intervals is equally as important as regarded in human exposure safety assessment. Studies have demonstrated that the precision of the BMD confidence interval directly influences the interpretation of the result, with large BMDU/BMDL ratios being generally unusable (Slob and Setzer, 2014).

In the vast number of cases, bound confidence intervals yield reliable estimates of the BMD, however unbound upper confidence limits indicate that there is no reliable dose-response upon which one can infer conclusions (Slob and Setzer, 2014; Wills et al., 2016a). One should therefore consider it naturally imperative to account for the precision of the BMD confidence interval when drawing inferences from BMD dose-response datasets. This is especially important when applying the 'combined covariate' BMD modelling approach with a shared covariate within or across and endpoint, since the outcome of the analysis is usually used to compare potencies across the covariate levels that were analysed. This can be illustrated by referring to the BMD confidence intervals obtained by Wills et al. (2016a) from the BMD analysis of select aneugens micronucleus frequency data that was already presented in the introduction chapter (**Figure 1.10**). 3 compounds yielded unbound BMDUs and hence it would be inappropriate to determine the potency of said compounds relative to the other compounds with bound BMD confidence intervals plotted in the same graph.

Moreover, Wills et al. (2016a) outlined a crucial point that due to uncertainty in the underlying dataset, the BMD confidence interval represents the range in which the true BMD point estimate lies. Wills et al. (2016a) continues to explain how alleged potency differences are only statistically defendable when the confidence intervals do not overlap, and hence this bolsters the argument that looking solely at BMD values would be a poor practice when making inferences from BMD modelling results.

4.1.3 BMD Confidence Intervals in Comparative Potency Analyses

The BMD 'combined covariate' approach has been used in several research publications to compare potencies among subgroups since being introduced by Slob and Setzer (2014). Plotting confidence intervals ordered by the geometric mean (midpoint) is a visually intuitive way to compare BMDs and assess differences in potency across the subgroups of the analysis (Wills et al., 2016a). Most published studies that draw conclusions on potency differences do so by visually comparing the position of the confidence intervals in comparison with another. For example, Allemang et al. (2018) performed BMD analyses to evaluate the relative genotoxic potency of 15 pyrrolizidine alkaloids via *in vitro* micronuclei formation in HepaRG cells. The authors were able to show 4 distinct separate potency groupings of the pyrrolizidine alkaloid chemicals tested (**Figure 4.2**) and was a valuable addition to the scientific literature demonstrating quantitative genotoxic potency interpretation of the pyrrolizidine alkaloids class of compounds.
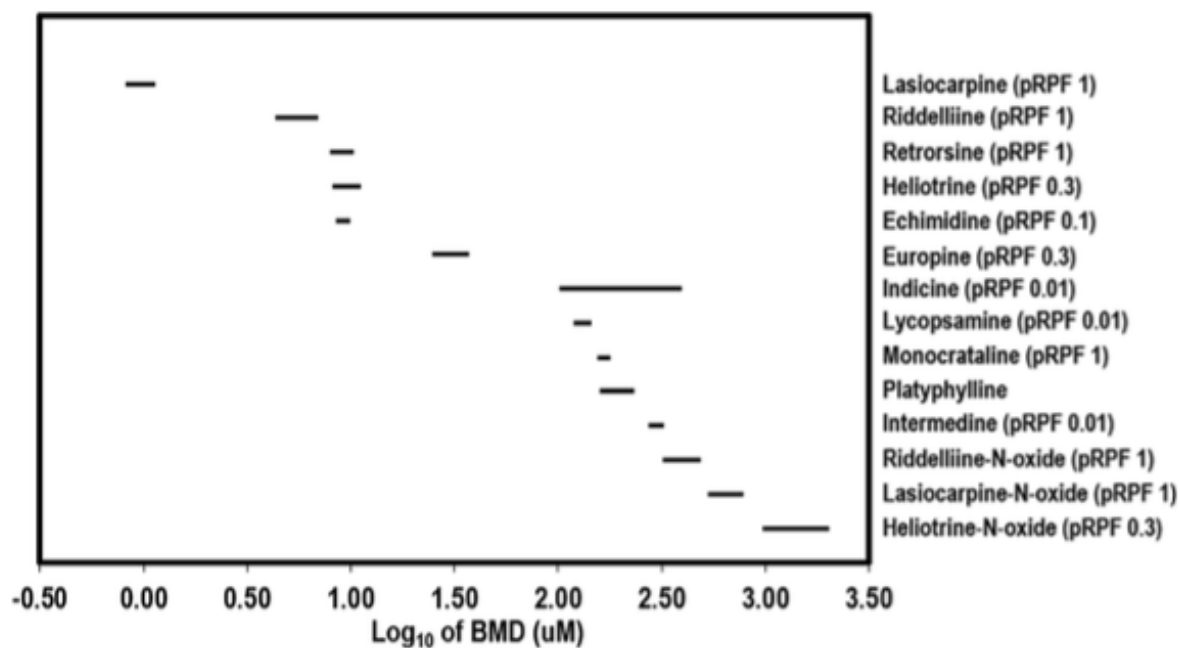
Figure 4.2. Plotted confidence intervals for the BMD analysis of 15 pyrrolizidine alkaloids studied in the *in vitro* micronucleus assay. 4 distinct potency groupings are visible being lasiocarpine as the most potent, a subsequent lower potent group of compounds (riddelliine to echimidine), subsequently europine, and the final 8 compounds forming the lowest potency group. Figure adapted from Allemang et al. (2018).

BMD comparative potency analysis extends further than testing the potency comparison of compounds within the same endpoint. The comparison of potency of a compound between different test systems can also be an interesting hypothesis, for example the comparative potency analysis of how a genotoxic compound performs in different *in vitro* cell lines or under different *in vivo* treatment duration regimens. Seo et al. (2019) were interested in how DNA damage is expressed between two different metabolically competent HepG2 and HepaRG *in vitro* liver cell lines. The BMD confidence intervals for 4 carcinogens were plotted in the Log scale (**Figure 4.3**) and the resulting plots were used to compare the resulting genotoxic potency compared between the two cell lines for the 4 compounds that were tested. Comparative potency analysis based on the BMD confidence intervals plots revealed that generally both the HepaRG and HepG2 cell lines are equally sensitive at detecting genotoxic responses for the tested compounds. This was inferred since the BMD confidence intervals largely overlap.
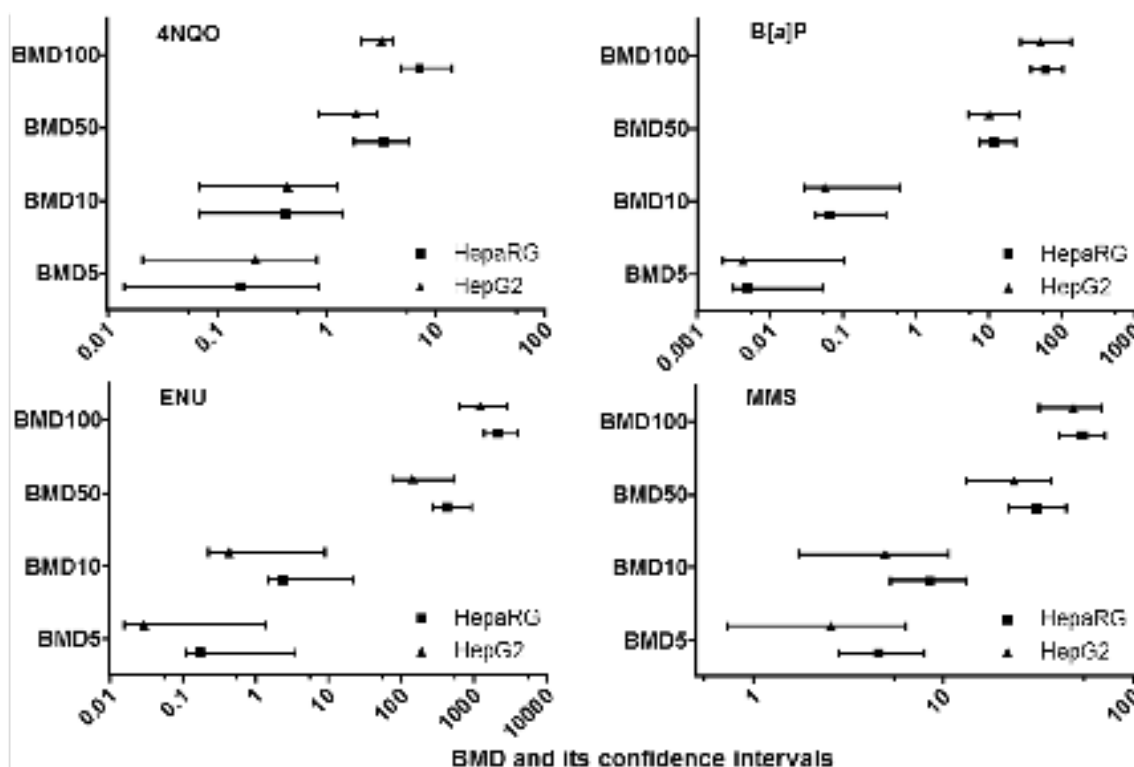
Figure 4.3. BMD confidence interval plots of the 4 genotoxic carcinogens 4NQO (4-Nitroquinoline 1-oxide), B[a]P (Benzo[a]Pyrene), ENU (N-ethyl-N-nitrosourea), and MMS (methyl methanesulfonate) exposed to two different *in vitro* liver cell lines (HepaRG and HepG2) and evaluated for DNA damage *via* the comet assay. Plot shows how the measured genotoxic BMD potency measurements are comparable with overlapping BMD confidence intervals in the different *in vitro* liver cell lines. Interestingly, the authors chose to identify BMD point estimates with triangular or square points on the confidence interval plots. Figure adapted from Seo et al. (2019).

Other authors used the BMD confidence interval plots from the combined covariate approach to assess reproducibility from an interlaboratory validation study of the *in vivo* Pig-a assay (Johnson et al., 2016). Quantitative dose-response data from 13 different Japanese was procured where rats were exposed to N-nitroso-N-ethylurea (ENU) under the same conditions. Red blood cell (RBC) and reticulocyte (RET) mutant frequency data was subjected to BMD dose-response modelling. In this example, laboratory was selected as the covariate in the BMD model. BMD confidence intervals were plotted on a Log scale and scrutinized for overlap or non-overlap. The confidence interval plot is shown in **Figure 4.4** and the authors concluded that the overlap of confidence intervals can be interpreted as "equipotency" which demonstrates a reasonable level of interlaboratory reproducibility.
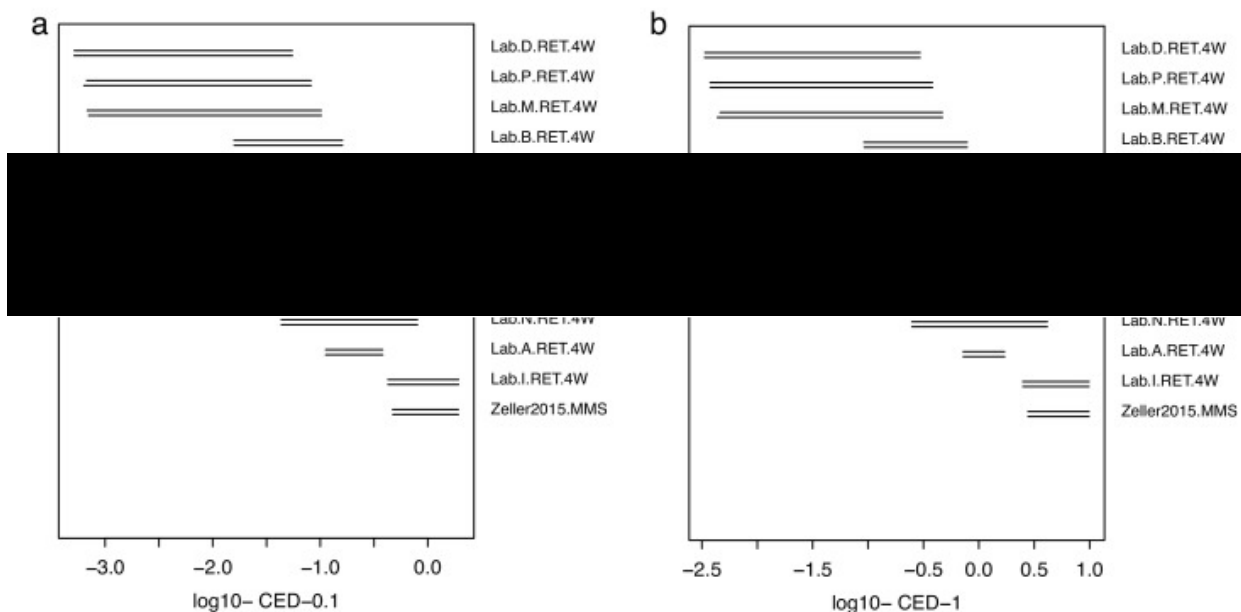
Figure 4.4. BMD confidence intervals of dose-responses from rats exposed to ENU and sampled at the week-4 period and analysed for RET mutant frequency. The laboratory was selected as the covariate in the BMD analysis. The lab code is shown on the right-hand side of each plot corresponding to the confidence interval. Exponential and Hill models were used in the analysis and plotted concomitantly. Overlapping confidence intervals were interpreted as equipotent demonstrating laboratory reproducibility. Figure taken directly from Johnson et al. (2016).

Another important question is which organ and/or tissue is the most sensitive to analyse to assess the effects of a genotoxic compound following exposure in an *in vivo* model. Specifically, this question must be answered when evaluating tissues from a 2-year or transgenic rodent carcinogenicity study for regulatory submission purposes in compliance with the ICH S1B guideline (ICH, 1998). To this end, Wills et al. (2016b) performed combined covariate BMD analysis with a tissue-specific covariate from 28-day repeat dose benzo[a]pyrene oral exposure *in vivo* MutaMouse studies. BMD confidence intervals for MutaMouse *LacZ* mutant frequencies for each tissue-specific level of the covariate were plotted on a Log scale in order of decreasing potency (**Figure 4.5**). The order of potency can be interpreted as the tissue-specific sensitivity to exposure from the genotoxic benzo[a]pyrene. The rank order from greatest to lowest for tissue-sensitivity was small intestine > bone marrow > glandular stomach or lung > liver.
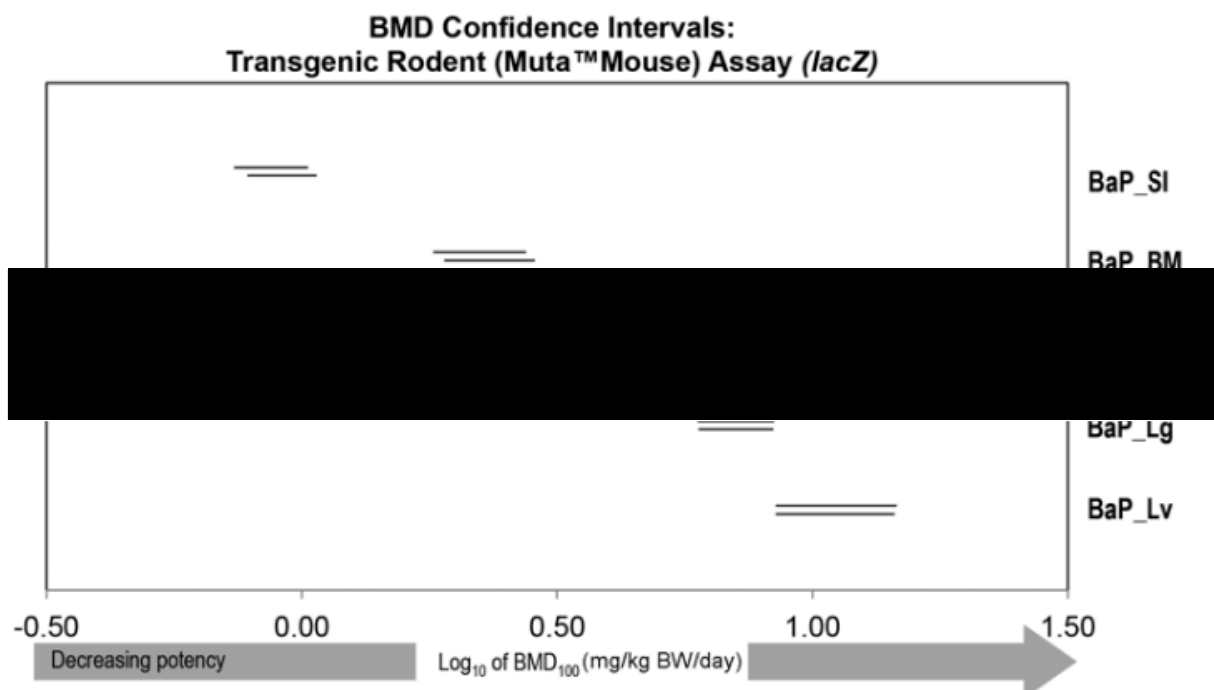
Figure 4.5. BMD confidence intervals of dose-responses from 28-day repeat dose benzo[a]pyrene oral exposure *in vivo* MutaMouse studies analysed for LacZ mutant frequencies. Tissue type served as the covariate in the BMD analysis. BMD confidence intervals were plotted for the exponential and Hill fitted dose-response models and decreasing orders of potency was interpreted as decreasing tissue-specific sensitivity to exposure to the test compound. Abbreviations: SI = small intestine; BM = bone marrow; GS = glandular stomach; Lg = lung; Lv = liver. Figure taken directly from Wills et al. (2016b).

In the 4 studies just summarised, potency comparisons were performed by visually comparing the relative positions of the confidence intervals. A benefit of performing combined covariate BMD analysis is that the precision of the BMD improves thus increasing the likelihood of having distinctly separate BMD confidence intervals that can been used to draw conclusions on statistically significant differences in compound potency (Slob and Setzer, 2014; Wills et al., 2016a). This is generally sufficient to make conclusions such as compound potency rank order within an endpoint, interlaboratory variability, or the cell- or tissue-specific sensitivity to a genotoxic compound. However, some recent publications have aimed to derive numerical values from the BMD comparative potency analyses. It is hypothesised that a numerical value adds an extra element of quantitative meaning to the result of the comparative potency analysis. The recent published examples are highlighted in

the next subsections of this chapter, along with the associated issues pertaining to the interpretation of the approaches.

4.1.4 Instances where BMD Confidence Intervals Have Not Been Used in BMD Potency Comparative Analysis

In contradiction to the previously referenced pivotal studies where potency comparisons were made by visually comparing BMD confidence intervals, other studies have used the BMD metrics values (BMD, BMDL and BMDU) to make potency comparisons. One example highlighted in this section includes using the BMD-BMD ratio to make comparative potency decisions without considering the uncertainty associated with the BMD measurements.

4.1.4.1 Quantitative Comparisons of *In Vitro* Genotoxicity Potential of Lipophilic Phycotoxins

Le Hegarat et al. (2020) were interested in the comparative potency of the phycotoxins okadaic acid (OA) and dinophysistoxins 1 and 2 (DTX-1 and 2) with respect to DNA damage responses including γH2AX and phosphohistone-H3 (p-H3) in proliferative *in vitro* liver HepaRG cells. Briefly, the experimental method included exposing multiple concentrations of the test compounds to *in vitro* cultures of HepaRG cells for 24-hours and sampling the cells followed by a sequence of washing, tagging of fluorescent antibodies, and plating for evaluation by cellular imaging and high content analysis for the γH2AX and p-H3 biomarkers. Data analysis included BMD modelling of the γH2AX and p-H3 biomarker dose-responses to compare the genotoxic potencies. The researchers concluded that DTX-1 was the most potent compound based on the compound having the lowest BMD value. Taking a closer look at the results in the research article shows that DTX-1 and DTX-2 have overlapping BMDLs. As stated by Wills et al. (2016a), the BMD confidence interval represents the range in which the true BMD can lie, and hence the drawing a conclusion based on a BMD point estimate disregards the level of uncertainty related by the BMD confidence intervals. Another issue with the results presented in the article is that the BMDU values were not reported so it is not conclusive whether bound or unbound confidence intervals were obtained.

4.1.4.2 BMD Ratios to Determine Study Duration Adjustment Factors

Other researchers have also disregarded the BMD measures of uncertainty by using the BMD ratio in comparative analysis. White et al. (2020) used different published dose-response datasets for the peripheral blood *Pig-a* gene mutation assay from rats treated with aristolochic acid (AA) or methyl methanesulfonate (MMS) for 3 and 28-days. BMD modelling was performed to compare the BMD ratio of *Pig-a* bed blood cell (RBC) mutant frequency from the 3-day and 28-day treatment regimens. Comparing BMD point estimate ratios across the 2 treatment regimens yielded BMD ratios (in the original scale) of 11.7 for aristolochic acid exposure, and 5.2 for MMS exposure (**Figure 4.6**).
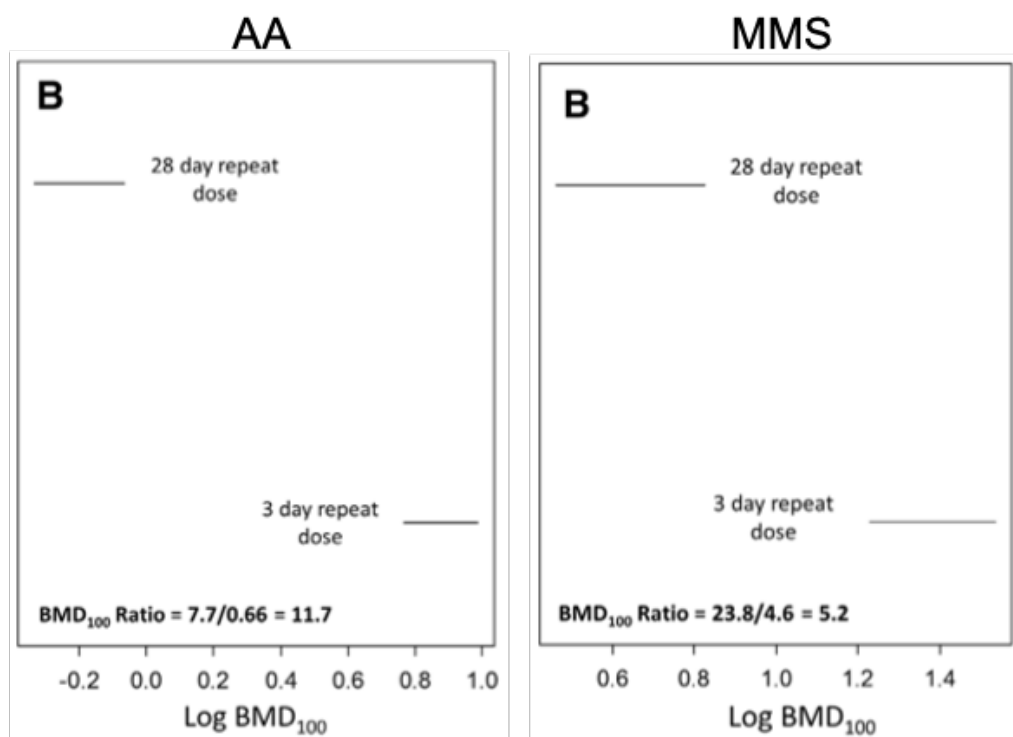
Figure 4.6. BMD Confidence Intervals from *Pig-a* bed blood cell (RBC) mutant frequency from the 3-day and 28-day treatment regimens of aristolochic acid (AA) or methyl methanesulfonate (MMS). Figure adapted from White et al. (2020).

The context of the comparative analysis was to assess whether study duration adjustment factors (AFs) in the range of 6-10 (recommended by regulatory guidance) are appropriate for determining health-based guidance values when applied to PoDs derived from nonclinical studies where the treatment duration is less than chronic. Comparison of the BMD ratio suggests that a conservative AF value of 10 is appropriate; however, the use of the

BMD point estimate disregards the uncertainty (potency range) of the dose-response. The true potency range would be expressed as the comparison between the 28-day BMDU and 3-day BMDL, to the 28-day BMDL and 3-day BMDU. This represents the minimum and maximum difference in potency between the 2 test conditions. Therefore, for aristolochic acid exposure, the potency ratio between 28-days and 3-days exposure is 6.74 to 20.64 mg/kg/day. Similarly, for MMS exposure, the potency ratio between 28-days and 3-days exposure is 2.31 to 11.83 mg/kg/day. One now ponders whether an AF >10 would be more appropriate in these instances given the upper range of uncertainty in the BMD metrics from this comparative analysis? The issues concerning AF derivation is beyond the scope of this work and the reader is encouraged to refer to the original article by White at al. (2020).

4.1.4.3 Published MultiFlow Data Comparing S9 Concentration Impact on Genotoxic Potency

Tian et al. (2020) published their findings on the use of a low concentration rat liver S9 enzyme/cofactor mix and how it influences the outcome of compound genotoxic potencies when evaluated using the MultiFlow assay. MultiFlow assay results have been extensively published for the p53-proficient TK6 human lymphoblastoid cell line. These cells cannot reliably detect pro-genotoxicants without a source of exogenous metabolic activation. The conventional approach for overcoming this limitation is to test compounds in the presence and absence of rat liver microsomes (known as the S9 fraction) with required cofactors (Ames et al., 1973). Tian and colleagues (2020) recently presented several problem statements associated with the use of rat liver S9. Supply continuity of Aroclor-induced rat liver S9 provides a challenge, where users will have to transition to an alternate induction scheme, for example phenobarbital/β-napthoflavone. High concentration S9 (2-4% v/v final) is inherently cytotoxic, and most assay protocols call for a washing step to remove test chemical and S9 enzymes after a relatively short exposure period (generally 3-4 hours). This is suboptimal to high-throughput systems since the step is not easily automated. Moreover, centrifugation and aspiration can potentially impact cell health and numbers which may be incorrectly attributed to test compound toxicity as opposed to the impact of processing.

Tian et al. (2020), investigated the use of phenobarbital/β-napthoflavone-induced rat liver S9 at maximal non-cytotoxic concentration (0.25% v/v final) in the MultiFlow assay. The laboratory was able to maintain the S9 enzyme/co-factor mix with cells and test compound for the entire exposure period of 24-hour in the TK6-based assay. The

investigators focused on 15 chemicals: 8 of which are clastogens that are known to require metabolic activation to maximize formation of DNA-reactive metabolites; 5 are cytotoxicants; and 2 are direct acting clastogens that do not require metabolic activation. In addition to determining compound mode of action using biomarker responses, the authors applied the BMD approach with the aim of calculating a numerical 'S9 potentiation ratio' value, which served as a comparison metric obtained by division of the BMD value in the absence of S9 by the BMD value in the presence of S9. The authors presented the 'S9 potentiation ratio' values in a table alongside the compound and S9 concentration condition.

4.1.5 A Hypothesis Developed from the Tian et al. (2020) Publication

The primary shortcoming highlighted in the 'S9 potentiation ratio' technique proposed by Tian et al. (2020) is that comparing BMD point estimates ignores the associated uncertainty related to the BMD estimation. Using the same logic highlighted by Wills and colleagues (2016), it is irreconcilable to rely upon a BMD point estimate to robustly compare potencies, since the associated BMD confidence interval represents the range in which the true BMD lies.

Whilst reflecting on this primary hypothesis, the authors of the Tian et al. (2020) manuscript at Litron laboratories were consulted for a discussion on the merit for further analysis of the same dataset. Whilst the authors initially acknowledged much of the same shortcoming related to the BMD point estimate relative to the uncertainty reported in the BMD confidence intervals, the authors were insistent that a numerical value is a practically favourable approach to easily inform the potency differences of a compound tested in the MultiFlow system under different S9 concentration conditions. Whilst the scientific literature advocates that graphical representation of BMD confidence intervals is the simplest method to interpret potency differences between covariates (Slob and Setzer, 2014; Wills et al., 2016a; Wills et al., 2016b; Wills et al., 2017; White et al., 2020), a common ground was reached and a further hypothesis developed which questioned whether it was possible to determine a more statistically robust value compared to the 'S9 potentiation ratio' proposed by Tian et al. (2020).

Primarily, it was stressed that the use of a BMD point estimate value does not provide an accurate representation of the likely potency range of the test compound. Said another way, the BMD point estimates and associated 'S9 potentiation ratio' comparison metrics did

not convey information about the uncertainty of the measurements that are consistent with the BMD uncertainty measurement approach that is advocated in the scientific literature.

This chapter focuses on the reanalysis and augmentation of the published Tian et al. (2020) dataset. Specifically, the BMD uncertainty measurements that relate to the relative genotoxic potency of the 10 clastogenic compounds are conveyed in detail. To this end, an 'S9 potency ratio CIs' is calculated using the BMD uncertainty measurements (BMDL and BMDU) between S9 exposure conditions (the presence and absence of S9) based on a novel developed algorithm, and said ratios were utilised to derive robust potency conclusions as a follow up to the Tian et al. (2020) work.

4.2 Materials and Methods

The following flow chart (**Figure 4.7**) summarises the methodology steps that were followed to arrive at the results presented in this chapter. Each subsection of this materials and methods chapter provides detail on steps 2 to 7 from the flow chart.

```
┌─────────────────────────────┐
│ 1: Consultation with Litron  │
│ for Hypothesis Development   │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 2: In Vitro Genotoxicity     │
│ Dataset Database Curation    │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 3: Combined Covariate BMD    │
│ Analyses                     │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 4: Confidence Interval       │
│ Determination and Plotting   │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 5: BMD Metrics Plotting      │
│ Radar Plots and Potency Plots│
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 6: 'S9 potency ratio CIs'    │
│ Algorithm Development        │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 7: Unsupervised Clustering   │
└─────────────────────────────┘
```
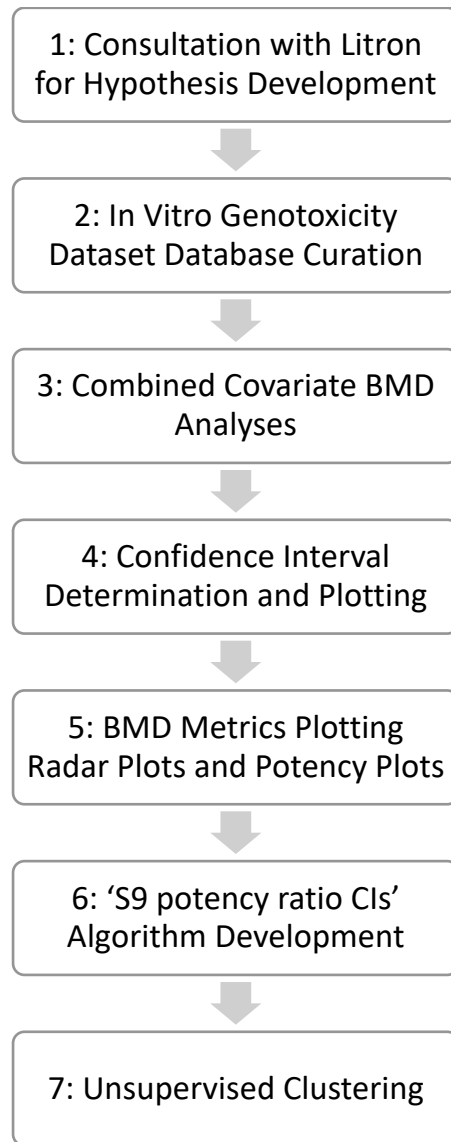
Figure 4.7. Materials and Methods Flow Chart

4.2.1 *In Vitro* Genotoxicity Dataset

The data were derived from a previously published article in which 15 compounds were studied using the *in vitro* MultiFlow DNA Damage Assay in the presence and absence of low dose (0.25% v/v) S9 (Tian et al. 2020) and provided by Litron Laboratories in Excel format. As previously mentioned, the Tian et al. (2020) data included 8 direct acting clastogens that require metabolic activation, 5 cytotoxicants and 2 direct acting clastogens that do not require metabolic activation. In this reanalysis the focus was on the 10 clastogens since this is where Tian et al. (2020) suggested the greatest differences in potency between S9 exist. The raw data were provided by Litron Laboratories and included 4- and 24-hour p53, γH2AX, and 24 hour RNC responses for the 10 clastogens: 2-acetylaminofluorene, 2-

aminoanthracene, 7,12-dimethylbenzanthracene, benzo[a]pyrene, cyclophosphamide, dibenzo[a,l]pyrene (also known as dibenzo[def,p]chrysene), diethylnitrosamine, mitomycin C, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), and resorcinol. The identity of the 10 clastogen compounds is provided in **Table 4.1**.

Table 4.1. Clastogen Compounds with MultiFlow Dose-Response Datasets with and without Low Concentration S9 (0.25% v/v)

Compound	CAS No.
2-Acetylaminofluorene	53-96-3
2-Aminoanthracene	613-13-8
7,12-Dimethylbenzanthracene	57-97-6
Benzo[a]pyrene	50-32-8
Cyclophosphamide monohydrate	6055-19-2
Dibenzo[a,l]pyrene	191-30-0
Diethylnitrosamine	55-18-5
Mitomycin C	50-07-7
2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (PhIP)	105650-23-5
Resorcinol	108-46-3

Biomarker responses included γH2AX and p53 since these biomarkers have been demonstrated responsive for the clastogenic mode of action in the MultiFlow assay (Bryce et al., 2016; Bryce et al., 2018), thus excluding the p-H3 biomarker from analysis. Moreover, the same biomarker dose-response dataset was included in the original Tian et al. (2020) publication.

4.2.2 Database Curation

The PROAST software requires data to be formatted for analysis including the following specific strict rules:

- All decimal points in point format not comma format
- Empty cells are not allowed
- Spaces between text or units are not allowed
- The raw data takes the format of the data matrix. In this case, the data matrix was the compound names, concentration values, a priori mode of action information, median fluorescence, median relative frequency, or percent values.
- Each column must have a header which describes the data in the data matrix
- Four rows must exist above the data matrix and complete according to the following rules:
    - First line: a one-word title that describes the data. In this case the MicroFlow data had the heading "metabolism".
    - Second line: the number of columns in the data matrix
    - Third line: includes a numerical code for the data type in the column (see below)
    - Fourth line: a one-word title for the column. The column title choices for the MultiFlow dataset are presented in **Table 4.2**.
- Data types (third line)
    - 0 – for all columns not representing a response variable
    - 1 – continuous data. In our case, all MultiFlow biomarker response variables (γH2AX, p53).
    - 2 – binary data (e.g., yes/no response in individual animals)
    - 3 – ordinal data (e.g., histopathology scores of severity)
    - 4 – quantal data (e.g., number of responding animals per number of animals in the group)
    - 5 – continuous data that are nested (e.g., fetal weights within litters)
    - 6 – quantal data that are nested (e.g., number of fetuses affected within litters)
    - 10 – mean (continuous) response in a dose group

Data types 2 to 10 are out of scope of the MultiFlow assay result variables. Hence the third line for the MultiFlow databases all included either a value of 0 or 1.

Only the pertinent biomarker response raw data was provided by Litron Laboratories including median fluorescence raw data for the γH2AX and p53 biomarker responses at both the 4-hour and 24-hour timepoints, the percent RNC frequency data at both the 4-hour and 24-hour timepoints, chemical name, concentration (μM), and S9 condition [with or without low concentration S9 (0.25% v/v)]. **Table 4.2** summarises the final selected database columns, rows, and headers formats for the MultiFlow dataset. Once the Excel database was formatted according to the strict rules prescribed by PROAST, the Excel sheet was transformed into tab delimited file (.txt) and stored in a working repository.

Tian et al. (2020) performed BMD analysis on individual compounds with the S9 exposure (with or without [+/-] 0.25% v/v S9) condition serving as the covariate. Hence, individual dose-response curves on a compound basis already existed for this dataset. Here, the individual dose-response curves were scrutinized so that compounds and/or endpoints with little to no evidence of a dose-response could be disqualified from reanalysis. Hence, the following compound/biomarker combinations were excluded from reanalysis: 4-hour p53 dose responses for diethylnitrosamine; 24-hour p53 dose-responses for diethylnitrosamine; and 24-hour γH2AX dose responses for 7,12-dimethylbenzanthracene.

Table 4.2. Replication of MultiFlow database columns, rows, and headers formatted per PROAST strict data formatting requirements

Row Number									
1	metabolism								
2	9								
3	0	0	1	1	1	1	1	1	1
4	Chemical	S9.condition[6]	Conc.µM	p53.4hr	p53.24hr	g.H2AX.4hr	g.H2AX.24hr	Perc.RNC.4hr	Perc.RNC.24hr
5 and more…	Data Matrix…								

---

[6] S9.condition was denoted as "without.S9" or "with.S9" in the data matrix following the strict PROAST data format conditions

4.2.3 Working in PROAST

PROAST v69.1 software package (http://www.proast.nl) operating in the R statistical suite was used for the BMD analyses presented in this chapter. The 'MENU' version of PROAST was used to navigate through the data and variables selection process in the BMD analyses. Firstly, the .txt file was imported into PROAST for analysis. The 'continuous' data type was selected from the MENU option, and 'concentration' selected as the 'independent variable'.

4.2.3.1 Obstacles Associated with the Data Processing

Previous experience (illustrated in Chapter 3) where compound serving as the covariate demonstrated that good quality potency ranks can be determined for compound studied with the MultiFlow assays biomarkers responses. For the study reported in this chapter similar high quality potency ranks were desired, and hence combined covariate analysis was the preferred method. In this respect, multiple analyses were performed to overcome several challenges in the data analysis.

Firstly, a low processor speed of one of the computers that was used impeded the concomitant analysis of all 10 chemicals in a single combine covariate analysis. To overcome this obstacle and use the same standard computer (8GB RAM), combined covariate analysis was performed using groups of compounds by grouping compounds for analysis into broadly similar chemical structural similarities where preferable, or by hypothesized mode of action. Compounds were therefore grouped into; 1) the aromatic amines 2-acetylaminofluorine, 2-aminoanthracene, and PhIP ((2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine)); 2) the polyaromatic hydrocarbons benzo[a]pyrene, dibenzo[a,L]pyrene, and 7,12-dimethylbenz[a]anthracene (dimethylbenzanthracene); 3) the alkylators diethylnitrosamine, and mitomycin c; and a 4th group containing cyclophosphamide and resorcinol that did not belong into the other groups. Although performing BMD analysis by small groups of similar compounds appears to be a viable option, it became apparent that the influence of the groupings needed to be evaluated for the introduction of bias.

Secondly, a more powerful computer (16 GB RAM) was utilised to perform BMD analysis of all 10 compounds concurrently with 'chemical' selected as the covariate in the combined covariate analyses. In doing so, this ensured that there was no introduction of bias into the analysis whilst being consistent with the state-of-the-art literature where compound

as the combined covariate results in precise BMD confidence intervals due to conserved dose-response shape parameters (Slob and Setzer, 2014; Wills et al., 2016a).

For both of these groups of analyses where either BMD analysis was performed by sub-groupings or with all 10 compounds, settings where changed in the PROAST MENU option to analyse "with.S9" separately from "without.S9" since this is considered a separate covariate. In the Tian et al. (2020) study, compounds were analysed separately with S9 condition (with/without S9) serving as the covariate. This meant that 10 sets of BMD analyses had to be performed (one for each compound) for each biomarker (4- and 24-hour p53, 4- and 24-hour γH2AX, and 24-hour RNC), totalling 50 BMD analysis outputs. One contended that including all 10 compounds in a single analysis would significantly reduce the workload of a laborious process of repeating analysis on a compound basis with S9 condition serving as the covariate, as well as increasing the precision of the BMD.

4.2.4 BMD Analysis

The MENU option of PROAST v69.1 was navigated to select the appropriate parameters to perform BMD analysis of the MultiFlow biomarker response dataset. The responses included each of the MultiFlow biomarkers consistent with the Tian et al. (2020) analyses – being 4- and 24-hour p53, 4- and 24-hour γH2AX, and 24-hour RNC responses.

PROAST v69.1 included the option to select either model 3 or model 5 when performing BMD analysis to reflect the latest BMD guidance recommended by EFSA (EFSA, 2017). In doing so, the model with the lowest Akaike Information Criterion (AIC) is selected for calculating the BMD confidence interval. In brief, AIC provides the means for model selection based on the quality of each model considering the prediction error within each model (McElreath, 2018; Taddy, 2019). The AIC analyses resulting in the selection of the either model 3, or maximal 4-parameter exponential model (model 5) based on the models containing the appropriate number of parameters for describing the dose responses (Slob and Setzer, 2014; EFSA, 2017). The exponential model equations are presented in the materials and methods chapter of this thesis (Chapter 2.3). Briefly, the model parameters include the following: a) the response within the controls; b) potency/efficacy of the compound on the shape of the response; c) the maximum fold-change in the response; and d) the rate of change in the response. The four parameter model contains all parameters (a,b,c,

and d), whereas model 3 lacks parameter 'c' where the maximum fold change in the response is not considered.

4.2.4.1 Critical Effect Size Selection

Consistent with the approach utilized by Tian et al. (2020), an arbitrary critical effect size (CES) of 0.3 was selected for this reanalysis. A CES of 0.3 represents a 30% change in response compared to the concomitant control. CES -0.3 for the RNC endpoint represents a 30% decrease in response for this endpoint analyses. There is a lack of consensus on the appropriate choice of CES reported in the literature (White et al., 2020) and a lengthy discussion on CES falls outside the scope of this thesis. Briefly, Chapter 3 of this thesis demonstrated that a CES of 0.5 produces compound genotoxicity potency ranks for the MultiFlow biomarker responses, and a CES of 0.3 is in the same order of magnitude as 0.5. In any event, we justify the use of CES 0.3 for these *in vitro* endpoints since the resulting BMDs do not lie in the extremities of the dose response curves where the associated uncertainties could be relatively high.

4.2.4.2 Confidence Interval Determination and Plotting

PROAST default 90% BMD confidence intervals were obtained for each compound and endpoint combination in the presence and absence of S9 and were graphically plotted by the software. The value of the BMD is a point estimate with an associated level of precision (the ratio between the BMDL and BMDU) and to demonstrate that the BMD is not the geometric midpoint of the confidence interval, the BMD was included in re-plotting of the confidence intervals in Excel. The confidence intervals were plotted on a compound basis.

The MultiFlow biomarkers response fold-increase values have been graphically represented in radar plots in a previous publication to rapidly discriminate non-responsive biomarkers from those that are responsive for a particular compound (Dertinger et al., 2019). Thus, it was explored if the same radar plots could add value to confidence interval discrimination of the high-content MultiFlow compound/biomarker response combination datasets in scope of this study. To this end, radar plots were created in excel using the BMDL and BMDU values obtained in this BMD analysis. One issue associated with the plotting of the BMD metrics in radar plots was how to present unbound confidence intervals in the form of infinite BMDU values. This issue was overcome by restricting the infinite BMDU to the BMDL value and marking the data point with a different marker.

4.2.5 'S9 potency ratio CIs' Algorithm

In keeping with the position expressed by Litron Laboratories that numerical values are valuable in potency comparison, the BMD confidence intervals obtained in the analysis were used to calculate an 'S9 potency ratio CI' that is derived for each compound and endpoint BMD analysis that returned a dose response by comparing the BMDU +S9 to BMDL -S9, to the BMDL +S9 to BMDU -S9. The 'S9 potency ratio CI' is synonymous with the magnitude of the potency difference (range) exhibited for each compound/endpoint combination after exposure to S9. Although there were only 2 experimental conditions (+/-S9) included here, the same algorithm could apply to 2 or more experimental conditions where one potency is compared to another potency of interest. The 'S9 potency ratio CIs' were calculated in the original scale so that the values can be compared to the 'S9 potentiation ratios' values obtained from BMD-BMD ratios by Tian et al. (2020). Values greater than 1 represent increased compound potency for a particular endpoint after S9 exposure, whilst values less than 1 confer the inverse. 0 values represent CIs that overlap and hence potency differences are statistically indefensible.

4.2.6 Unsupervised Hierarchical Clustering

Previous publications have used unsupervised hierarchical clustering techniques to derive more information from the MultiFlow biomarker response datasets (Dertinger et al., 2019). The data outputs can be regarded as unbiased interpretations of the underlying datasets that does not rely solely on human interpretation. To the same point and inspired by Litron Laboratories use of hierarchical cluster analysis of MultiFlow biomarker responses, it was hypothesised whether unsupervised hierarchical clustering could add unbiased interpretation of the comparative potency of compounds studies in the MultiFlow assay with and without low concentration S9.

Lower and upper values of the 'S9 potency ratio CIs' (algorithm described in the results and discussion sections) were evaluated using JMP software's unsupervised clustering platform (JMP, v12.0.1). Lower and upper 'S9 potency ratio CI' values associated with the following 5 biomarkers were used as variables: 4-hour p53, 24-hour p53, 4-hour γH2AX, 24-hour γH2AX and 24-hour RNC. The analysis options were set as follows: clustering method = hierarchical; method for calculating distances between clusters = 'Ward'; data as usual = 'Standardize Robustly'; data visualization = 'Dendrogram,' with 'two-way clustering'.

4.3 Results

4.3.1 Problems Associated with Compound Subgrouping BMD Analyses

To demonstrate whether unfavourable bias would be introduced by performing BMD analysis on the small subgroups of compounds based on structural similarities, the BMD confidence intervals in rank order for aromatic amines and polycyclic aromatic hydrocarbons in the 4- and 24-hour p53 timepoint/biomarker combination is included in **Figure 4.8** where compounds were evaluated either as a large group or small subgroup combined covariate. This timepoint/biomarker combination is shown as a general example out of the other timepoint/biomarker combinations. The figure shows that although the confidence intervals for each compound/biomarker combination reside in the same order of magnitude compared between the large group or small subgroup combined covariate, the confidence intervals are generally more precise when the compounds are analysed in the larger group combined covariate.
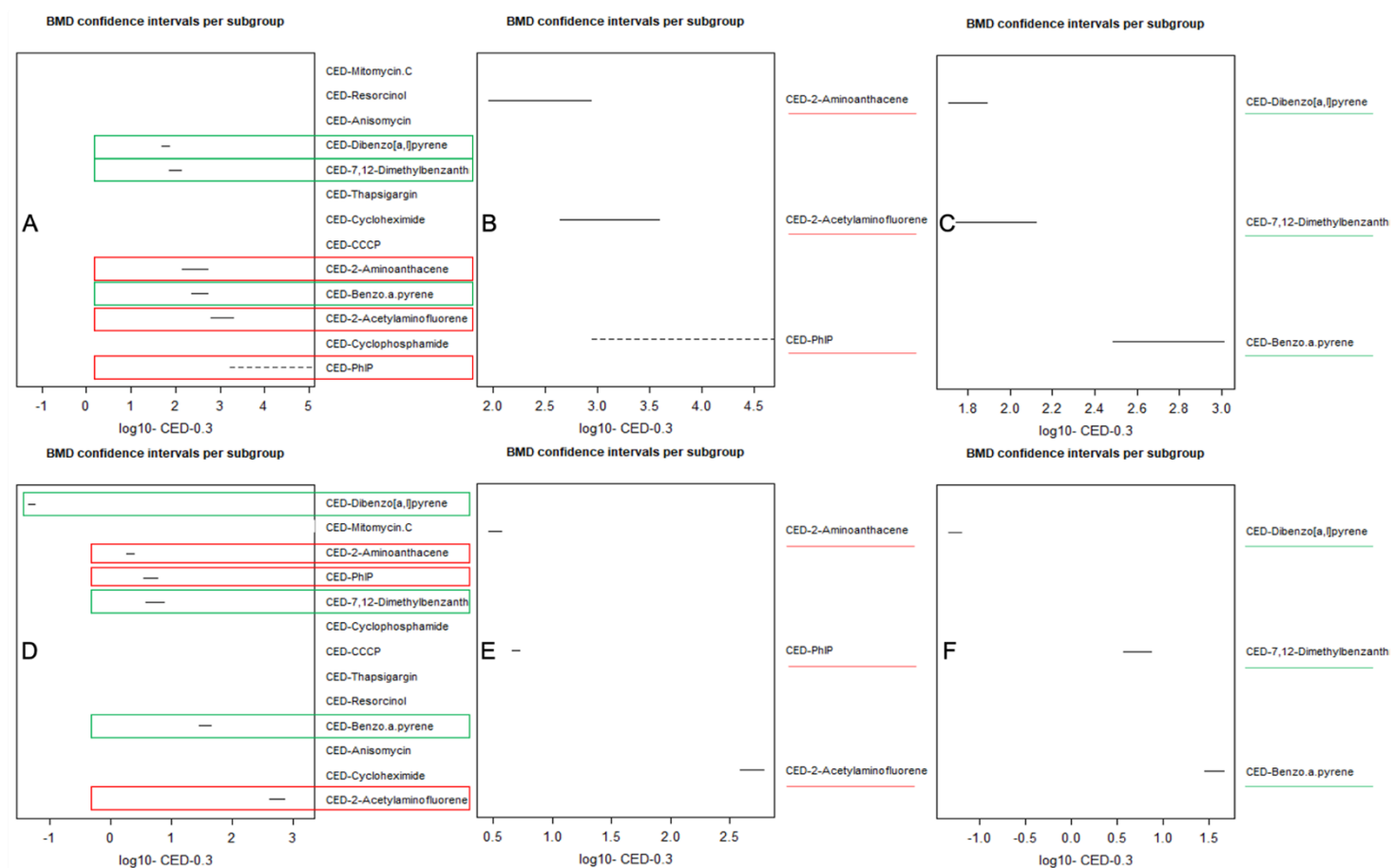
Figure 4.8. BMD confidence intervals plotted in rank order by Log midpoint with CES 0.3. X-axis Log10 concentration μM. Panels A and D show the results of analysis by combined covariate as a large group of compounds. Panels B and E are the aromatic amine (AA) subgroup, and panels C and F are the polycyclic aromatic hydrocarbon (PAH) subgroup. Panels A and D only show the compounds in scope of this comparison to reduce clutter in the image. AAs highlighted in red and PAHs in green.

4.3.2 BMD Confidence Interval Potency Plots

Values of the BMD, BMDL and BMDU were estimated and collated for all compounds (+/- S9) and biomarker responses that were included in the BMD analysis. The BMD confidence intervals were plotted for each compound with +/-S9 confidence intervals side-by-side (**Figure 4.9** to **Figure 4.18**). The BMD point estimates are included in the comparative potency plots to aid in graphical representation of the BMD point estimate relative to the corresponding BMDL and BMDU. This is important since some individuals misinterpret the BMD to be at the midpoint (or geometric mean) between the BMDL and BMDU. This is a misconception as evident by the differing lengths of the confidence intervals either side of each BMD point estimate displayed in **Figure 4.9** to **Figure 4.18**.
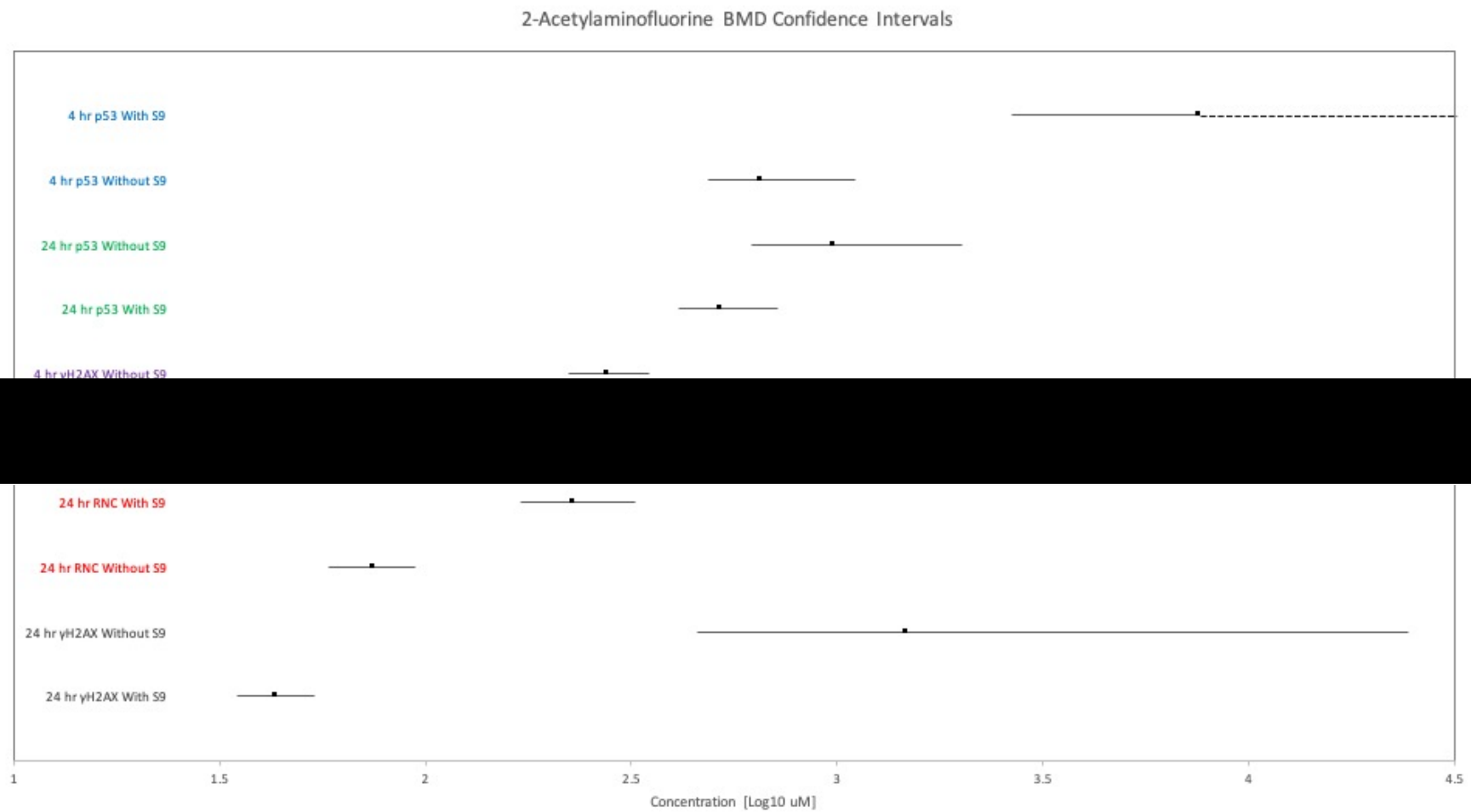
Figure 4.9. BMD Confidence Intervals for 2-acetylaminofluorene. 4-hour p53 biomarker With S9 infinite BMDU indicated with dashed positive direction lines. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
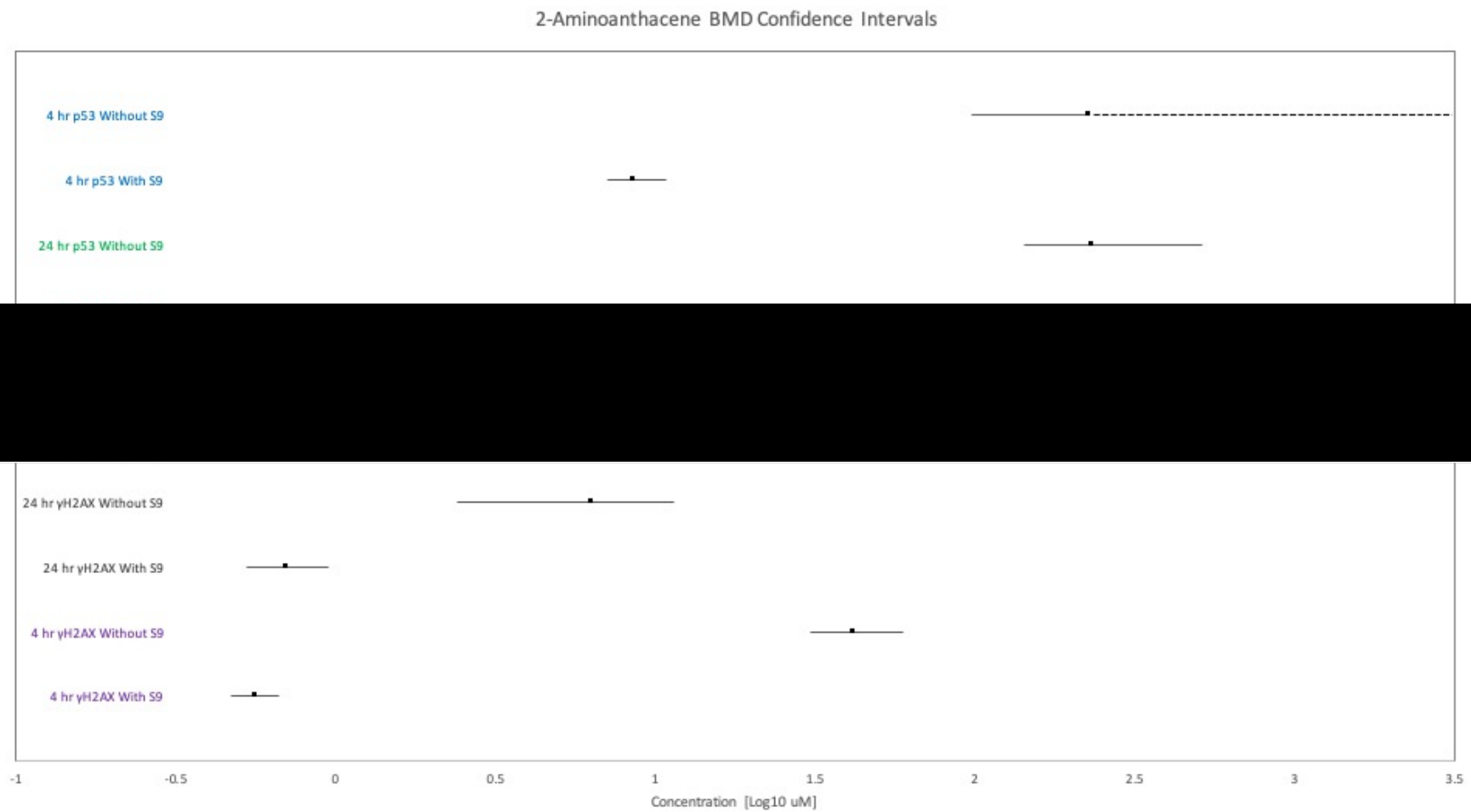
Figure 4.10. BMD Confidence Intervals for 2-aminoanthracene. 4-hour p53 biomarker Without S9 infinite BMDU indicated with dashed positive direction line. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
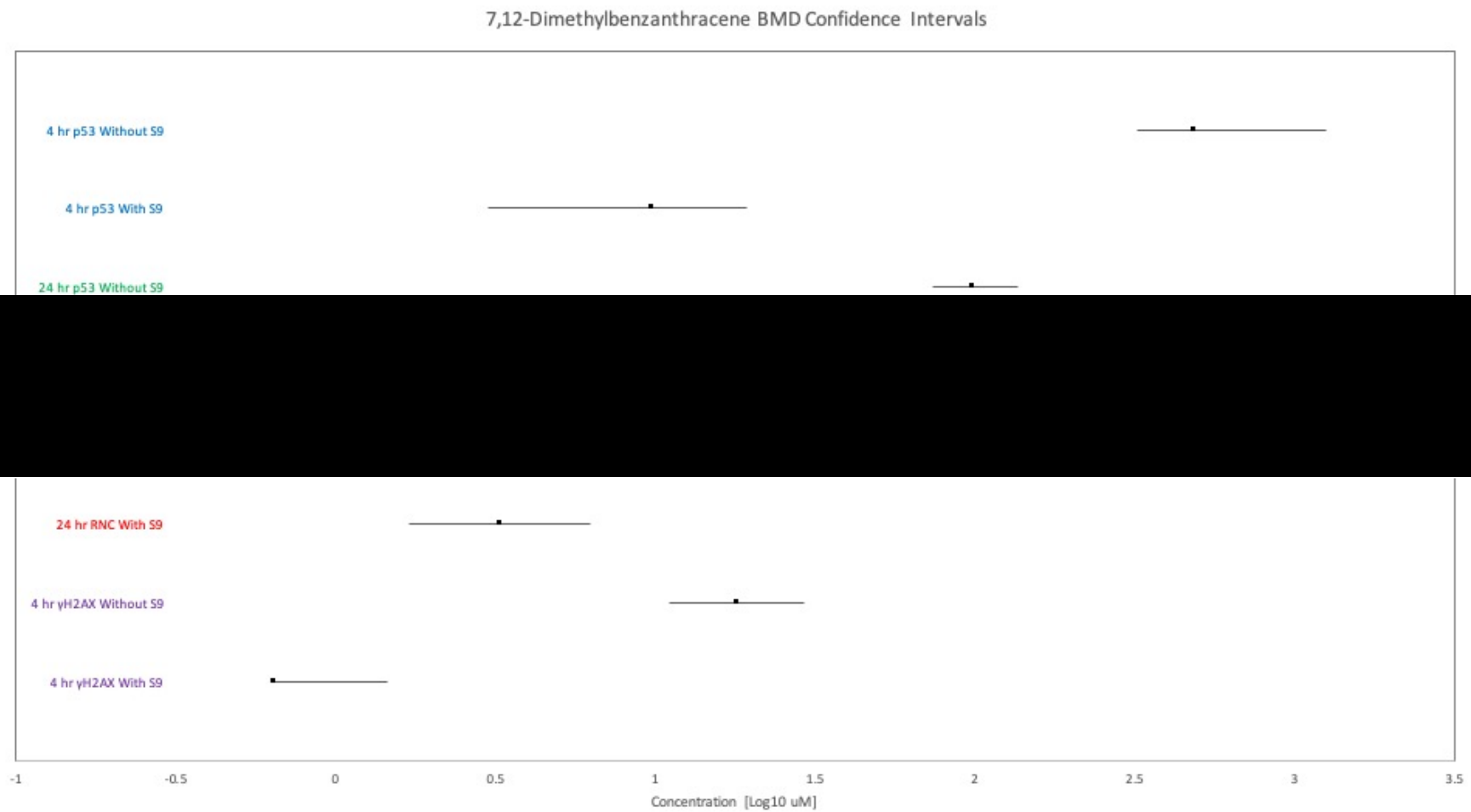
Figure 4.11. BMD Confidence Intervals for 7,12-dimethylbenzanthracene. 4-hour γH2AX biomarker With S9 yielded a zero value lower confidence bound. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
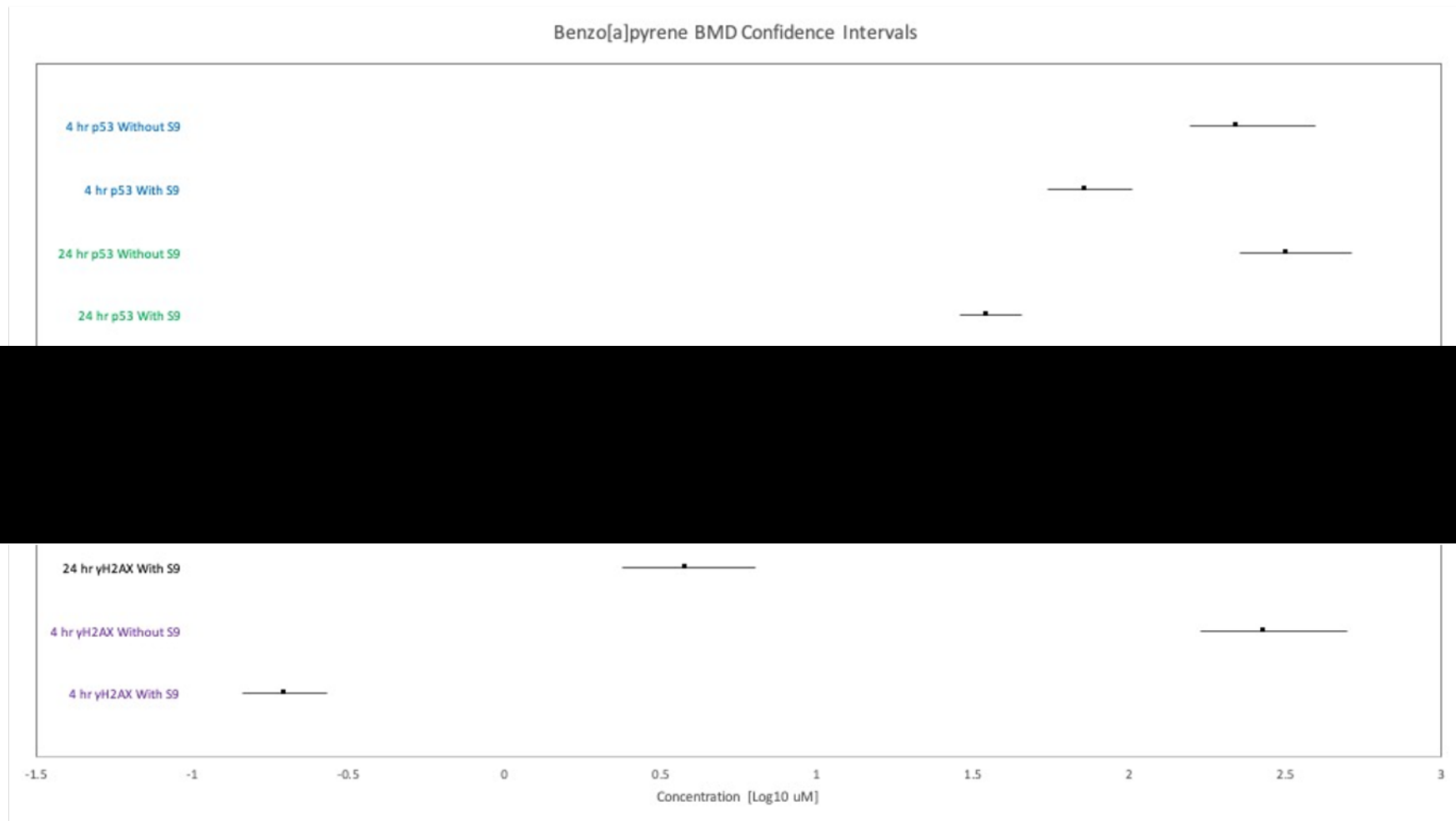
Figure 4.12. BMD Confidence Intervals for benzo[a]pyrene. 24-hour RNC endpoint Without S9 yielded a disproportionately high BMD with infinite BMDU indicated as a dashed line spanning the width of the plot. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
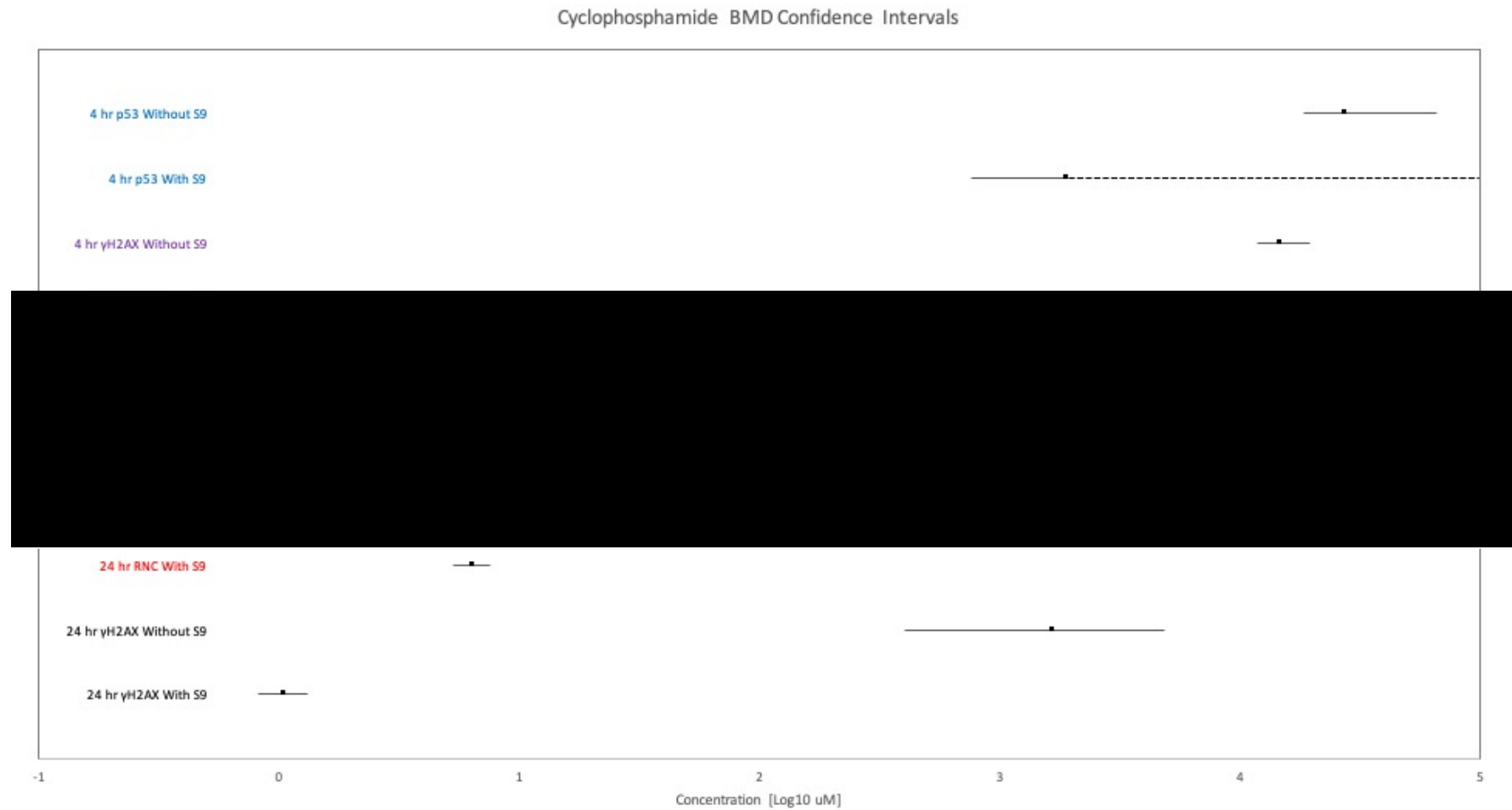
Figure 4.13. BMD Confidence Intervals for cyclophosphamide. 4-hour p53 biomarker With S9 infinite BMDU: indicated with dashed positive direction line. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
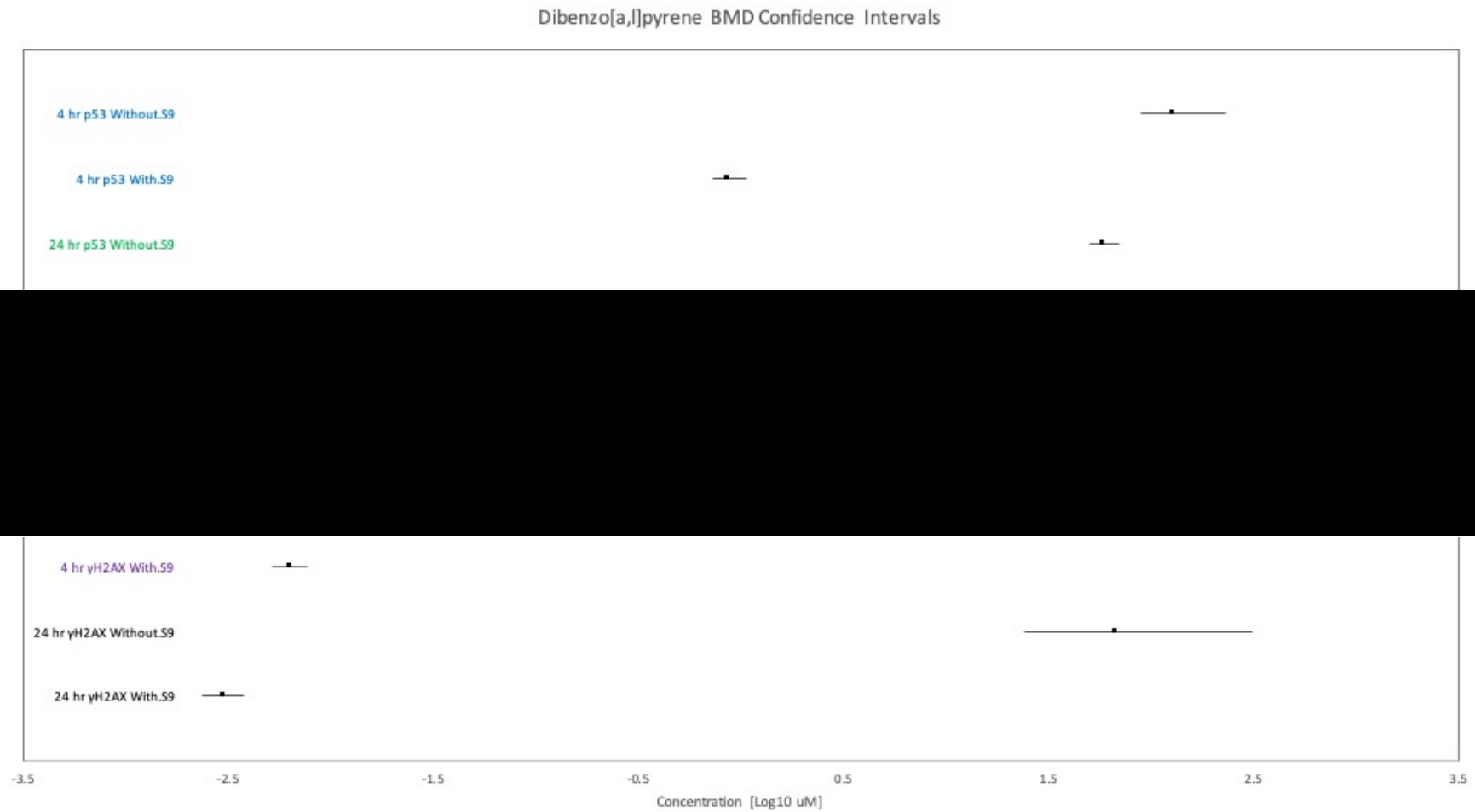
Figure 4.14. BMD Confidence Intervals for dibenzo[a,l]pyrene. 24-hour RNC biomarker Without S9 yielded a disproportionately high BMD with infinite BMDU: BMD restricted and only displaying the BMDL. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).

Diethylnitrosamine BMD Confidence Intervals

24 hr RNC Without.S9

24 hr RNC With.S9

24 hr yH2AX Without.S9

24 hr yH2AX With.S9

2.5   3   3.5   4   4.5   5

Concentration [Log10 uM]
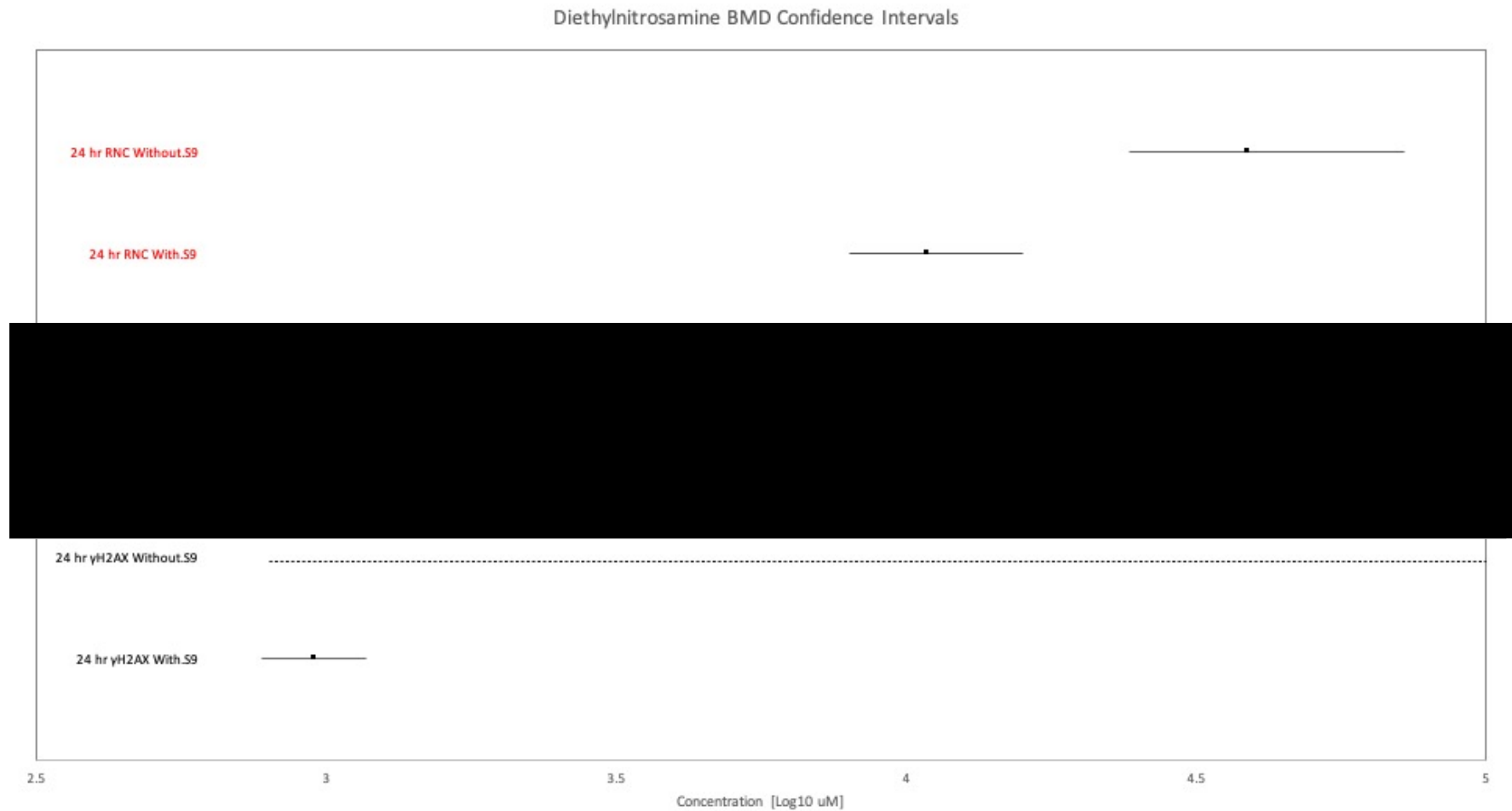
Figure 4.15. BMD Confidence Intervals for diethylnitrosamine. 24-hour γH2AX endpoint Without S9 yielded a disproportionately high BMD with infinite BMDU: indicated as a dashed line spanning the width of the plot. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
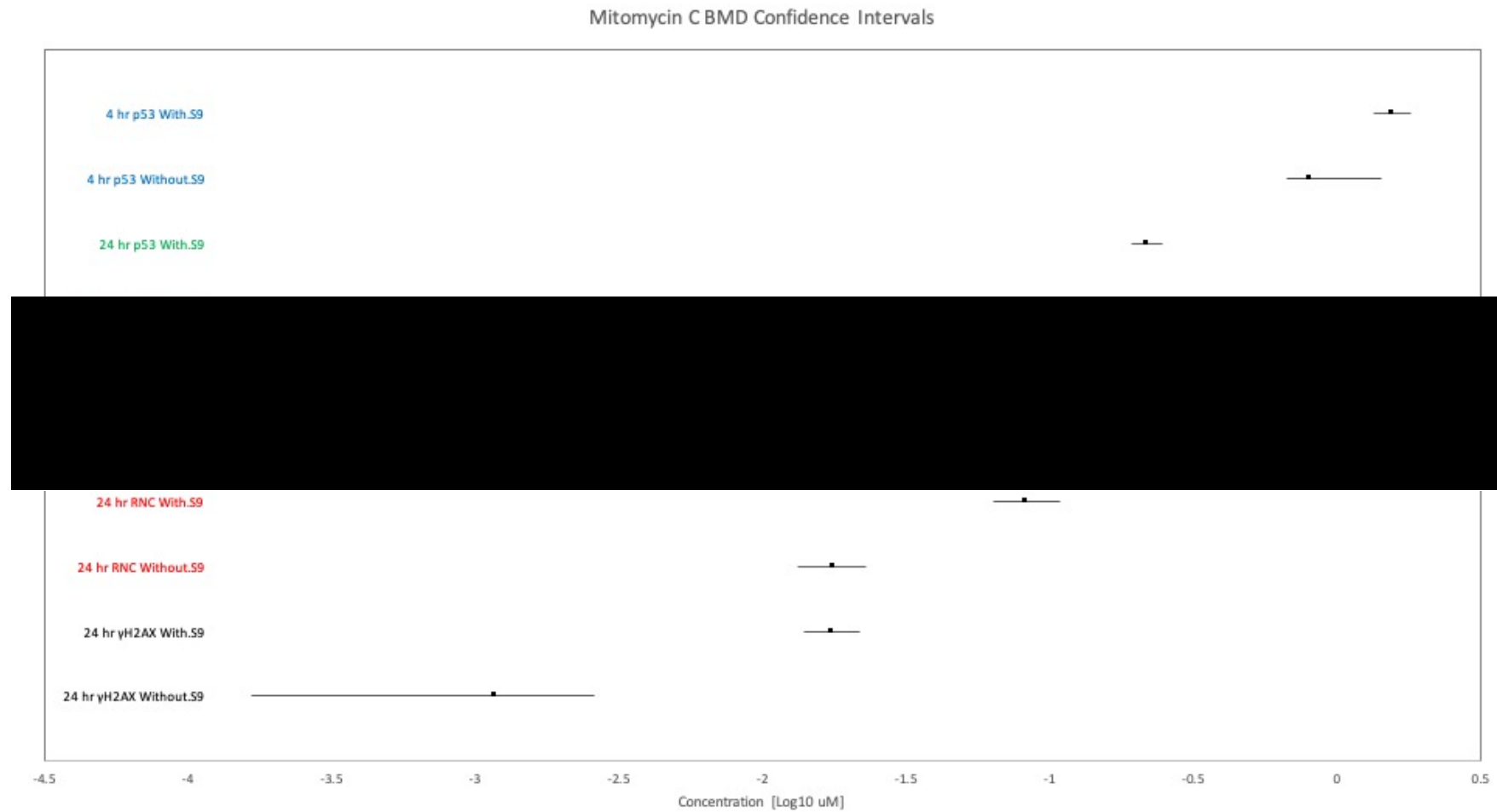
Figure 4.16. BMD Confidence Intervals for mitomycin C. 2-sided confidence intervals were obtained for all endpoint's BMD analyses. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
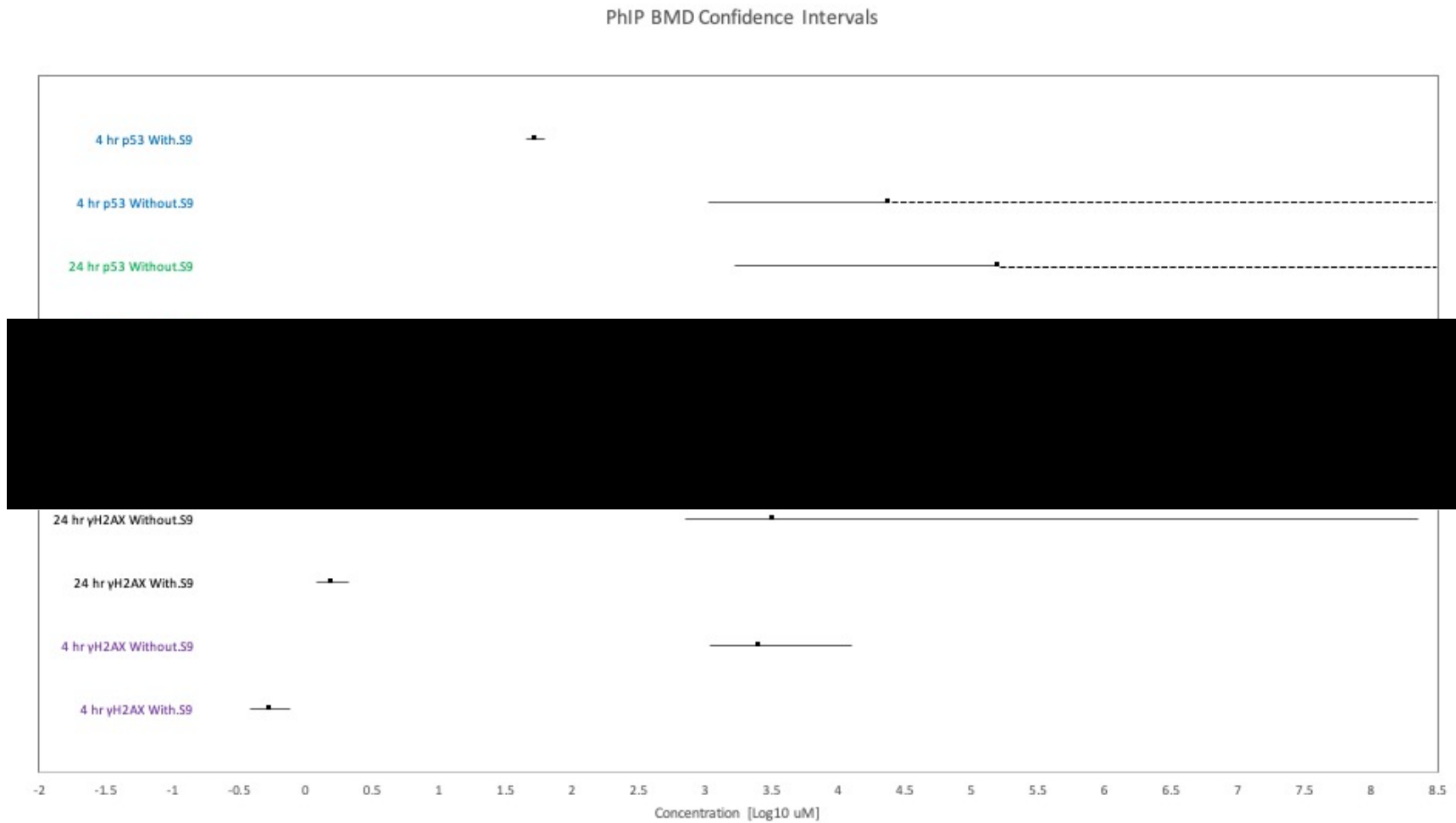
Figure 4.17. BMD Confidence Intervals for PhIP. 24-hour p53 biomarker Without S9 and 4-hour p53 biomarker Without S9 infinite BMDUs indicated with dashed positive direction lines. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).
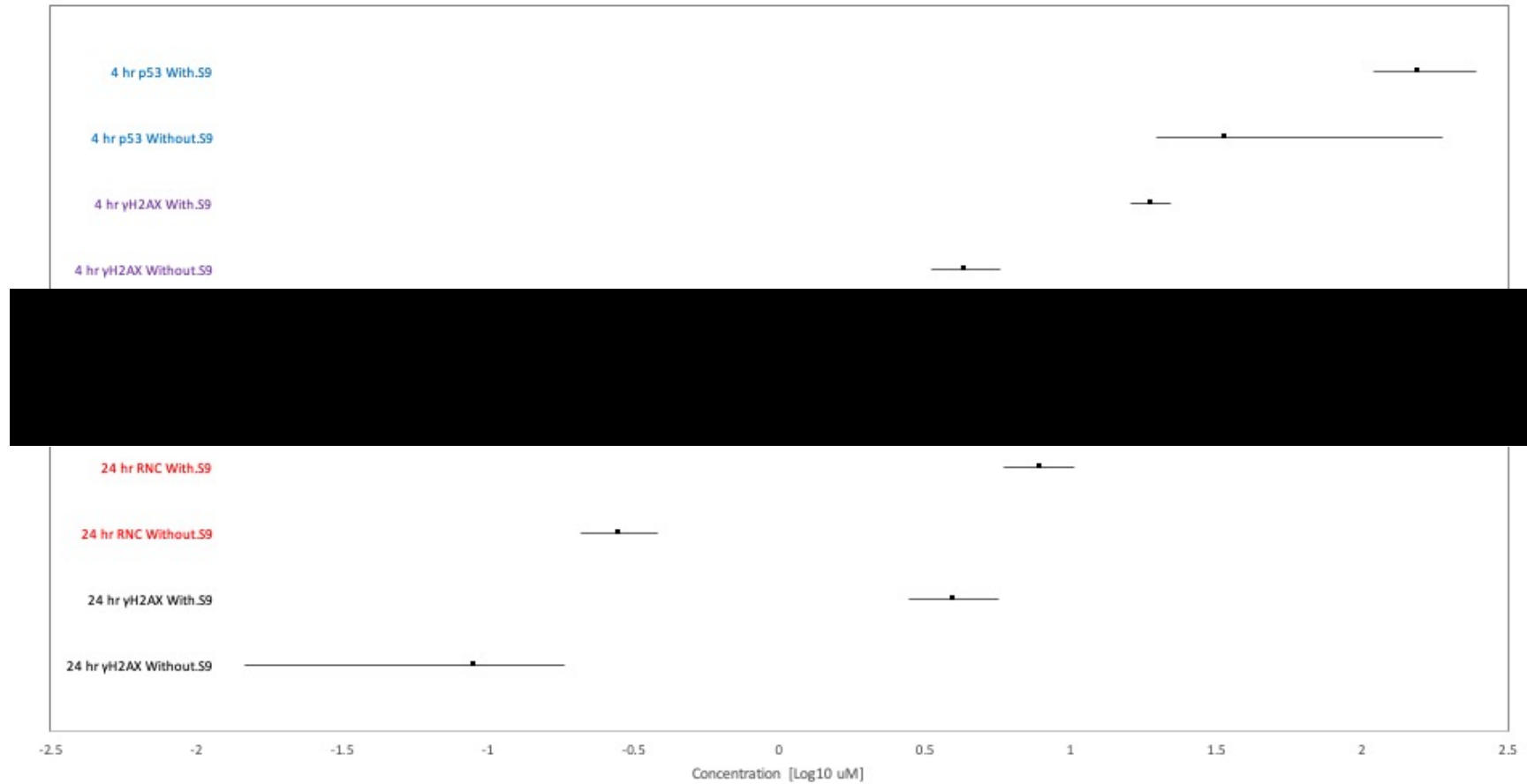
Figure 4.18. BMD Confidence Intervals for resorcinol. 2-sided confidence intervals were obtained for all timepoint/biomarkers' BMD analyses. BMD point estimates are displayed as data points. Figure from Wheeldon et al. (2021).

4.3.3 BMD Confidence Intervals Radar Plots

The same BMD confidence intervals (values of the BMDLs and BMDUs) were plotted for each compound +/-S9 in radar plots (**Figure 4.19** to **Figure 4.28**).
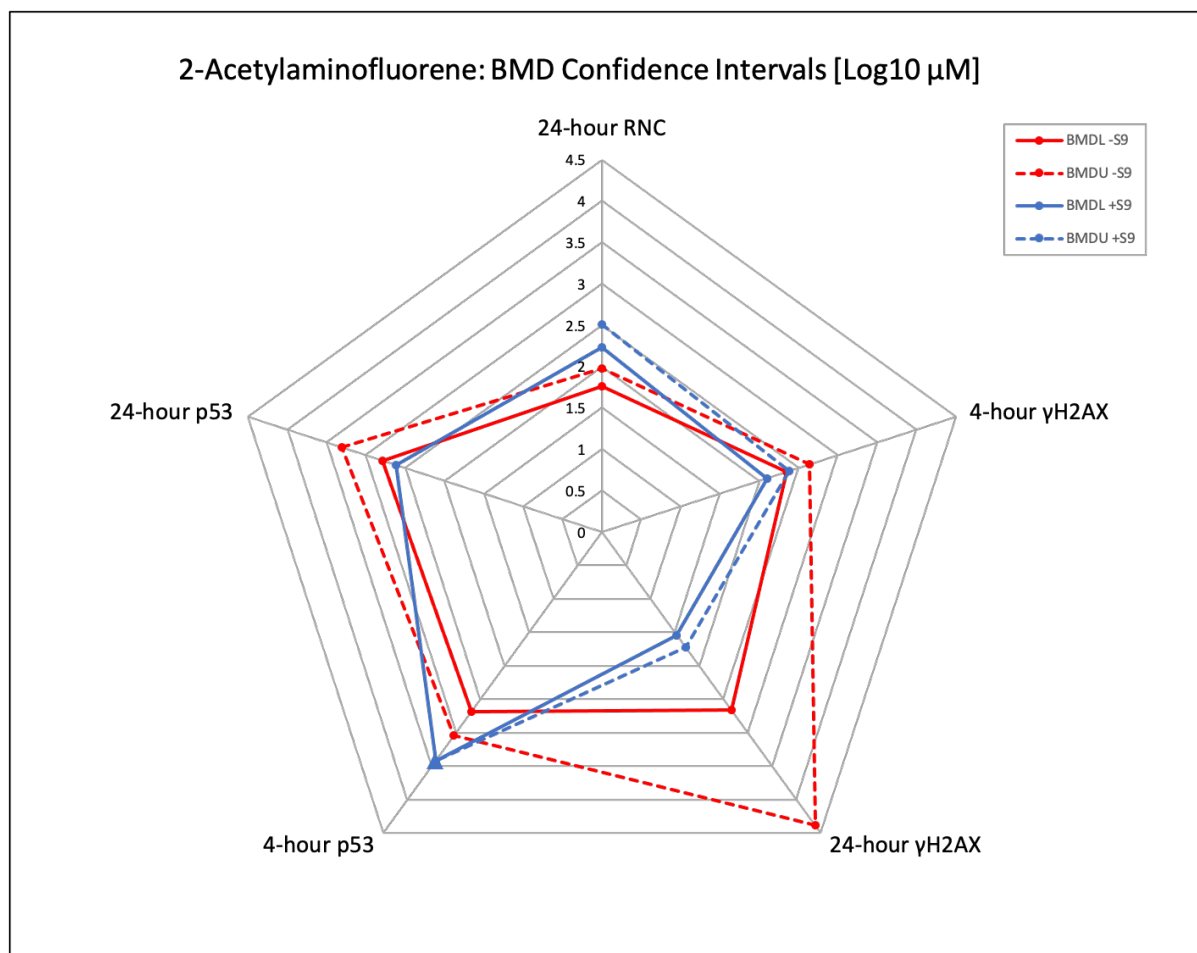
Figure 4.19. Radar plot for 2-Acetylaminofluorine. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 µM scale. An unbound confidence interval limit (infinite, BMDU) in the 4-hour p53 biomarker response is depicted with a large triangle datapoint fixed to the same value as the BMDL. There is a marked increase in potency in the presence of S9 in the 24-hour γH2AX biomarker response.
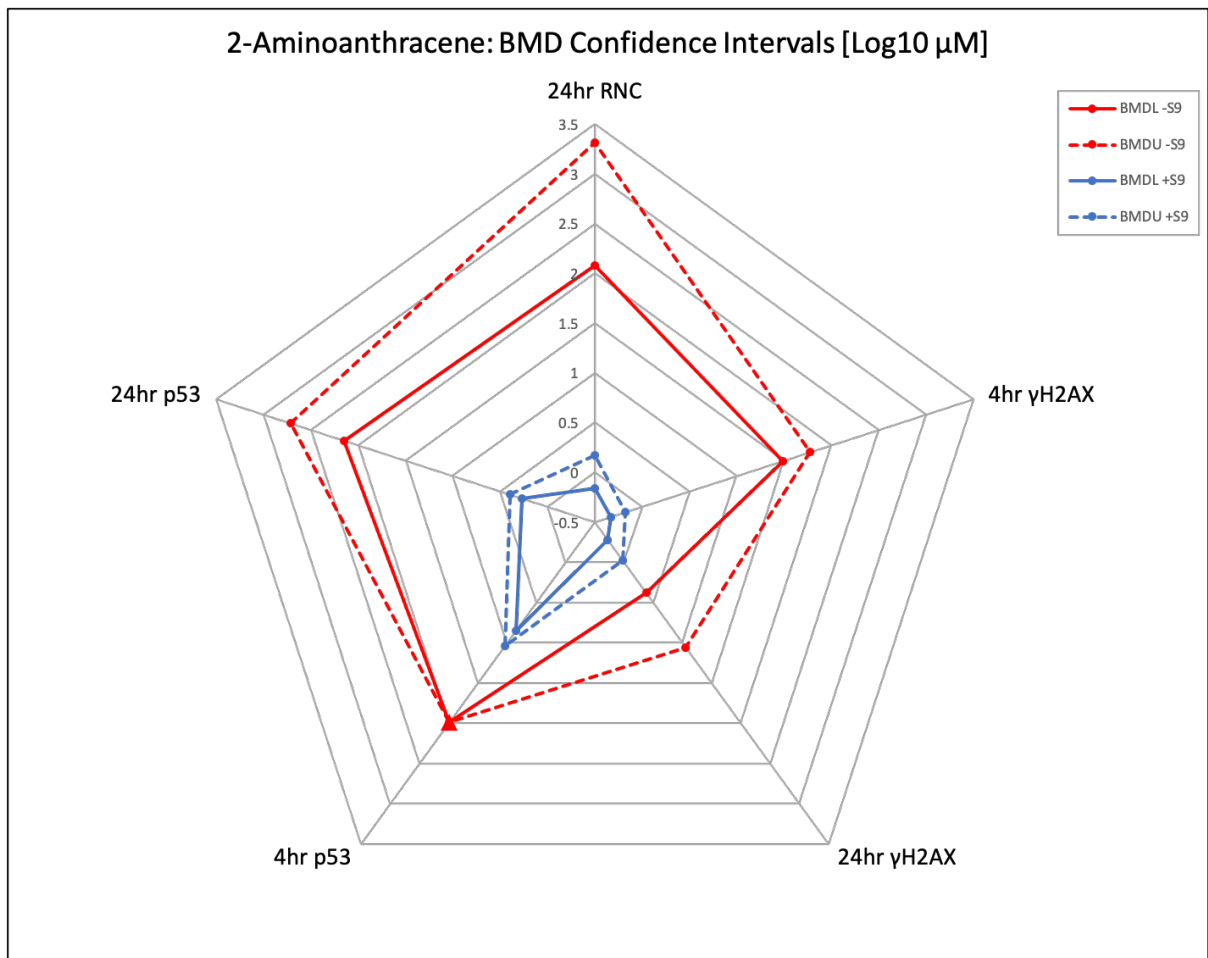
Figure 4.20. Radar plot for 2-Aminoanthracene. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. An unbound confidence interval limit (infinite, BMDU) in the 4-hour p53 biomarker response is depicted with a large triangle datapoint fixed to the same value as the BMDL.
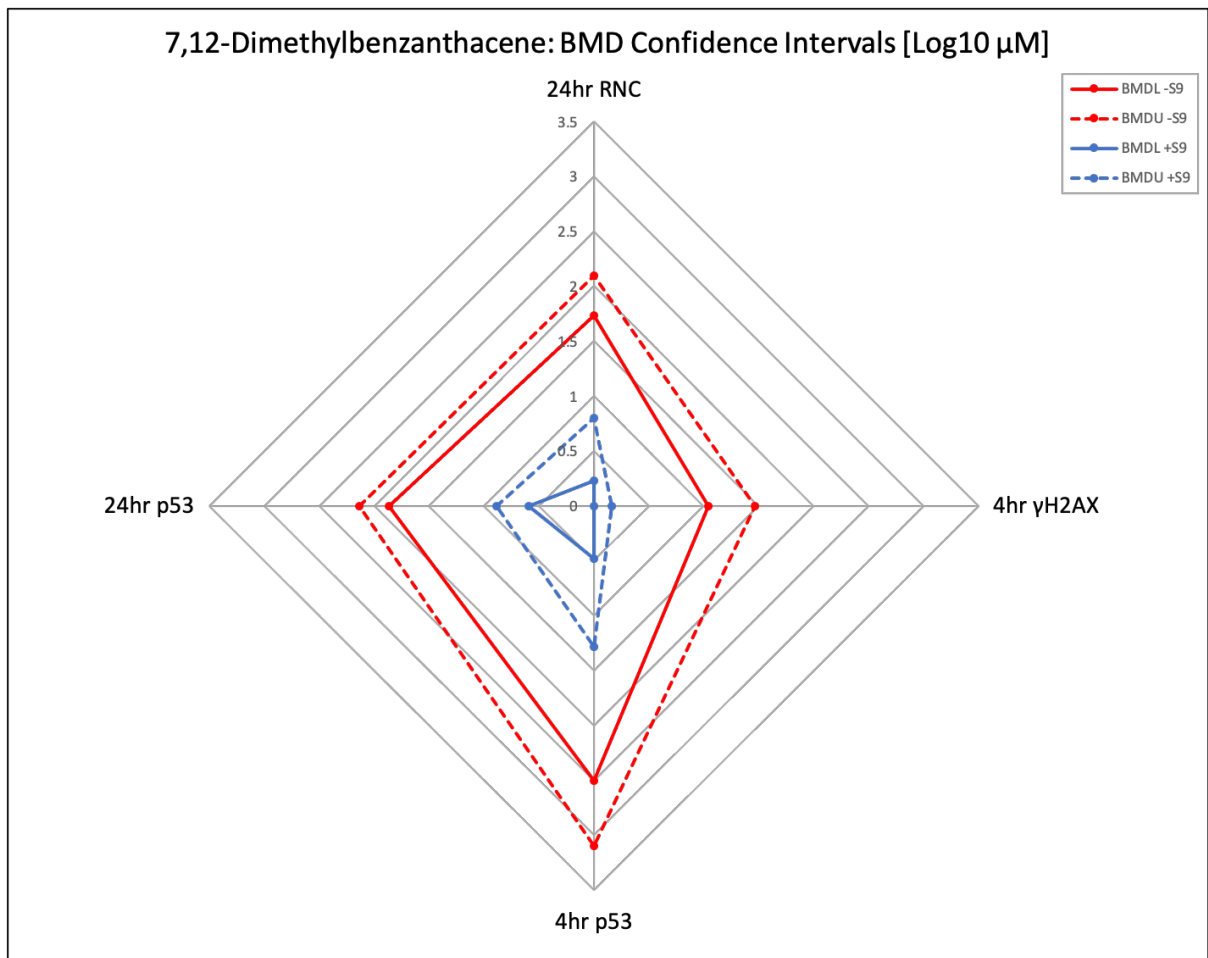
Figure 4.21. Radar plot for 7,12-Dimethylbenzanthracene. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale.
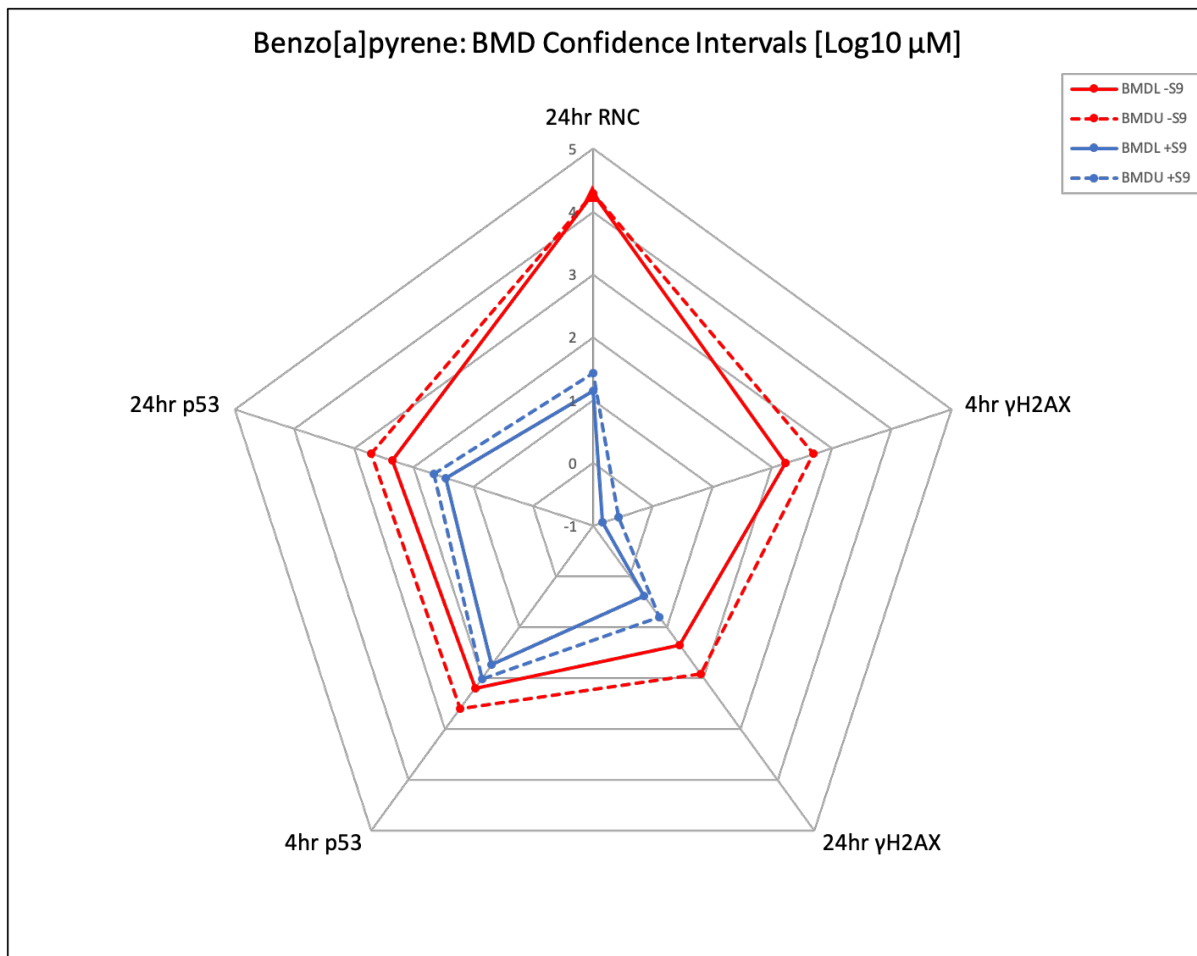
Figure 4.22. Radar plot for Benzo[a]pyrene. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. An unbound confidence interval limit (infinite, BMDU) in the 24-hour RNC biomarker response is depicted with a large triangle datapoint fixed to the same value as the BMDL.
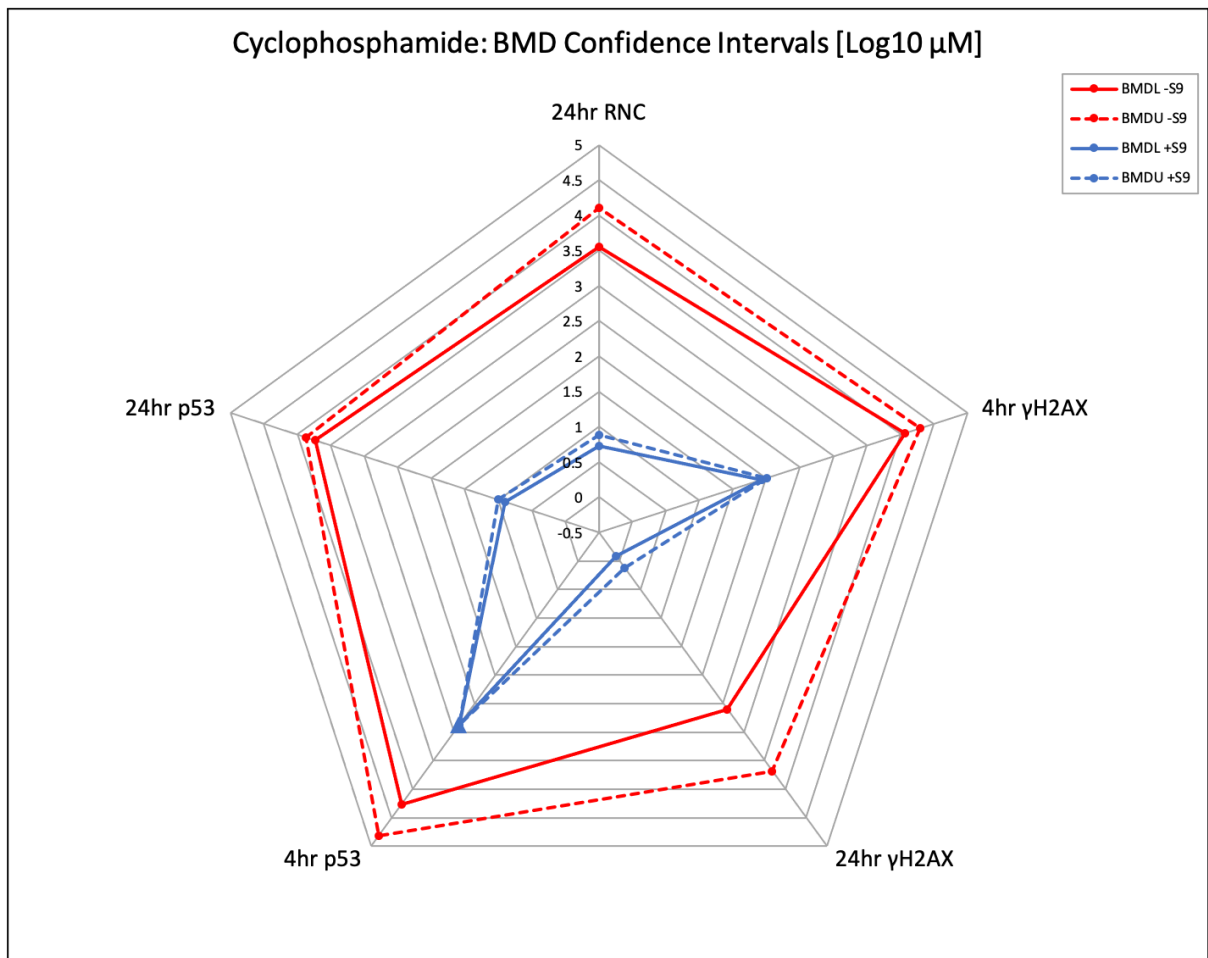
Figure 4.23. Radar plot for Cyclophosphamide. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. An unbound confidence interval limit (infinite, BMDU) in the 4-hour p53 biomarker response is depicted with a large triangle datapoint fixed to the same value as the BMDL.
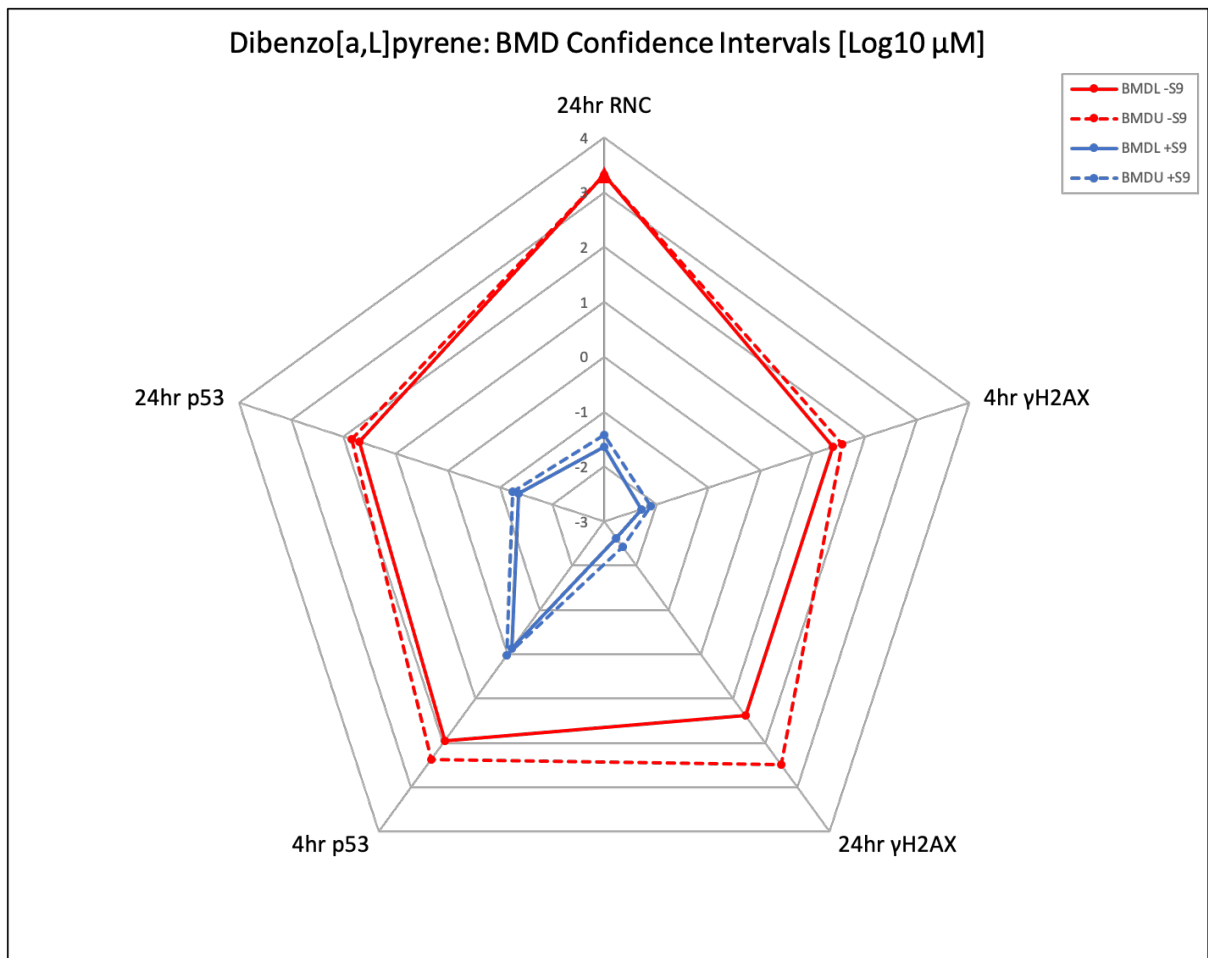
Figure 4.24. Radar plot Dibenzo[a,L]pyrene. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. An unbound confidence interval limit (infinite, BMDU) for the 24-hour RNC biomarker response is depicted with a large triangle datapoint fixed to the same value as the BMDL.
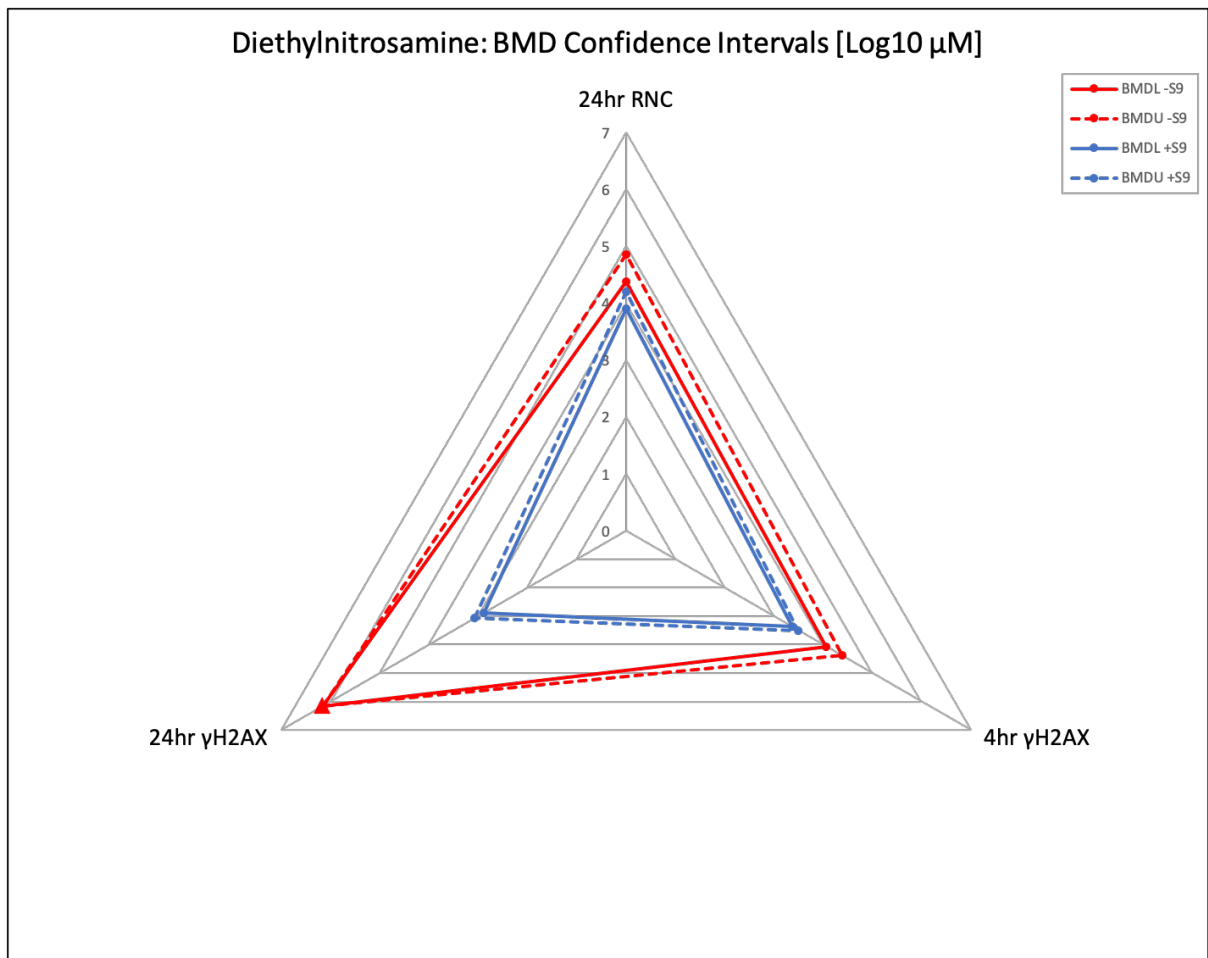
Figure 4.25. Radar plot for Diethylnitrosamine. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. An unbound confidence interval limit (infinite, BMDU) in the 24-hour γH2AX biomarker response is depicted with large triangle datapoint fixed to the same value as the BMDL.
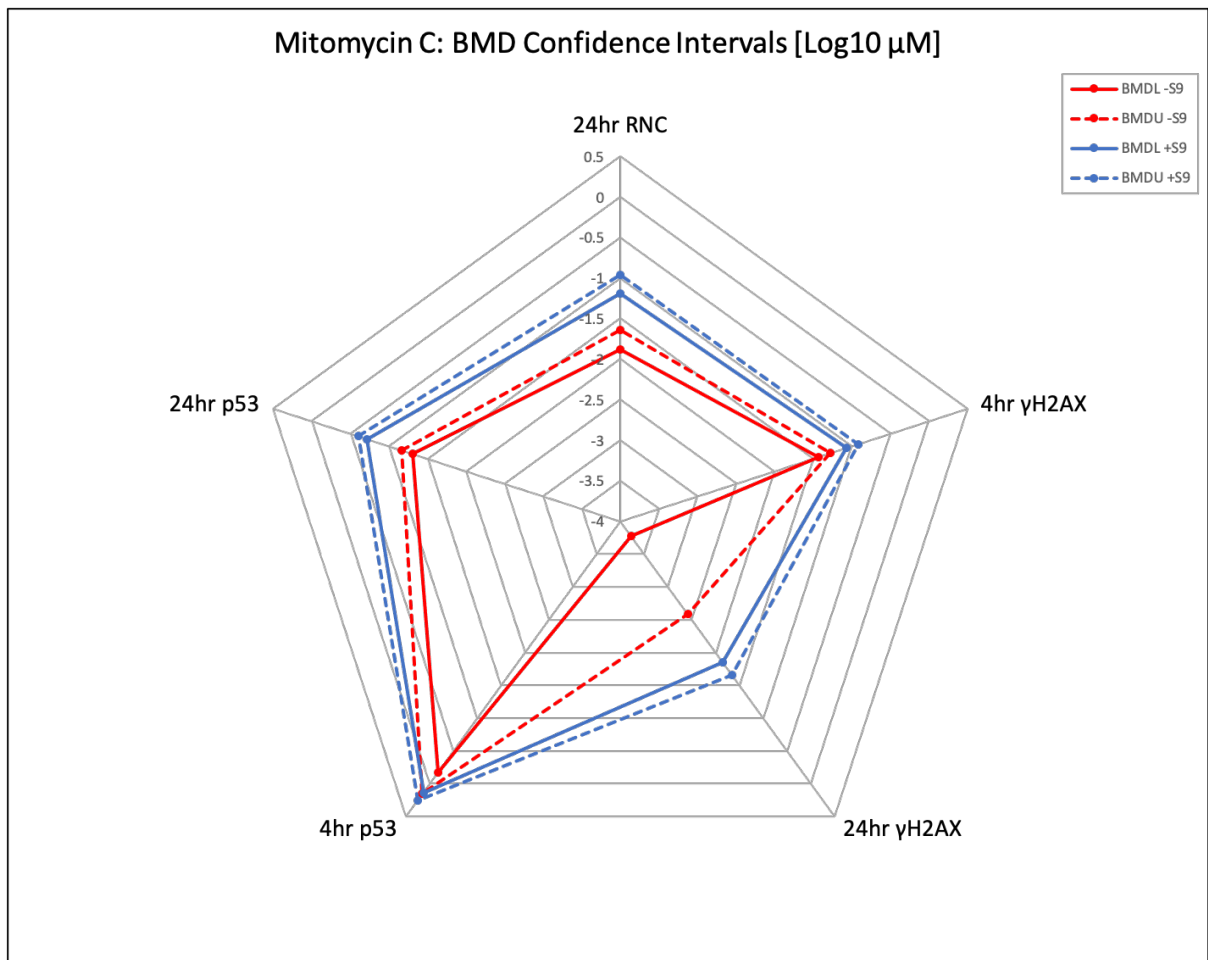
Figure 4.26. Radar plot for Mitomycin C. MultiFlow biomarkers depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 µM scale.
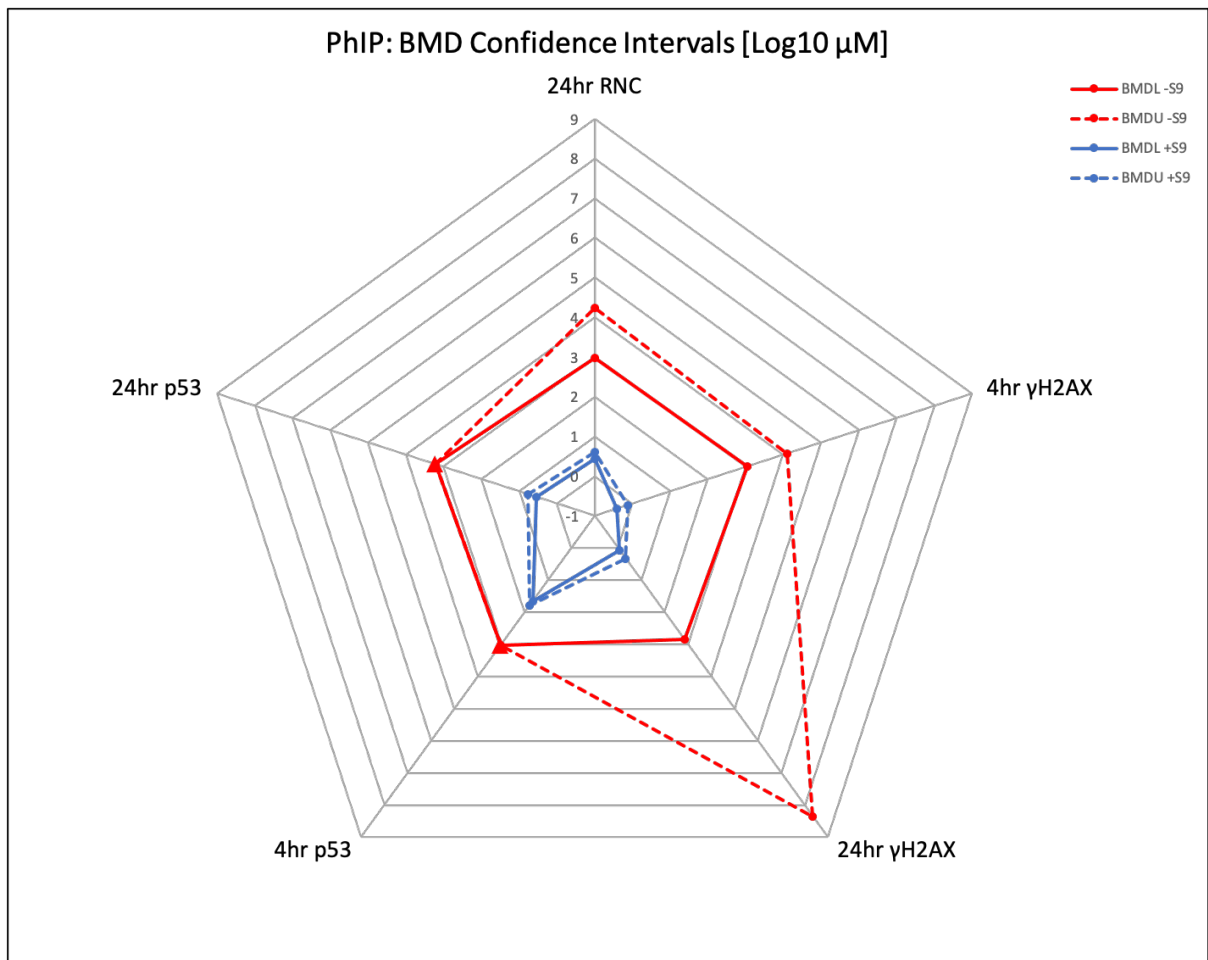
Figure 4.27. Radar plot for PhIP. MultiFlow biomarkers are depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 μM scale. Unbound confidence interval limits (infinite, BMDUs) for the 4-hour p53 and 24-hour p53 biomarkers are depicted with large triangle datapoints fixed to the same values as the BMDLs.
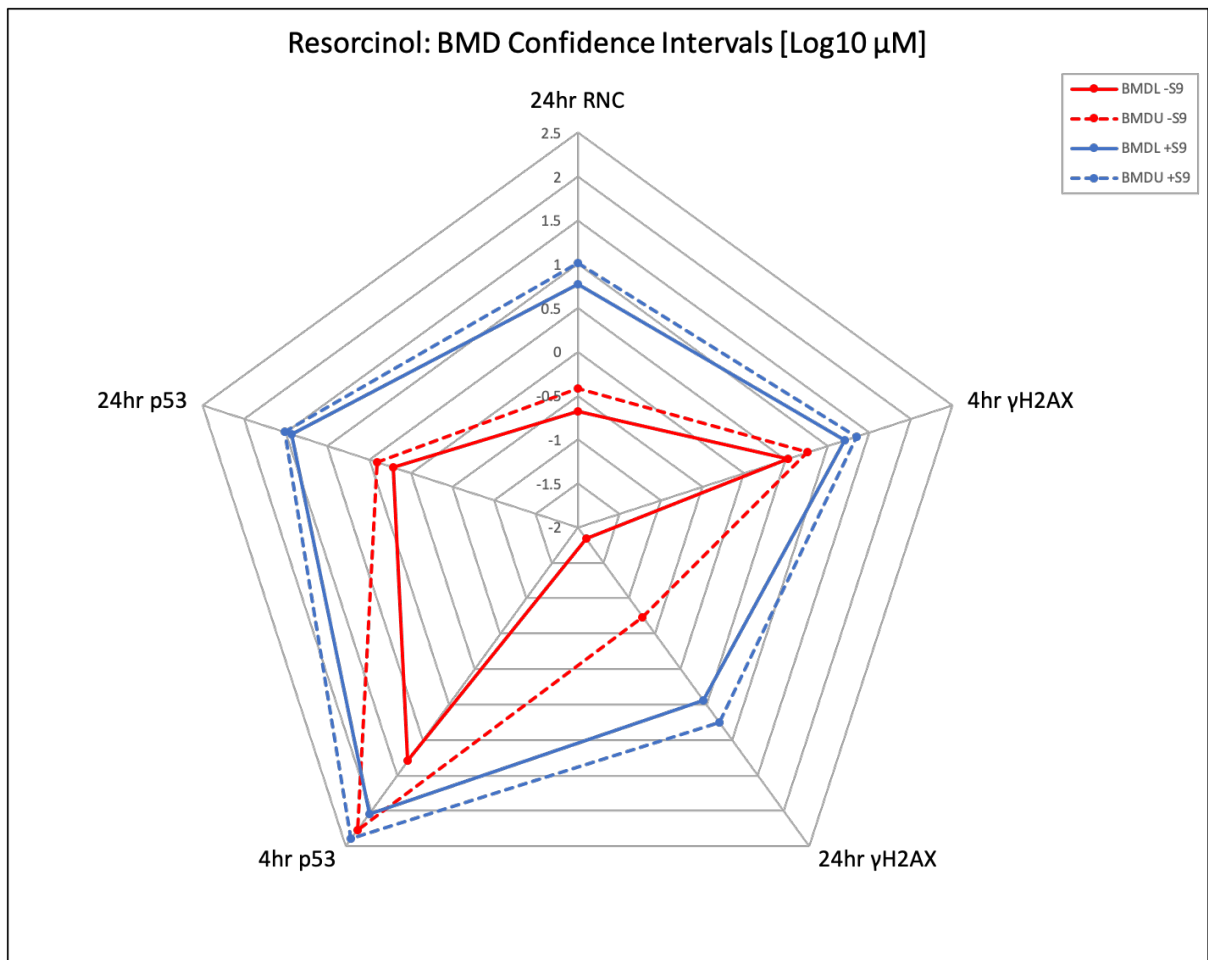
Figure 4.28. Radar plot for Resorcinol. MultiFlow biomarkers depicted on the Y axes. BMDL and BMDU values plotted on the X axes in a Log10 µM scale.

4.3.4 S9 Potency Ratio CIs

Tian et al., (2020), derived S9 potentiation ratios by comparative analysis of the BMD point estimates across the S9 exposure conditions for each compound and endpoint. There is a concern that the use of BMD point estimates can misrepresent the potency effect of S9 in the *in vitro* MultiFlow system that the authors examined since the BMD measure of uncertainty is disregarded.

The BMD CIs obtained in this re-analysis were used to calculate an 'S9 potency ratio CI' that is derived for each compound and endpoint BMD analysis that returned a dose response by comparing the BMDU +S9 to BMDL -S9, to the BMDL +S9 to BMDU -S9. The 'S9 potency ratio CI' is synonymous with the magnitude of the potency difference (range) exhibited for each compound/endpoint combination after exposure to S9. More context and

commentary on this concept are provided in the discussion section 4.4.4 of this chapter. The values are displayed in **Table 4.3** in comparison with the S9 potentiation ratios from Tian et al. (2020).

Table 4.3. Compound/Biomarker Combination BMD Ratios and 'S9 Potency Ratio CIs' (original and Log scale) compared with the Tian et al. (2020) 'S9 Potentiation Ratios'

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [µM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [µM]	S9 Potency Ratio CI Range [Log10 µM][3]
2-Acetylaminofluorine	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	NC	1.89	0	0
	4 hr γH2AX	>2.6	1.14	0	0
	24 hr γH2AX	>8.9	34.23	8.53 - 704.02	0.93 – 2.85
	24 hr RNC	0.53	0.33	0.18 - 0.55	-0.74 to -0.26
2-Aminoanthracene	4 hr p53	>7	No DR	No DR	No DR
	24 hr p53	>22	108.72	57.26 - 278.80	1.76 – 2.45
	4 hr γH2AX	167	74.55	46.03 - 126.81	1.66 – 2.10

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [μM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [μM]	S9 Potency Ratio CI Range [Log10 μM][3]
	24 hr γH2AX	16.3	8.95	2.51 - 21.59	0.40 – 1.33
	24 hr RNC	>26.2	295.11	79.87 - 3004.35	1.90 – 3.48
7,12-Dimethylbenzanthracene	4 hr p53	>32.9	49.90	16.63 - 418.60	1.22 – 2.62
	24 hr p53	23.7	18.06	9.52 - 34.87	0.98 – 1.54
	4 hr γH2AX	88.7	28.07	No DR	No DR
	24 hr γH2AX	NC	No DR	No DR	No DR
	24 hr RNC	21.8	24.43	8.59 - 73.37	0.93 – 1.87
Benzo[a]pyrene	4hr p53	NC	3.06	1.54 - 7.23	0.19 – 0.86
	24 hr p53	>80	9.14	4.98 - 17.88	0.70 – 1.25

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [μM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [μM]	S9 Potency Ratio CI Range [Log10 μM][3]
	4 hr γH2AX	>463	1357.44	629.63 - 3448.28	2.80 – 3.54
	24 hr γH2AX	1.8×10^6	11.18	3.47 - 34.73	0.54 – 1.54
	24 hr RNC	>31.4	No DR	No DR	No DR
	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	870	702.02	535.64 – 923.65	2.93 – 2.97
Cyclophosphamide	4 hr γH2AX	>101	158.03	115.69 – 227.54	2.06 – 2.36
	24 hr γH2AX	3249	1589.81	305.34 – 5922.33	2.48 – 3.77
	24 hr RNC	797	963.44	466.40 – 2405.30	2.67 – 3.38
Dibenzo[a,l]pyrene	4 hr p53	>56	148.74	88.94 – 316.87	1.95 – 2.50

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [μM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [μM]	S9 Potency Ratio CI Range [Log10 μM][3]
	24 hr p53	>962	1162.83	50.24 – 1578.48	1.70 – 3.20
	4 hr γH2AX	22,727	4666.876	24.49 – 6,967.37	1.39 – 3.84
	24 hr γH2AX	>38,462	22167.08	6325.46 – 130,672.27	3.80 – 5.12
	24 hr RNC	>1220	No DR	No DR	No DR
Diethylnitrosamine	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	NC	No DR	No DR	No DR
	4 hr γH2AX	>3.5	5.90	3.64 – 10.47	0.56 – 1.02
	24 hr γH2AX	>10	No DR	No DR	No DR
	24 hr RNC	NC	3.58	1.53 – 9.10	0.18 – 0.96

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [μM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [μM]	S9 Potency Ratio CI Range [Log10 μM][3]
Mitomycin C	4 hr p53	0.51	0.52	0.37 – 1.06	-0.43 – 0.03
	24 hr p53	0.73	0.27	0.20 – 0.36	-0.70 – -0.44
	4 hr γH2AX	0.82	0.44	0.31 – 0.62	-0.51 – -0.21
	24 hr γH2AX	0.25	0.07	0.01 – 0.19	-2.00 – -0.72
	24 hr RNC	0.31	0.21	0.12 – 0.36	-0.92 – -0.44
PhIP	4 hr p53	>3.1	No DR	No DR	No DR
	24 hr p53	>34.6	No DR	No DR	No DR
	4 hr γH2AX	>1633	4719.67	1436.10 – 33,246.07	3.16 – 4.52

Compound	Biomarker	S9 Potentiation Ratio from Tian et al. (2020) [μM]	BMD-BMD Ratio	S9 Potency Ratio CI Range [μM]	S9 Potency Ratio CI Range [Log10 μM][3]
	24 hr γH2AX	>781	2067.12	336.19 – 189,075,630.25	2.53 – 8.28
	24 hr RNC	>18.2	703.43	230.83 – 6142.75	2.36 – 3.79
Resorcinol	4 hr p53	NC	0.22	0	0
	24 hr p53	0.11	0.07	0.0498 – 0.0937	-1.30 – -1.03
	4 hr γH2AX	0.26	0.23	0.1502 – 0.3594	-0.82 – -0.44
	24 hr γH2AX	0.05	0.03	0.0026 – 0.0659	-2.59 – -1.18
	24 hr RNC	0.047	0.04	0.0202 – 0.0645	-1.69 – -1.19

4.3.5 Unsupervised Hierarchical Clustering Results

As previously mentioned, the 'S9 potency ratio CI' values span several orders of magnitude. To objectively define the range of 'S9 potency ratio CIs' obtained, the lower and upper values (Log scale) of each 'S9 potency ratio CI' were analyzed *via* unsupervised hierarchical clustering. The clustering was based on squared Euclidean distance 'Ward's method' (Ward and Hook 1963) between points. The resulting groups are presented in **Figure 4.29** in the form of a two-dimensional dendrogram with accompanying heat map. The upper and lower values of the 'S9 potency ratio CIs' associated with the *in vitro* biomarkers (4-hour p53, 24-hour p53, 4-hour γH2AX, 24-hour γH2AX, and 24-hour RNC) are displayed on the X axis. Compounds are plotted on the Y axis. The heat map represents the order of magnitude difference in 'S9 potency ratio CIs'. Compound/biomarker combinations where S9 exposure increased genotoxic potency are displayed with varying intensities of red on the heat map, whereas compounds and endpoints where S9 exposure decreased potency are displayed with varying intensities of blue on the heat map. Compound/biomarker combinations where the dose-response following S9 exposure was not statistically significantly different were plotted as zero values and displayed as grey blue on the heatmap. Additionally, the clustering platform could not accommodate missing values for compound/biomarker combinations that showed no dose-response. In these instances, a zero value was entered to accommodate the clustering method.

Figure 4.29. JMP software unsupervised hierarchical clustering results are shown as a two-dimensional dendrogram with heatmap for the 10 clastogens 'S9 potency ratio CIs'. The values of the lower and upper confidence limits of the 'S9 potency ratio CIs' were included in the analysis. The lower and upper bound of the 'S9 potency ratio S9 CIs' are plotted on the *X* axis per biomarker combination, and compounds on the *Y* axis. Increasing intensities of red indicate a strong tendency for S9 exposure to increase the genotoxic potency of a compound for a specific biomarker response. An increasing intensity of blue indicates the converse, where the presence of S9 decreases the genotoxic potency of a compound for a specific biomarker response. Grey blue represents compound/biomarker combinations where the dose-response following S9 exposure was not statistically significantly different (overlapping CIs), or where a zero value was included to accommodate the clustering method in instances where no S9 potency ratio CIs were obtained (infinite BMDUs). There are 4 distinct clades that group compounds into 1) high, 2) low, 3) zero, and 4) negative (as a subset of clade 3), effects on potency because of S9 exposure. Abbreviation: DMBA = 7,12-Dimethylbenzanthracene. Figure from Wheeldon et al. (2021).

4.4 Discussion

4.4.1 Problems Associated with Compound Subgrouping BMD Analyses

Whilst the confidence intervals calculated by either large group or small subgroup combined covariate analysis generally resided in the same order of magnitude, the general trend was that confidence intervals calculated from large group combined covariate analysis were more precise. This is consistent with the published literature that shows that increasing the number of levels in the covariate results in more precise confidence intervals (Wills et al., 2016a). Since the hypothesis under test presented in this chapter was to determine if a more statistically robust comparative potency metric could be developed, it was concluded that bias could not be ruled out by performing BMD analysis on the small compound subgroups. To this end, a more powerful computer (16GB RAM) had to be used to conduct the analyses with up to 10 compounds in the combined covariate.

4.4.2 BMD Confidence Interval Potency Plots

CIs that span a maximum of approximately 1-2 Log units (i.e., 1-2 orders of magnitude) are considered good quality and consistent with other BMD CI spans that are reported in the literature for other *in vitro* genotoxicity systems (Soeteman-Hernández et al. 2015; Bemis et al. 2016; Wills et al. 2016; Allemang et al. 2018). While most compounds' BMD analysis yielded tightly bound CIs, some displayed wide CIs or dose responses with unbound upper confidence limits. Specifically, the 4-hour p53 biomarker response for 2-acetylaminofluorene -S9 yielded an infinite BMDU (BMD 3.88 Log10 μM, BMDL 3.42 Log10 μM, BMDU Infinite Log10 μM) (**Figure 4.9**). Secondly, both the 4-hour and 24-hour p53 (**Figure 4.17**) biomarker responses for PhIP -S9 yielded unbound BMDUs (BMD 4.39 Log10 μM, BMDL 3.03 Log10 μM, BMDU Infinite Log10 μM; BMD 5.20 Log10 μM, BMDL 3.22 Log10 μM, BMDU Infinite Log10 μM, respectively). The 4-hour γH2AX biomarker response for 7,12-Dimethylbenzanthracene returned a zero BMDL value indicating that the lower bound of the BMD is not significant from zero (**Figure 4.11**). These limited instances where dose-response analysis yielded an infinite BMDU are considered to have exhibited no evidence of a dose-response relationship, and hence are unsuitable for drawing statistically robust conclusions from the analyses. In the single instance of a zero BMDL, the resulting BMDL to BMDU ratio is unobtainable (problems associated with

division by 0), and hence justifies omission of this compound/biomarker combination for further analysis.

In addition to having unbound BMDUs, some compounds also displayed BMD estimates that are grossly disproportionate to the BMDs from other biomarkers or S9 condition for the same compound. The BMD CIs for these instances were not plotted, since the relative disproportionality would make the other CIs appear comparatively small. These instances are indicated in the comparative potency plots with either dashed lines spanning the entire graph range, or the BMD restricted and only displaying the BMDL. Specifically, the 24-hour RNC biomarker response -S9 for benzo[a]pyrene [BMD 12.33 Log10 μM] (**Figure 4.12** dashed line), the 24-hour RNC biomarker response -S9 for dibenzo[a,l]pyrene [BMD 11.99 Log10 μM] (**Figure 4.14** BMD restricted and only displaying the BMDL), and the 24-hour γH2AX biomarker response -S9 for diethylnitrosamine [BMD 20.74 Log10 μM] (**Figure 4.15** dashed line). These biomarker responses also yielded infinite BMDUs, and consequently displayed no evidence of a dose response upon which one can then accurately define potency.

Consistent with the approach advocated for in the scientific literature, visual scrutiny of each compound's BMD potency plot allows efficient visual discrimination of the potency difference of each compound/biomarker combination that results from S9 exposure. Compound/biomarker combinations where the S9 exposure condition yielded BMD CIs that largely overlap results in potency differences that are statistically insignificant. Conversely, non-overlapping CIs are statistically significant and show that the S9 condition exerts an impact on the compound's potency in the same *in vitro* biomarker response.

For the most part, as one would expect, the tested clastogens that require metabolic activation to exert genotoxic effects show an increase in potency when exposed to the low concentration S9 system. This is evidenced by +S9 CIs that reside in more potent regions of the comparative potency plot by several orders of magnitude compared to -S9 CIs. An outlier is 2-acetylaminofluorene where all biomarker responses except for 24-hour γH2AX and 24 -hour RNC display CIs that reside in the same order of magnitude (**Figure 4.9**). This suggests that S9's potentiating effect was restricted to these endpoint/timepoint combinations. The comparative potency plots also show that the direct acting clastogens, mitomycin C and resorcinol, were not potentiated by S9 – rather, to a certain extent –, the presence of S9

slightly decreased their genotoxic potency (**Figure 4.16** and **Figure 4.18**). Both compounds do not require metabolic activation to exert genotoxic effects, and the decrease in potency in the presence of S9 may be due to cytotoxic effects from the use of S9.

4.4.3 BMD Confidence Interval Radar Plots

BMD metrics (BMDL and BMDU) are presented in radar plots to show the effect of S9 across the test compounds that elicited a relevant dose-response across each endpoint. This approach offers a simple and immediate visual interpretation of the effect on potency of S9 across the biomarkers tested in the MultiFlow assay, in a similar way to how Dertinger et al. (2019) uses radar plots to show the magnitude of effects in responsive MultiFlow biomarkers. The radar plots include the relevant biomarker responses on the Y axes with the BMDL and BMDU values plotted on the X axes in a Log10 μM scale. It is not possible to plot an unbound (infinite) upper confidence interval (BMDU) on the radar plot. This issue was handled by restricting the BMDU to the same value of the BMDL and highlighting this with a large triangular datapoint. Notice how these data points cause a distinct skew comparative to the trend observed in the bound BMD_U data points. This observation emphasises that the use of BMD point values does not accurately represent the range of potency effects in a dataset.

Two groups can be deciphered from the radar plots that are consistent with the test compound mode of action. The clastogens overall show a remarkable difference in the genotoxic potency of the test compound in the presence of S9 compared to its absence. This is evident by the distinct shift to higher potency (more negative values in the Log scale) by the blue lines (BMDL/U +S9) on the radar plots in **Figure 4.19** to **Figure 4.28**. The outliers include mitomycin C (**Figure 4.26**) and resorcinol (**Figure 4.28**). As opposed to the other clastogens, mitomycin c and resorcinol were less potent in the presence of S9 as evident by the blue lines (BMDL/U +S9) residing in the more positive region of the Log scales.

The plotting of the BMD confidence intervals in radar plots is novel and there are no published studies reporting the same method. The BMD confidence interval plots were presented during manuscript peer review with 2 reviewers in favour, and 1 reviewer vehemently opposed for reasons related to the historic precedent of displaying BMD confidence intervals in standard potency plots (like those in **Figure 4.9** to **Figure 4.18**). Whist one understands that most reviewers are used to seeing the standard potency plots, the

radar plots offer a nuanced method to viewing comparative potency metrics which some genetic toxicologists could become accustomed to.

4.4.4 S9 Potency Ratio CIs

As previously highlighted, Tian et al., (2020) derived S9 potentiation ratios by comparative analysis of the BMD point estimates across the S9 exposure conditions for each compound and endpoint. There is a concern that the use of BMD point estimates can misrepresent the potency effect of S9 in this *in vitro* system since the BMD measure of uncertainty is disregarded.

Tian et al. (2020), correctly pointed out that the BMD estimate can be beyond the top concentration used in the benchtop experiment. In these instances, the authors calculated the 'S9 potentiation ratio' by restricting the BMD to the top concentration (denoted with a greater than (>) symbol in the tabulated results section of their article). However, these tight BMD restrictions led to mischaracterization of the true BMD ratio. It is contended that one should only limit BMD values for graphical display where the value is disproportionately larger than other endpoints and S9 conditions for the same compound. In the BMD analysis reported in this re-analysis, disproportionately high BMD values also coincided with unbound BMDUs.

The BMD CIs obtained in this re-analysis were used to calculate an 'S9 potency ratio CI' that is derived for each compound and endpoint BMD analysis that returned a dose response by comparing the BMDU +S9 to BMDL -S9, to the BMDL +S9 to BMDU -S9. The 'S9 potency ratio CI' is synonymous with the magnitude of the potency difference (range) exhibited for each compound/endpoint combination after exposure to S9. Although there were only 2 experimental conditions (+/- S9) included here, the same algorithm could apply to 2 or more experimental conditions where one potency is compared to another potency of interest. The 'S9 potency ratio CIs' were calculated in the original scale so that the values can be compared to the 'S9 potentiation ratios' values obtained from BMD-BMD ratios by Tian et al. (2020). Values greater than 1 represent increased compound potency for a particular endpoint after S9 exposure, whilst values less than 1 confer the inverse. 0 values represent CIs that overlap and hence potency differences are statistically indefensible. The BMD-BMD ratios from this analysis are also shown for comparative purposes. The S9 potency ratios in the Log scale were also displayed for the sole purpose of aiding in demonstration of

differences in orders of magnitude. The values were displayed in **Table 4.3** in comparison with the S9 potentiation ratios from Tian et al. (2020).

Comparing the 'S9 potency ratio CIs' with the S9 potentiation ratio obtained by Tian et al. (2020) shows that in almost all cases, the 'S9 potentiation ratio' that was derived from BMD point estimates is either mischaracterized through tight restrictions on the BMD (where the researchers denoted these ratios with a > sign) or is in the upper range of the 'S9 potency ratio CI' range, indicating a tendency for over estimation of the S9 bioactivation effect.

Calculating an 'S9 potency ratio CI' for these experiments accurately conveys the uncertainty measurements which should be accounted for when assessing potency comparisons across conditions. The 'S9 potency ratio CIs' calculated here can be summarised by stating that S9 increases the potency of the compounds ranging from approximately 1-2 orders of magnitude (Log10 µM) for lower potency differences, to approximately 3-5 orders of magnitude (Log10 µM) for higher potency differences. The effect of S9 is statistically insignificant for some compound/endpoint combinations where BMD CIs overlap. In other instances, S9 exposure decreased potency (expressed as negative values) of compounds by approximately 0.4 to 2.0 orders of magnitude.

4.4.5 S9 Potency Ratio CI Range

As previously mentioned, the 'S9 potency ratio CI' values span several orders of magnitude. To objectively define the range of 'S9 potency ratio CIs' obtained, the lower and upper values (Log scale) of each 'S9 potency ratio CI' were analysed *via* unsupervised hierarchical clustering. The clustering was based on squared Euclidean distance 'Ward's method' (Ward and Hook 1963) between points. The resulting groups are presented in **Figure 4.29** in the form of a two-dimensional dendrogram with accompanying heat map. The upper and lower values of the 'S9 potency ratio CIs' associated with the *in vitro* biomarkers (4hour p53, 24-hour p53, 4-hour γH2AX, 24-hour γH2AX, and 24-hour RNC) are displayed on the X axis. Compounds are plotted on the Y axis.

The heat map represents the order of magnitude difference in 'S9 potency ratio CIs'. Compound/biomarker combinations where S9 exposure increased genotoxic potency are displayed with varying intensities of red on the heat map, whereas compounds and biomarkers where S9 exposure decreased potency are displayed with varying intensities of

blue on the heat map. Compound/biomarker combinations where the dose-response following S9 exposure was not statistically significantly different were plotted as zero values and displayed as grey blue on the heatmap. Additionally, the clustering platform could not accommodate missing values for compound/biomarker combinations that showed no dose-response. In these instances, a zero value was entered to accommodate the clustering method.

There are 4 distinct clades identified in the dendrogram: 1) Chemicals whose genotoxic potency was *dramatically* increased in the presence of S9 (dibenzo[a,l]pyrene, cyclophosphamide, 2-aminoanthracene, and PhIP); 2) chemicals whose genotoxic potency was increased in the presence of S9 (dimethylbenzanthracene, benzo[a]pyrene, and 2-acetylaminofluorene); 3) chemicals whose genotoxic potency was not increased in the presence of S9 (mitomycin c, resorcinol, and diethylnitrosamine); and 4) a subset of clade 3, clade 4, whose genotoxic potency was reduced in the presence of S9 (mitomycin c, and resorcinol; indicated by a strong dark blue in the heat map).

The clustering results demonstrate that distinct groups exist that best describe the range of S9-dependent effects into high, low, zero and negative categories. The division of the S9 potency ratio CIs into groups follows application of quantitative metrics allowing objective interpretation in a hazard-based scale context. It is likely that when a more diverse set of clastogens are analysed, it will be possible to detect further subdivisions of potency effect (e.g., low, medium, high, very high). The same clustering method could be applied to other dose-response comparative analyses. Several examples are envisioned, for instance the study of compound potency effects between different exposure durations (subchronic vs. chronic dosing regimens), between different target tissues (liver vs. bone marrow), or between different *in vitro* cell lines (liver HepaRG vs. HepG2, or V79 versus CHO).

4.5 Concluding Remarks

The results of the data analysis presented in this results chapter illustrate the necessity to utilise the full BMD CIs to draw conclusions from dose-response datasets. It has been demonstrated through re-assessment of the previously published MultiFlow dataset (Tian et al., 2020) that in comparative potency analysis, the use of BMD point estimates can yield mischaracterised or over-estimated potency ratios during instances where the width of the CIs varies considerably between conditions.

Interpretation of BMD CIs relies upon visual scrutiny by the genetic toxicologist, and it is speculated that one could be overwhelmed with the results from hundreds of compounds from *in vitro* screening experiments. Therefore, the genetic toxicologist may wish to use effective statistical methods such as hierarchical clustering to aid in data interpretation. When objectively considering the range of 'S9 potency ratio CIs' presented in the analyses presented in this results chapter, one is provided with a scale of high to low, zero, and negative S9 potency effects. While the potency effect of S9 has been illustrated in the *in vitro* MultiFlow genotoxicity test system, the same approach can be applied to any comparative genotoxicity potency analysis. It is proposed that investigators apply the following stepwise approach when performing comparative potency analysis of 2 or more experimental conditions using the BMD methodology:

- Utilise the results of combined covariate BMD analyses to plot BMD CIs that represent potency across the endpoints evaluated under different dependent variables. For graphical purposes only, limit disproportionately high BMDs that display infinite BMDU values.
- Derive the lower and upper values of 'potency ratio CIs' between conditions by comparing BMDU condition 1 to BMDL condition 2, and BMDL condition 1 to BMDU condition 2 (ad infinitum), respectively.
- Objectively identify groups in the data that best describe the magnitude of the difference in potency observed between conditions. Whilst a genetic toxicologist can successfully identify patterns in CI plots of a few experimental conditions (a small number of compounds, animal sex, limited number of cell lines, etc.) *via* visual techniques, it is contended that hierarchical clustering may add significant value to the interpretation of large datasets, particularly those from *in vitro* screening experiments of large numbers of compounds and biomarker combinations. Whilst this was shown in radar plots and hierarchical clustering techniques in this chapter, visual techniques could take the form of the novel ToxPi program graphical outputs (https://toxpi.org/).

There were several challenges that had to be overcome in this phase of the research project, namely:

- Overcoming the obstacle associated with computer low processing power and addressing this with the most scientifically robust solution. The resolution was using a higher-powered computer since the introduction of bias could not be ruled out.

- Influencing one peer reviewer in the submission of the manuscript that the comparative potency plots were best kept in compound specific format so that the side-by-side comparison of biomarker responses could be visualised.

Ultimately one of the peer-reviewers stated that "the clear guidance on data consolidation, BMD analysis, and subdivision into few categories is exemplary" which attests to the value that this work brings to the literature.

Chapter 5. *In Vitro* Mechanistic Genotoxicity and Potency Determinations for the Topoisomerase II Inhibitor Class of Compounds.

5.1 Introduction

Previous results chapters in this thesis demonstrated the general amenability of the MultiFlow biomarker responses to BMD modelling for compound potency ranking beyond hazard characterisation. Reported in this results chapter, two hypotheses were developed. The primary hypothesis was whether the multiplexed MultiFlow and Toxys ToxTracker biomarker responses could provide robust potency ranks within a specific compound mode of action (MOA). The secondary hypothesis was inspired by the work of Wills et al. (2016a) who generated robust potency ranks from *in vitro* micronucleus data for the benzimidazole class of compounds where the potency rank order was accounted for with knowledge of the compound structures and mechanisms of action. To this end, the Topoisomerase-II (Topo-II) Inhibitor class of compounds was selected for the specific case study reported in this results chapter to test the two hypotheses developed in this chapter[7].

5.1.1 Topo-II Inhibitors

Deweese and Osheroff (2008) explain that each human cell contains approximately 2 m of DNA that is condensed into a nucleus that is approximately 10 μm in diameter, and that accessing the genetic material is a topological challenge since it is anchored to the chromosome scaffold with the two strands of the double helix being supercoiled. The DNA's topological properties can only be altered when the double helix is broken. Two aspects of DNA topology affect nuclear processes. Firstly, DNA strand separation is required to overcome the negatively supercoiled nature of the DNA double helix. The linear movement of tracking enzymes such as helicases and polymerases causes compression of the double helix into a shorter region, leading to increasingly overwound sections in the DNA making it

---

[7] The data presented in this results chapter was published in "Benchmark Dose Analysis of DNA Damage Biomarker Responses Provides Compound Potency and Adverse Outcome Pathway Information for the Topoisomerase II Inhibitor Class of Compounds" (Wheeldon et al., 2020) and some passages in this chapter have been quoted verbatim from the same article. A copy of the research article is available in Appendix 3. The data was also presented at the 2019 Genetic Toxicology Association annual meeting in Delaware USA and won best poster award. A copy of the poster is available in Appendix 4.

more difficult to open the two strands of the DNA double helix. Secondly, intramolecular knots and intermolecular tangles that are produced during DNA replication need to be resolved to prevent cell lethality and to prepare the genetic material for the next cycle of mitosis or meiosis. **Figure 5.1** illustrates the supercoiling of DNA ahead of DNA tracking enzymes.
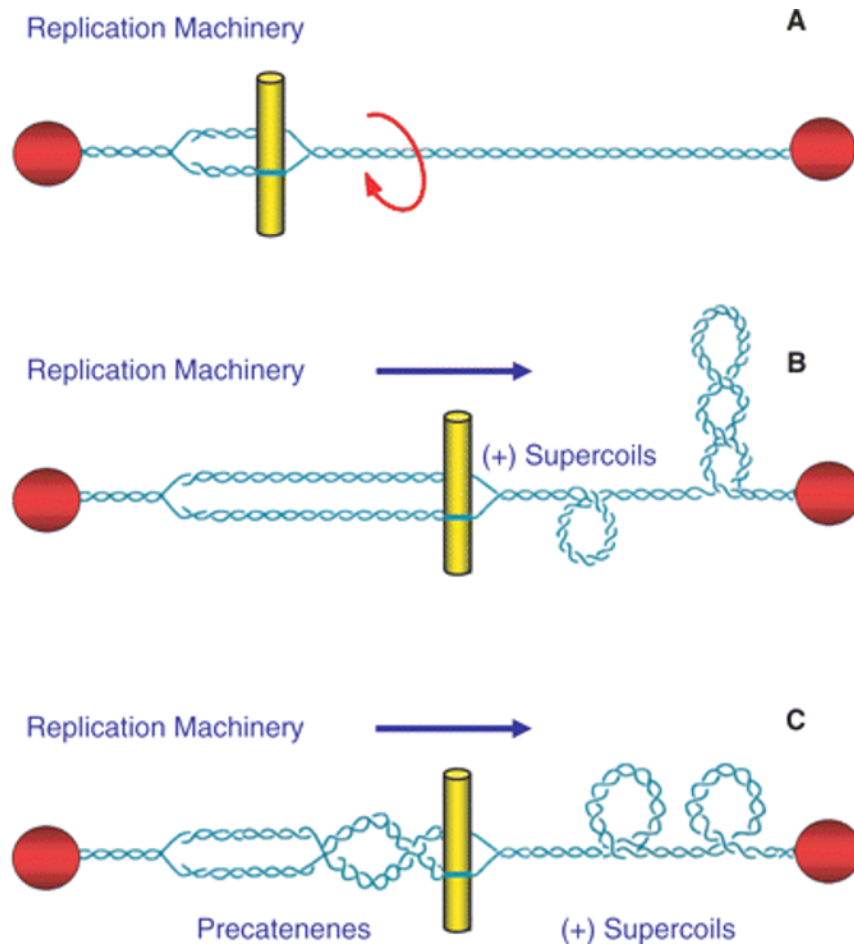
Figure 5.1. Schematic showing: A) Chromosomal DNA tethered to membranes or chromosome scaffold (red spheres) and unable to rotate; B) The yellow bars represent replication machinery which traverse along DNA causing compression of the DNA into small segments resulting in acute overwinding (positive supercoiling); and C) Precatenanes include translated torsional stress from the pre-replicated DNA which must be resolved. Figure adapted from Deweese and Osheroff (2008).

Type II topoisomerases are abundant in prokaryotic and eukaryotic cells (Holm et al., 1985; Earnshaw and Heck, 1985; Earnshaw et al., 1985; Dal Ben et al., 2007) and regulate the topological state of DNA during replication, transcription, and repair (Isaacs et al., 1998;

Nitiss, 2009). The enzyme covalently binds to DNA during replication and characterised by the enzymes catalytic cycle, an ATP-coupled reaction cleaves a temporary double-strand break in the DNA's backbone, thus creating a transient cleavage complex through which another DNA duplex can pass (Deweese and Osheroff, 2008). Strand passage is followed by subsequent DNA break ligation restoring the DNA structure. Therefore, the series of various DNA cleavage-religation reactions serve to resolve the torsional stress associated with DNA's super helical nature (Berger, 1998) explained above and in **Figure 5.1**.

Proliferating cells cannot survive without the function of the Topo-II enzyme, however the cleavage-complex associated DNA double strand breaks can be potentially deleterious. Deweese and Osheroff (2008) eloquently described Topo-II to assume a Dr. Jekyll/Mr. Hyde character where the enzyme has a dual persona; on the one hand being essential to cell viability, but conversely has the capacity to fragment the genome. Levels of cleavage complexes are tolerated *in vivo* as readily reversible intermediates. However, when levels of the enzyme drop below a critical concentration, daughter chromosomes can remain entangled following replication resulting in cell death from mitotic arrest (Ross et al., 1984; Deweese and Osheroff, 2008; Nitiss, 2009). Conversely, cleavage complex stabilisation intermediates can block DNA replication machinery such as replication forks, transcription complexes or DNA tracking enzymes. The resulting disrupted cleavage complexes can lead to conversion of the transient DNA strand breaks to permanent DNA double-stranded breaks. An accumulation of DNA breaks can result in potential chromosome aberrations and other DNA damaging effects. Other cellular events including apoptotic pathways (Solary et al., 1994) can be initiated in the amount of DNA breaks is overwhelming. **Figure 5.2** from Deweese and Osheroff (2008) summarises the effects when the Topo-II mediated critical balance is offset.
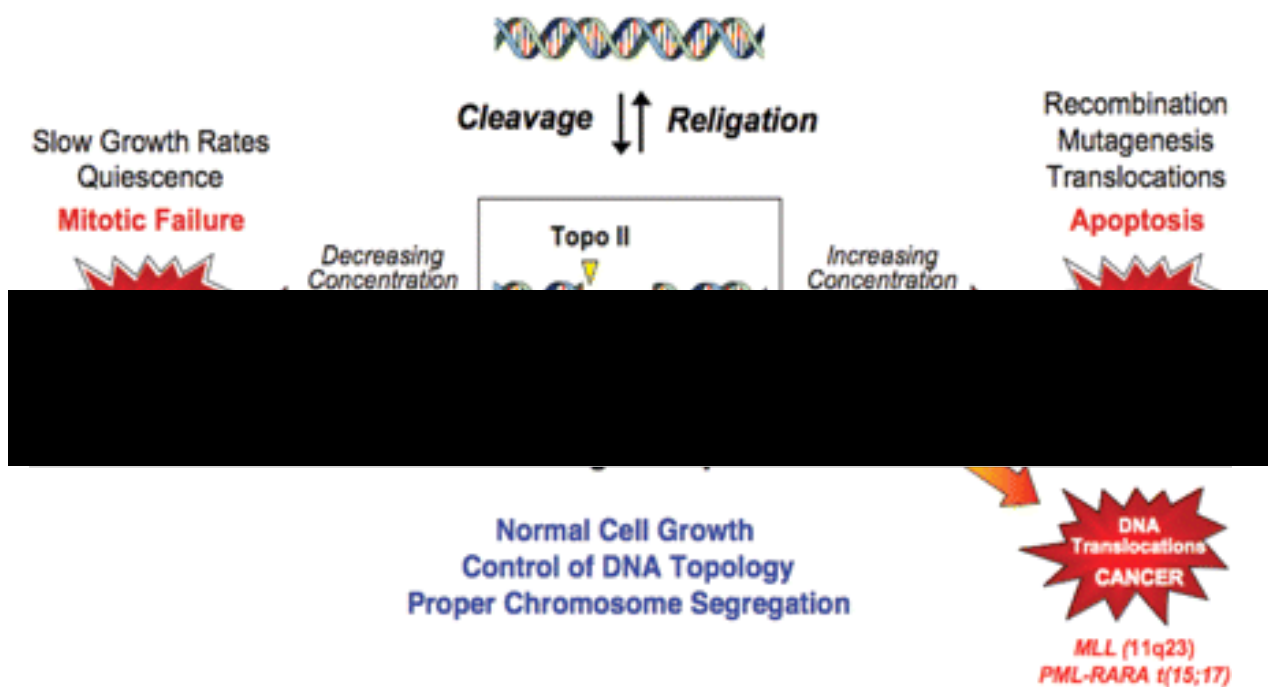
Figure 5.2. Schematic showing the effects if the level of Topo-II complexes become too low (left arrow) resulting in mitotic failure, and too high (right arrow) leading to apoptosis or permanent rearrangements of DNA in the form of translocations which can lead to specific cancers. Figure taken from Deweese and Osheroff (2008).

In scope of the genotoxicity risk assessment theme central to this thesis, a particular class of compounds of interest are the Topoisomerase-II (Topo-II) Inhibitors. This specific class of compounds were selected since they are potent genotoxicants with a pharmaceutical mechanism of action of interest (Hande, 1998) that target the Topo-II enzyme. Numerous compounds are known to perturb the DNA cleavage-religation action of Topo-II (Burden and Osheroff, 1998) and can be separated into two categories. Compounds belonging to the first category are known as catalytic inhibitors and tend to lower the overall activity of the Topo-II enzyme (Deweese and Osheroff, 2008) by hijacking the binding of the Topo-II enzyme to DNA or by inhibiting ATP action (Larsen et al., 2003). The second category includes compounds that increase levels of Topo-II-DNA cleavage complexes (Nelson and Kastan, 1994) that tend to 'poison' the Topo-II enzyme and convert it to a cellular toxin (Deweese and Osheroff, 2008). For this reason, these specific compounds are referred to as Topo-II Poisons with a distinguished mechanism of action different to the Topo-II Catalytic Inhibitors.

Many of the frontline chemotherapeutic agents used in the treatment of human systemic and solid tumours such as etoposide, the most active broad-spectrum antibiotics such as ciprofloxacin, and naturally occurring dietary isoflavonoid compounds such as genistein, target the Topo-II enzyme as Topo-II Poisons. Topo-II Poisons act via two different pathways by either inhibiting the ability of the enzyme to ligate cleaved DNA molecules, or by enhancing the forward rate of base scission. There are two mechanisms in which the Topo-II Poisons interact with the Topo-II-DNA complex, *via* forming noncovalent interactions at the Topo-II-DNA interface, or in a redox-dependent manner to covalently adduct to the enzyme at amino acid residues outside the active site (Wilstermann and Osheroff, 2003; McClendon and Osheroff, 2007; Deweese and Osheroff, 2008). The focus of this chapter is limited to Topo-II Poisons.

A particular Topo-II Poison and its exposure impact to human health includes the well-studied Etoposide compound, which is a podophyllotoxin toxin derivative from the American Mayapple (Hande, 1998). The compound was first synthesised in 1966 and approved for use in cancer therapy by the US FDA in 1983 (Montecucco et al., 2015). Epidemiological studies performed in the 1980s and 1990s demonstrated that patients treated with high dose etoposide in their chemotherapy regimen for metastatic germ cell tumours were at a significantly higher risk of developing secondary acute myeloid leukaemia in the years following initial etoposide exposure (Winick et al., 1993; Kollmannsberger et al., 1998; Felix, 2001). It was later shown that low dose etoposide exposure leads to MLL gene rearrangements *in vitro* in mouse embryonic stem cells, foetal liver-derived CD34+ hematopoietic stem cells (HSCs), and cord blood-derived CD34+ HSCs. The MLL gene rearrangements lead to chromosomal abnormalities and genomic instability (Baldwin and Osheroff, 2005; Bueno et al., 2009). Therefore, it is important to understand the dose-response relationship and potential adverse health outcomes that could result from Topo-II Poison exposure.

5.1.2 Genotoxic Concentration-Response Studies Topo-II Poisons

Some recent studies have assessed the concentration-response profile of some Topo-II Poisons in genotoxicity test systems. A study by Lynch et al. (2003) used micronucleus frequency dose-response data from mouse lymphoma cells exposed to varying concentrations of the Topo-II Poisons etoposide, doxorubicin, genistein and ciprofloxacin. The researchers

applied the broken stick model to demonstrate that a threshold dose-response existed for the tested Topo-II Poisons in the micronucleus frequency datasets.

A more recent study by Gollapudi et al. (2019) aimed to investigate the *in vitro* concentration-response relationships for the Topo-II Poison etoposide by performing BMD analysis of a flow-cytometric TK6 cell micronucleus dataset. The researchers found that increasing concentrations of etoposide exposure vastly increased micronucleus frequency compared to control in which they were able to derive BMD metrics from the dose-responses. The results with respect to etoposide genotoxic potency published by Gollapudi et al. (2019) are similar to the results obtained from genotoxic potency ranking of MicroFlow micronucleus frequency data presented in Chapter 3 of this thesis and published in Dertinger et al. (2019) where etoposide was one of the training set compounds.

One of the shortcomings of using micronucleus frequency data for Topo-II Poison potency ranking is that the frequency endpoint and dose-response relationships are representative of other clastogens which does not discriminate the mode of action from other genotoxic clastogens. An important aspect of indirect induced DNA damage from Topo-II mediated genotoxicity is the concept of a threshold of clastogenicity (Müller and Kasper, 2000). So, to add mechanistic information to the Topo-II mediated induced clastogenicity weight of evidence, one of the advantages of the Lynch et al. (2003) study was the inclusion of the TARDIS (trapped in agarose DNA immunostaining) assay which the authors explain is an immunohistochemical method that localises Topo-II-DNA *in situ* with the nucleus using specific antibodies. In doing so, a relative measure is provided of stabilised cleavage complex formation that can help to differentiate Topo-II mediated clastogenicity from other clastogenic mechanisms. However, performing multiple assays in parallel increases the time and resource investment needed to characterise a compound's mechanism of action. To this end, the advent of high throughput multiplexed assays may provide mechanistic information to add to the weight of evidence supporting a compound's hypothesised mechanism of action in a multi-compound assessment.

5.1.3 Primary Hypothesis: High-Throughput Mechanistic Genotoxicity Assays for Robust Topo-II Poison Comparative Genotoxic Potency Analysis

5.1.3.1 MultiFlow and ToxTracker Assay Mechanistic Information

Based on the known mechanism of action of the Topo-II Poisons in which DNA double strand breaks precede mitotic arrest, one would expect signals to be responsive in the main *in vitro* high-throughput genotoxicity test systems that are the focus of this thesis. For instance, the MultiFlow assays γH2AX biomarker should be responsive to Topo-II Poison mediated DNA double strand breaks. Replication fork collapse would lead to delays in cell cycle progression and associated nuclear translocation of p53 (Yoon et al., 2004), hence one would expect that the MultiFlow assay's p53 biomarker signals should increase with exposure to Topo-II Poisons eliciting cell cycle delay. Mitotic cells are associated with phosphorylation of Histone H3 at serine 28 (Hans and Dimitrov, 2001), and hence one would expect Topo-II Poison mediated mitotic arrest to be associated with a decrease in expression of p-H3 in the same MultiFlow biomarkers response compared to healthy dividing cells.

With respect to existing data from the MultiFlow assay for Topo-II Poison compounds, previously published data included mode of action classification which classified select Topo-II Poisons into the clastogen class of compounds by Dertinger et al. (2019). Although the authors interestingly identify the group of Topo-II Poison compounds as a distinct clade separate from other clastogens *via* the unsupervised hierarchical clustering technique (**Figure 5.3**), this was an observation based on *a priori* knowledge of the compound's mechanism. Further in-depth information to support the Topo-II Poison classification of compounds was not provided in the article, probably because it was out of scope of the same publication.
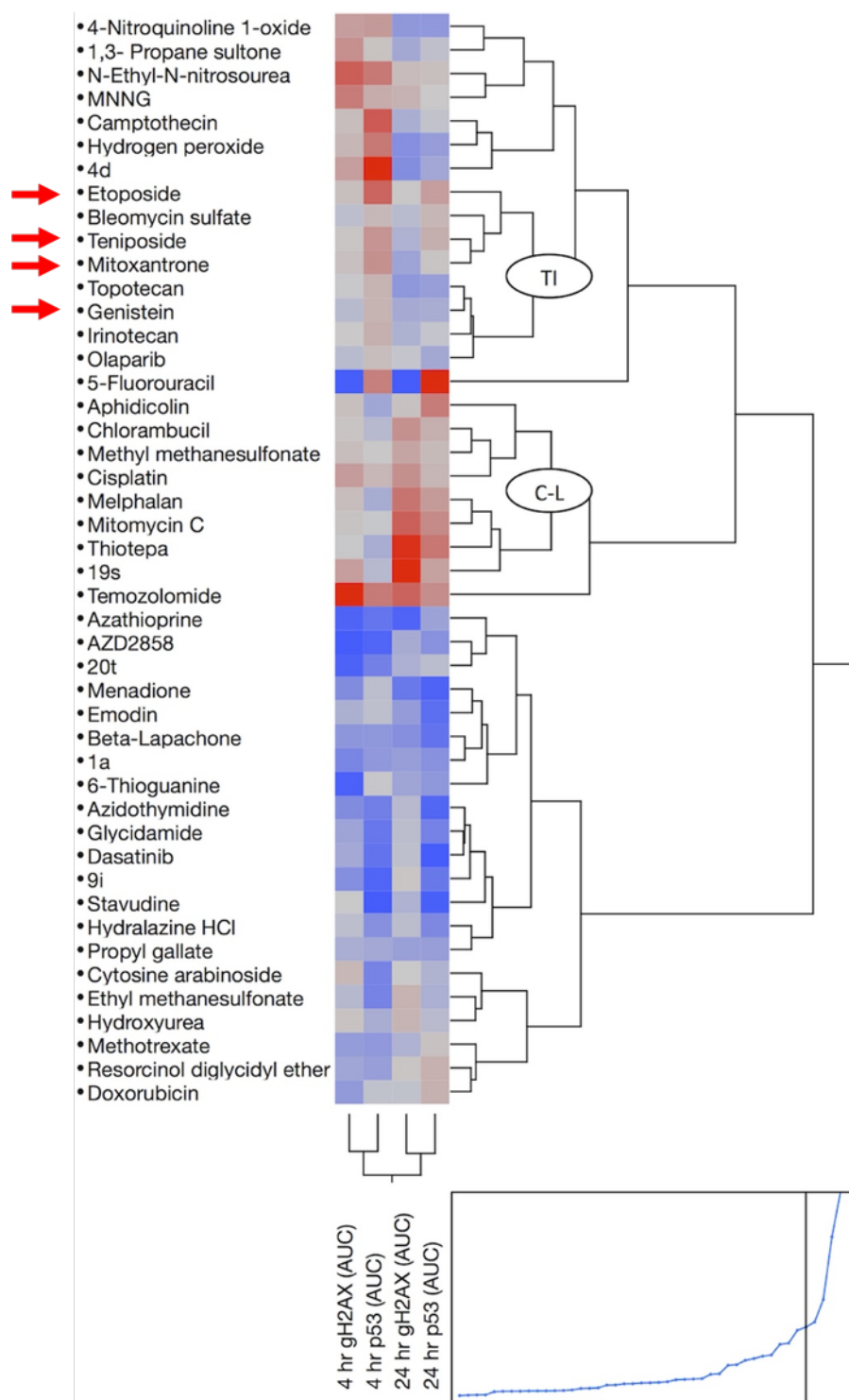
Figure 5.3. 2-Dimensional dendrogram from unsupervised hierarchical cluster analysis of the MultiFlow biomarker dose-response Area Under the Curve (AUC) measurements from tested compounds with *a priori* mode of action information. The Topo-II Poison compounds etoposide, teniposide, mitoxantrone, and genistein all cluster in the same region depicted with the clade labelled TI. The red arrows have been added to the figure to highlight the location

of the Topo-II inhibitor compounds. Key: TI = Topo-II Inhibitor; C-L = Cross-Linker. Figure adapted from Dertinger et al. (2019).

Additionally, the Toxys ToxTracker assay's biomarker responses could provide additional or complementary mechanistic information for the Topo-II Poison compounds in addition to the MultiFlow biomarker responses. Again, based on the known mechanism of action of the Topo-II Poison compounds, the ToxTracker biomarker responses are expected to perform in line with the following events specific to biomarkers responsive to DNA damage effects: The Rtkn biomarker should exhibit an increase in response associated with ATM recruitment to DNA double strand breaks; the Btg2 biomarker should exhibit an increase in response associated with p53 activation upon cell cycle stalling; and, the Bscl2 biomarker should exhibit an increase in response indicative of being activated by the ATR-Chk1 signalling pathway in response to stalled DNA replication forks (Hendriks et al., 2011).

With respect to existing data from the ToxTracker assay on the responsiveness of the biomarkers to exposure to Topo-II compounds, Hendriks et al. (2015) included exposure of etoposide and doxorubicin to the assay's six green fluorescent protein (GFP) tagged mouse embryonic stem cell reporter lines. GFP expression in genes associated with DNA damage response was evaluated by flow cytometry. A compound is classed as genotoxic if a greater than 1.5-fold increase in GFP-induction is achieved in the Bscl2 and Rtkn DNA damage reporters. **Figure 5.4** shows the GFP-biomarker response heat maps for the tested compounds including etoposide and doxorubicin.
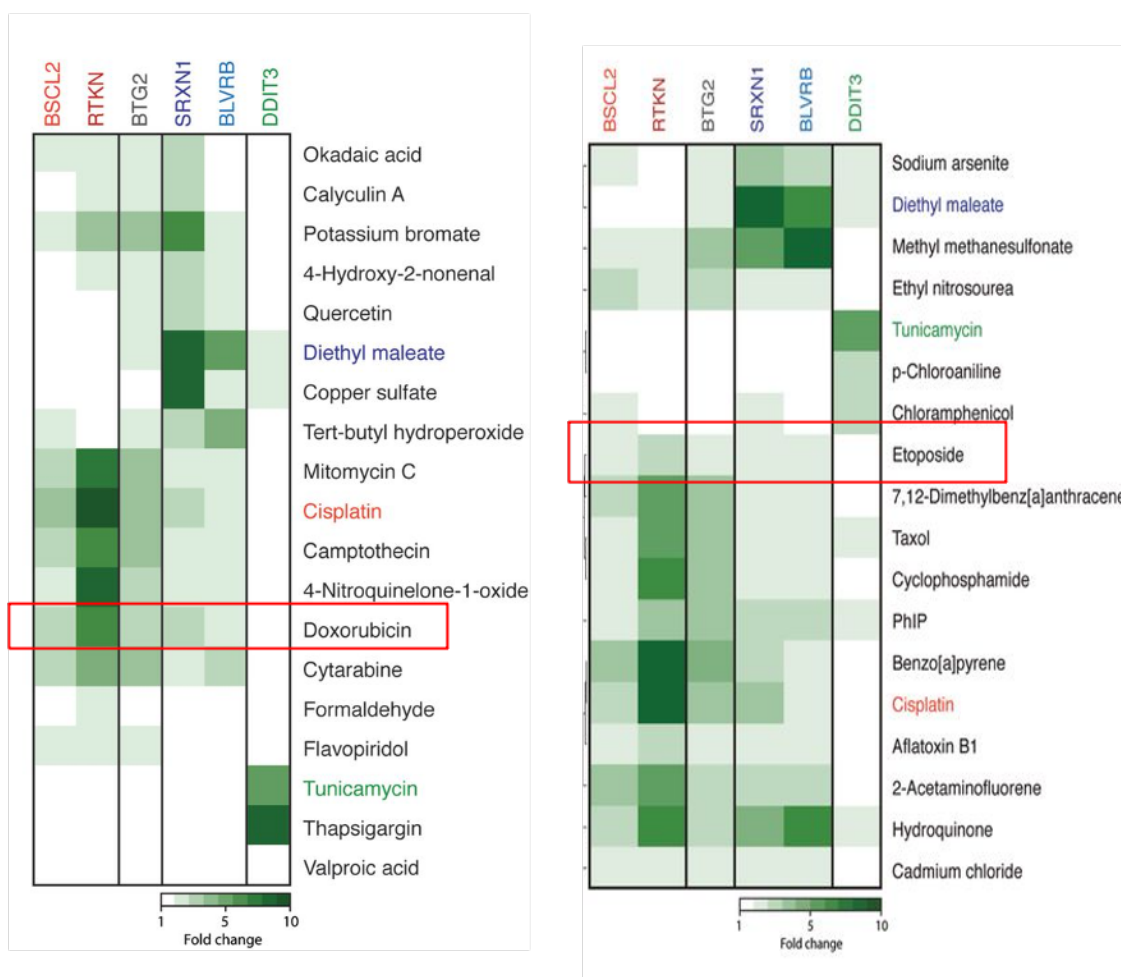
Figure 5.4. Heat maps showing the relative fold-increase levels in GFP induction in the ToxTracker 6 mES cell line biomarker responses. The Topo-II Poison compounds Etoposide and Doxorubicin (outlined here with red boxes) show strong very strong induction in the DNA damage response biomarkers Rtkn and Bscl2. There is some activation of the Btg2 reporter indicative of stalled replication forks, as well as the oxidative stress response reporters Srxn1 and Blvrb. There appears to be no expression of the unfolded protein response Ddit3 biomarker for both compounds in these heat maps. Figure adapted from Hendriks et al. (2015).

5.1.3.2 MultiFlow and ToxTracker Assay Potency Information

The same comparison can be made between the MultiFlow and ToxTracker assays that the results are interpreted in the paradigm of hazard characterisation. Little potency information can be derived from the biomarker response metrics presented in the assay data interpretation. The most one can conclude is that perhaps doxorubicin is 'more potent' than etoposide based on stronger induction of the DNA damage response biomarkers **Figure 5.4**

studied in the ToxTracker assay, and that etoposide is 'more potent' than teniposide due to the same argument in the MultiFlow assay **Figure 5.3**; however, the magnitude and levels of uncertainty are missing in the comparative potency determinations based on comparing fold-increase metrics.

It has already been demonstrated that the MultiFlow assay biomarker responses can produce robust potency metrics when evaluated by BMD modelling across a range of compound classes [Chapter 3 and Dertinger et al. (2019)]. There is limited existing published data on BMD analysis of the ToxTracker biomarker responses. One recent article by Cappellini et al. (2018) published the results of their study of exposure of cobalt-based nanoparticles to the ToxTracker assay. The researchers performed BMD analysis of the resulting biomarker dose-response datasets which proved amenable to BMD analysis when dose-responses were evident. Some other researchers have initiated collaborative projects with Toxys including BMD analysis of the ToxTracker datasets (personal communication Giel Hendriks, Toxys B.V.) This provides some indication that the ToxTracker assay biomarker dose-response datasets for the Topo-II Poison compounds should be responsive to BMD analysis.

Thus, the primary hypothesis pertains to the suggestion that the combination of biomarker potency information from the *in vitro* high throughput assays could add valuable detailed quantitative mechanistic information about the Topo-II Poison class of compounds beyond the published *in vitro* micronucleus (Lynch et al., 2003; Gollapudi et al., 2019) and MultiFlow and ToxTracker assay hazard characterisation data (Hendriks et al., 2015; Dertinger et al., 2019). No previously published literature existed where novel biomarker responses had been assessed *via* BMD modelling for robust potency determinations for the Topo-II Poison class of compounds.

5.1.4 Secondary Hypothesis: Potency Ranks Accounted for with Structural and Mechanistic Information

It is important to acknowledge that the findings of Wills et al. (2016a) inspired the secondary research hypothesis developed in this results chapter. Specifically, Wills et al. (2016a) demonstrated that the BMD combined covariate approach applied to *in vitro* micronucleus dose-responses for the benzimidazole class of compounds permitted robust compound potency ranks (BMD confidence interval potency plot presented in Introduction

Chapter **Figure *1.10***). The researchers scrutinised patterns in the confidence intervals in the potency rank plot and suggested that the observed potency differences are on account of compound structure. The authors proposed that the large moiety in the 5-position of oxibendazole, mebendazole, and flubendazole may account for the higher potency than benomyl and carbendazim that lack the same structural feature – which are all benzimidazole derivatives that are known to competitively inhibit colchicine binding. Colchicine binding site inhibitors act by a common mechanism *via* binding to the colchicine site on tubulin, and subsequently inhibit tubulin assembly and suppressing microtubule formation (Lu et al., 2012). Wills et al. (2016a) describes that diethylstilboestrol and the structurally similar compound bisphenol-A display lower potencies likely due to the former compound's non-specific action at the colchicine binding site, and the latter compound's effect on promoting spindle poison polymerisation through stabilisation of the microtubule. The other benzimidazoles albendazole oxide and thiabendazole showed lower potency whilst also acting at the colchicine binding site. Wills et al. (2016a) concluded that emphasis on compound mechanism of action and potency is needed in multi-compound genotoxicity assessments.

To this end, it was hypothesised if the same approach could be applied to generating genotoxic potency ranks within the Topo-II Poison class of compounds, however with dose-response data from the MutliFlow and ToxTracker assay's DNA damage response biomarkers to provide more mechanistic information to support potential genotoxic potency ranks of the same chemicals.

5.2 Materials and Methods

The following flow chart (**Figure 5.5**) summarises the methodology steps that were followed to arrive at the results presented in this chapter. Each subsection of this materials and methods chapter provides detail on steps 2 to 6 and 8 from the flow chart.
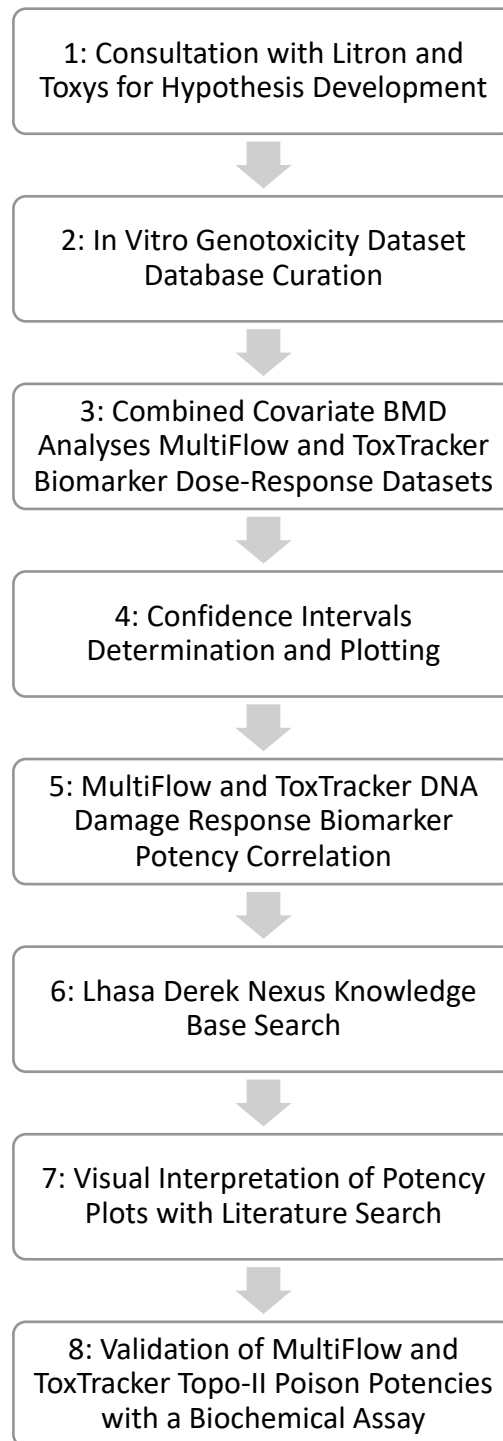
Figure 5.5. Materials and Methods Flow Chart

5.2.1 Datasets

Raw data for BMD analysis was provided by the collaborating scientists at Litron Laboratories and Toxys. Raw data for BMD analysis in this chapter is defined as tabulated median fluorescence, or median relative frequency data outputs from flow cytometric

analysis performed at Litron Laboratories and Toxys B.V. Raw data for analysis other than BMD analysis was either obtained from the Lhasa Derek Nexus software programme (section 5.2.5) or from literature (section 5.2.6).

MultiFlow assay raw data took the format of an Excel spreadsheet with headings including compound name, concentration (μM), median fluorescence data for each compound and concentration level for the biomarkers γH2AX and p53, median relative frequency values for p-H3, polyploidy, RNC (Relative Nuclei Count), and RPD (Relative Population Doubling). All measurements were performed at the 4-hour and 24-hour sampling timepoint and indicated in the column heading.

ToxTracker assay raw data took the format of an Excel spreadsheet with a maximum of 2 individual compounds reported separately per Tab. Tab headings including well, reporter name, treatment (compound name), concentration (μg/mL), mean GFP values, induction compared to control, cell concentration compared to control and survival compared to control for all biomarker responses specific to Bscl2, Btg2, Rtkn, Srxn1, Blvrb, and Ddit3. All measurements were performed at the 24-hour timepoint.

5.2.1.1 MultiFlow Assay Specific Information Pertaining to the Dataset

Litron Laboratories performed the benchtop experiments using Topo-II Poisons to acquire the MultiFlow dataset for analysis. The assay has been extensively discussed in the published literature (Bryce et al., 2014; 2016; 2017; 2018). Briefly, human lymphoblastoid TK6 cells were prepared using the reagents and instructions specified in the MultiFlow DNA damage kit – p53, γH2AX, Phosphohistone-H3. Cell suspensions were exposed to a total of 8 Topo-II Poison compounds (**Table 5.1**) and processed for flow cytometric analysis at the 4- and 24-hour timepoints. Reported raw data were restricted to those concentrations that did not exceed the MultiFlow assay's cytotoxicity limit, that is, the top concentration of each chemical had to exhibit ≤80% reduction to RNC at the 24-hour time point. In the absence of excessive cytotoxicity, the top concentration was 1 mM, or the lowest precipitating concentration, whatever was lower.

Table 5.1. Topo-II Compounds Studied in the MultiFlow Assay and their Identification

Compound	Compound Abbreviation	CAS No.
Ciprofloxacin	cip	85721-33-1
Doxorubicin	dox	23241-92-8
Emodin	emo	518-82-1
Etoposide	etp	33419-42-0
Flumequine	flm	42835-25-6
Genistein	gen	446-72-0
Mitoxantrone	mit	65271-80-9
Teniposide	ten	29767-20-2

5.2.1.2 ToxTracker Assay Specific Information Pertaining to the Dataset

Toxys B.V. performed the benchtop experiments to acquire the ToxTracker dataset for analysis. The assay has been extensively discussed in the published literature (Hendriks et al., 2011; 2015). Briefly, mouse embryonic stem (mES) cells that were GFP tagged for the assays specific biomarkers were exposed to a test set of 6 Topo-II Poison compounds (**Table 5.2**) at 5 test concentrations. Emodin and etoposide were exposed with and without S9. Although neither compound requires metabolic activation to form genotoxic metabolites, the exposure to S9 was included in the standard protocol. Following a 24-hour period of incubation, the induction of GFP was measured by flow cytometry. The reporter activity was determined as the mean fluorescent intensity from 5,000 intact cells.

Table 5.2. Topo-II Compounds Studied in the MultiFlow Assay and their Identification

Compound	CAS No.
Doxorubicin	23241-92-8
Daunorubicin	20830-81-3
Emodin (+/- S9)	518-82-1
Etoposide (+/- S9)	33419-42-0
Mitoxantrone	65271-80-9
Teniposide	29767-20-2

5.2.2 Database Curation

The PROAST software requires data to be formatted for analysis including the following specific strict rules:

- All decimal points in point format not comma format
- Empty cells are not allowed
- Spaces between text or units are not allowed
- The raw data takes the format of the data matrix. In this case, the data matrix was the compound names, concentration values, median fluorescence, median relative frequency, or percent values.
- Each column must have a header which describes the data in the data matrix
- Four rows must exist above the data matrix and complete according to the following rules:
  - First line: a one-word title that describes the data. In this case the MicroFlow data had the heading "MicroFlow", and the ToxTracker data had the heading "ToxTracker".
  - Second line: the number of columns in the data matrix

- o Third line: includes a numerical code for the data type in the column (see below)
- o Fourth line: a one-word title for the column. The column title choices for the MultiFlow and ToxTracker datasets are presented in **Table 5.3** and **Table 5.4**, respectively.
- o Fifth and further lines: the data matrix.
- Data types (third line)
  - o 0 – for all columns not representing a response variable
  - o 1 – continuous data. In our case, all MultiFlow biomarker response variables (γH2AX, p-H3, p53), and all ToxTracker biomarker (Bscl2, Btg2, and Rtkn) GFP values are continuous data.
  - o 2 – binary data (e.g., yes/no response in individual animals)
  - o 3 – ordinal data (e.g., histopathology scores of severities)
  - o 4 – quantal data (e.g., number of responding animals per number of animals in the group)
  - o 5 – continuous data that are nested (e.g., fetal weights within litters)
  - o 6 – quantal data that are nested (e.g., number of fetuses affected within litters)
  - o 10 – mean (continuous) response in a dose group

Data types 2 to 10 are out of scope of the MultiFlow and ToxTracker assay result variables. Hence the third line for the MicroFlow and ToxTracker databases all included either a value of 0 or 1.

Since not all variables are relevant to the hypotheses that were being tested in this study, some columns were excluded from analyses. For example, although the MultiFlow assay provides information relative to polyploidy through the evaluation of propidium iodide gating logic – said column was excluded from the database. Percent RNC and RPD columns were maintained since the data are indicative of treatment related cytotoxicity effects which could prove important in interpreting the BMD analysis outputs. Additionally, the ToxTracker data includes information on relative GFP induction compared to control, cell concentration compared to control and survival compared to control. GFP and induction scores are not in scope of the hypotheses tested in this study and so the columns were removed. However, the relative survival column was maintained as well as an indicator of test article induce cytotoxicity. Additionally, the focus was on the DNA damage response

biomarkers including Bscl2, Btg2, and Rtkn that are hypothesised to be responsive to Topo-II Poisons (5.1.3.1), and hence the oxidative damage Srxn1 and Blvrb, and the unfolded protein response biomarker Ddit3 were not included in the analyses. **Table 5.3** and **Table 5.4** summarise the final selected database columns, rows, and headers formats for the MultiFlow and ToxTracker datasets, respectively.

The ToxTracker data concentration units were provided in μg/mL. One of the latter steps in the process was to compare the potencies obtained from the ToxTracker biomarker responses with those obtained from the MultiFlow biomarker responses. Hence, to normalise the concentration to the same molar concentration units, the ToxTracker concentrations were converted from μg/mL to μM using the molecular weight of the compounds per PubChem (https://pubchem.ncbi.nlm.nih.gov/) by searching the compound using the CAS numbers provided in **Table 5.2**. Transformation from μg/mL to μM is achieved by normalising the mass concentration to g/L and then division by the molecular mass in g/mol, followed by conversion to μM units. For example, one of the tested concentrations of emodin was 125 μg/mL, which equals 0.125 g/L, and the same compound has a molecular weight of 270.24 g/mol (https://pubchem.ncbi.nlm.nih.gov/#query=emodin). Therefore, the molar concentration is 0.46 μM (0.125 g/L ÷ 270.24 g/mol × 1000).

Table 5.3. Replication of the MultiFlow database columns, rows, and headers formatted per PROAST strict data formatting requirements

Row Number										
1	MultiFlow									
2	10									
3	0	1	1	1	1	1	1		1	1
4	Compound	Conc.µM	4hr.gH2AX	24hr.gH2AX	4hr.pH3	24hr.pH3	4hr.p53	24hr.p53	RNC.perc	Cytotoxicity.perc[8]
5 and more…		Data Matrix…								

---

[8] Cytotoxicity was calculated at 100% minus the %RNC at 24-hours.

Table 5.4. Replication of the MicroFlow database columns, rows, and headers formatted per PROAST strict data formatting requirements

Row Number								
1	ToxTracker							
2	8							
3	0	0			0	1		1
4	Compound	Conc.µM	Bscl2_GFP	Bscl2_Surv[9]	Btg2_GFP	Btg2_Surv[9]	Rtkn_GFP	Rtkn_Surv[9]
5 and more…	Data Matrix…							

---

[9] Surv = Relative Survival metrics provided in the ToxTracker datasets

Once the Excel databases were formatted according to the strict rules prescribed by PROAST, the Excel sheets were transformed into tab delimited files (.txt) and stored in a working repository.

5.2.3 Working in PROAST

PROAST v61.2 (for MultiFlow data) or v69.1 (for ToxTracker data) software package (http://www.proast.nl) operating in the R statistical suite was used for the BMD analyses presented in this chapter. The 'MENU' version of PROAST was used to navigate through the data and variables selection process in the BMD analyses. Firstly, the .txt files were imported into PROAST and analysed independently. The 'continuous' data type was selected from the MENU option, and 'concentration' selected as the 'independent variable'. The MENU option of the PROAST package was navigated to select the appropriate parameters to perform BMD analysis of the MultiFlow and ToxTracker biomarker response dataset. The MultiFlow responses included each of the MultiFlow biomarkers – being 4- and 24-hour p53, 4- and 24-hour γH2AX, and 4- and 24-hour p-H3. The ToxTracker responses included each of the ToxTracker biomarkers under test – being Bscl2-GFP, Btg2-GFP, and Rtkn-GFP.

5.2.4 BMD Analysis

For the MultiFlow biomarker responses (24-hour γH2AX, p53, and p-H3) consistent with the approach reported in Chapter 3.2.4 of this thesis, the selection of the maximal 4-parameter model (model 5) was selected from the PROAST menu (equation below). The maximal exponential model was selected based on it containing the greatest number of parameters for describing the dose response. A Hill model was not included in the analyses based on the work of (Slob and Setzer, 2014) who demonstrated that the shape and steepness of dose-response curves do not differ between the results of either exponential or Hill models.

Exponential Model 5: $y = a\left[c - (c-1)\exp\left(-\frac{x}{b}\right)^d\right] with\ a > 0, b > 0, c > 0, d \geq 1,$

where y denotes the continuous endpoint, and the dose by x. Parameter a = the level of the endpoint at dose 0; b = potency/efficacy of the test compound on the shape of the response; c = maximum fold-change in response; and d = rate of change in the response constrained to values equal to or larger than 1.

For the ToxTracker dataset PROAST v69.1 included the option to select either model 3 or model 5 when performing BMD analysis to reflect the latest BMD guidance recommended by EFSA (EFSA, 2017). In doing so, the model with the lowest Akaike Information Criterion (AIC) is selected for calculating the BMD confidence interval. In brief, AIC provides the means for model selection based on the quality of each model considering the prediction error within each model (McElreath, 2018; Taddy, 2019). The AIC analyses resulting in the selection of the maximal 4-parameter exponential model based on it containing the greatest number of parameters for describing the dose response (Slob and Setzer, 2014; EFSA, 2017). A Hill model was not included in the analyses based on the work of (Slob and Setzer, 2014), who demonstrated that the shape and steepness of dose-response curves do not differ between the results of either exponential or Hill models.

5.2.4.1 Critical Effect Size Determination

For the MultiFlow datasets, different values of CES were selected to determine if choice of CES impacted the potency rank order of the Topo-II poisons evaluated in the MultiFlow assay. To this end, the following CESs were trialled for the MultiFlow biomarker response BMD analyses: 4- and 24-hour γH2AX 0.1 (10%), 0.5 (50%) and 1.0 (100%); 4- and 24-hour p53 0.1 (10%), 0.2 (20%) and 0.5 (50%); and 4- and 24-hour p-H3 0.1 (10%), 0.2 (20%) and 0.5 (50%). The CES values for the γH2AX biomarker were selected to bracket the CES ranges (0.2-0.4) recommended by Zeller et al. (2016) for the same biomarker response. The CES values for the p53 and p-H3 biomarker responses were arbitrarily selected to determine if more conservative CES values impacts the potency rank order compared to the previous experience of CES 0.3 and 0.5 for the same biomarkers (Chapter 3 and Chapter 4). Potency rank orders were visually compared across all biomarker responses.

For the ToxTracker datasets, PROAST v69.1 includes the option to calculate and use an endpoint specific CES based on maximum response and within group variance described in Slob (2017). An endpoint specific CES can be termed a marker specific CES for these *in vitro* biomarker responses. Thus, the option was selected and resulted in marker specific CESs for the ToxTracker biomarker responses. For proof-of-concept, a CES of 0.5 (50%) was also selected to provide potency comparison with the MultiFlow biomarker responses under the marker specific CES conditions.

5.2.4.2 Confidence Interval Determination

The PROAST standard 90% confidence intervals were calculated for each level (compound) of the covariate and save in an R data file. The confidence intervals were automatically plotted by PROAST in a Log plot. Note that PROAST v61.2 standard output does not organise each level of the covariate (compound) in order of potency rank order. Manual rearrangement of the potency rank order was performed for the MultiFlow compounds by ranking the compounds by Log midpoint and numbering them from 1 to 8. PROAST v69.1 standard output ranks confidence intervals in rank order by Log midpoint.

5.2.4.3 System Cross-Correlation

To perform system cross-correlations, the previously determined 90% confidence intervals for each covariate of the MicroFlow and ToxTracker assay biomarker responses were re-imported into PROAST and the 'f.bigtable' function was used to collate a tabulated format for the confidence interval values. In doing so, irrelevant columns were removed from the file by following the steps in the PROAST menu options. A .txt file was generated by PROAST that contained the BMD confidence bounds and the relevant labels for each of the associated factors that were originally present in the other columns of the dataset. The file was saved in the working directory as a 'bigtable'. A 'bigtable' was created for each endpoint's BMD analysis, i.e., 4- and 24-hour γH2AX, p53 median fluorescence, p-H3 frequency data, Bscl2-GFP, Btg2-GFP, and Rtkn-GFP. Finally, a new working directory was created and the relevant 'bigtable' .txt files were copied into the new working directory and the 'f.cor.table' PROAST function was used to begin the process of iteratively creating BMD correlation plots. A 'crosstable' was created, where the confidence intervals from two systems are present for each dataset according to the chosen covariate. The 'crosstable' was used upon entering the 'f.cor.plot' command with 2 different biomarker responses of choice labelled on the x and y axes, respectively.

Since the goal was to correlate the confidence intervals from compounds assessed for similar mechanistic responses to DNA damage in the MultiFlow and ToxTracker assays, the following correlations were created based on the a-priori knowledge of Topo-II Poison genotoxic mechanisms of action:

- DNA double strand breaks: MultiFlow γH2AX induction versus ToxTracker Rtkn-GFP expression

- p53 dependent cell cycle delay: MultiFlow p53 induction versus ToxTracker Btg2-GFP expression
- Stalled replication forks cell cycle delay/mitotic arrest: MultiFlow p-H3 downregulation versus ToxTracker Bscl2-GFP expression

5.2.5 Derek Nexus Structural Activity Knowledge

An important step in performing risk characterization of chemicals is evaluating compound read across (analogue searching). The methodology involves comparing compounds of interest to one of familiarity with respect to compound structure and known toxicological effects (Dearfield et al., 2017). Expert rule-based Quantitative Structure–Activity Relationship [(Q)SAR] software programs have been developed and are used in industry genotoxicity assessments. Although these methodologies are typically utilized for classifying compounds based on their bacterial mutagenic potential, the expert rule-based software, Derek Nexus, can provide useful structural alert information to aid in further compound sub-categorization. Typically, the approach is restricted to qualitative hazard identification (ie, substance X has known mutagenic potential) (Amberg et al., 2016; Barber et al., 2017), however can provide structural alert information for other measures of genotoxicity (such as chromosome aberrations) and other toxicological endpoints such as skin sensitisation and hepatotoxicity (Segall and Barber, 2014). To this end, the Derek Nexus database was utilised to search for structural alerts and to provide compound structural information specific to the Topo-II Poison compounds tested in the MultiFlow and ToxTracker systems.

5.2.6 Fluorescence Anisotropy Assay Data

One step in the process was to compare the Topo-II Poison compound potencies obtained from the *in vitro* high throughput flow cytometric methods (MultiFlow and ToxTracker) with another quantitative test system with the aim of the comparison maybe providing more mechanistic information and to test for the introduction of bias. To this end, a study by (Shapiro and Austin, 2014) was identified in which the authors developed a high-throughput fluorescence anisotropy-based assay specific to human Topo-II. The assay measures ATP-dependent relaxation of supercoiled plasmid DNA by the Topo-II enzyme. The assay was used to compare potency amongst 19 compounds, 7 of which match with the compounds studied in the MultiFlow and ToxTracker assays.

5.2.6.1 Statistical Analysis

The reported summary data (mean, standard deviation, and number of replicates) from the Shapiro and Austin (2014) publication of the Topo-II α and β enzyme inhibitory potencies (IC_{50} μM) measured in triplicate were used to calculate the Log_{10} of the mean and subsequent standard errors and plotted in Excel. Log_{10} was used to transform the data to approximate normality (Feng et al., 2014). Standard error of the mean (SEM) was used since it provides the margin of error associated with the IC_{50} value, which can be regarded as somewhat related to the BMD confidence intervals used to relate the level of uncertainty associated with the BMD estimate. SEM for each compound's IC_{50} potency was calculated using the following formula:

$$SEM\ (\sigma\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

where:

σ = standard deviation

$\bar{x}$ = the sample's mean

n = the sample size.

5.3 Results

5.3.1 BMD Analyses MultiFlow Biomarker Dose-Responses

Four parameter exponential BMD models applied to the MultiFlow biomarker responses resulted in good dose-response for most compounds. The following exceptions were observed in all CES selected for BMD analysis: Ciprofloxacin yielding no dose response in the 4- and 24-hour γH2AX biomarkers (**Figure 5.6**), and marginal scatter in the 4-hour p53 and 4-hour p-H3 biomarkers dose-responses (**Figure 5.7**); and flumequine yielded no dose-response in the 24-hour p53 and 4-hour p-H3 biomarker (**Figure 5.8**), and marginal scatter in the 24-hour p-H3 biomarker dose-response (**Figure 5.9**).

start values

E5: y = a*[c-(c-1)exp(-bx^d)]


```
version: 61.2
loglik    NA
var-   0.000535
a-    1
b-    0.000519
c-    1.03
d-    1
conv : 0
scaling factor on x :  1
dtype :  1
  selected :
  Compound cip
removed: none
```

Exponential model

m5-


```
version: 61.2
loglik   25.61
var-   0.00452
a-   0.998
CED-   0.0417
c-   0.884
d-   1.05
CEDL   0
CEDU   9.505
b:  15.8
conv :  1
scaling factor on x :  1
dtype :  1
  selected :
  Compound cip
removed: none
```
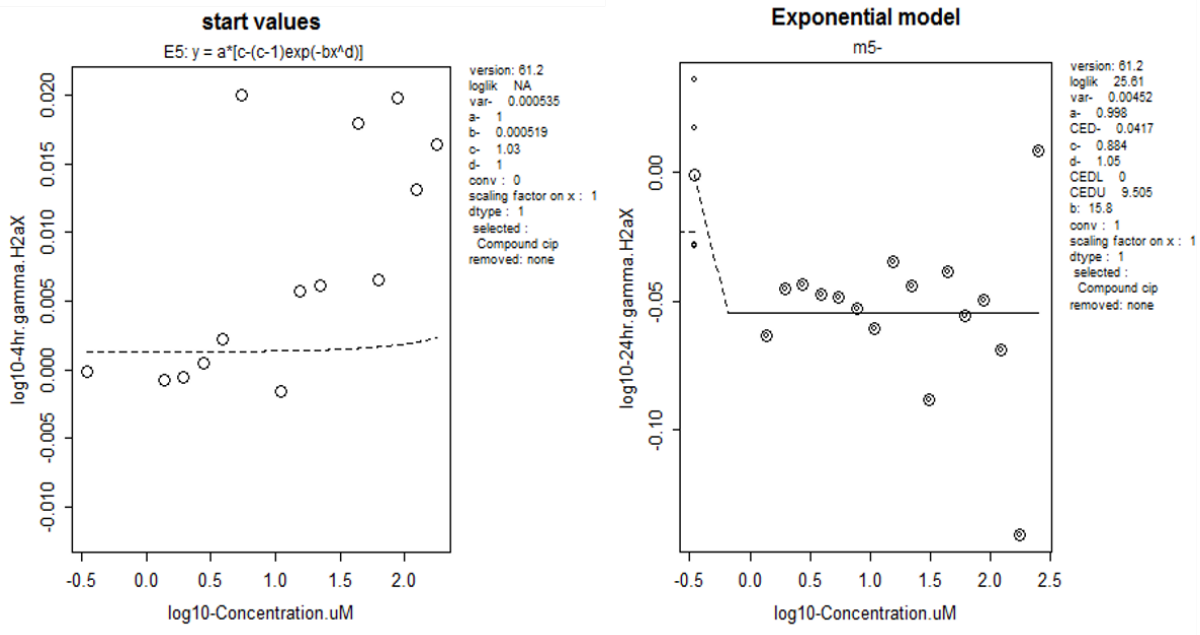
Figure 5.6. BMD dose-response models for ciprofloxacin in the 4-hour (left panel) and 24-hour (right panel) γH2AX biomarker responses. Large amounts of scatter are visible across all concentrations tested yielding no dose-response. Abbreviation: cip = ciprofloxacin.
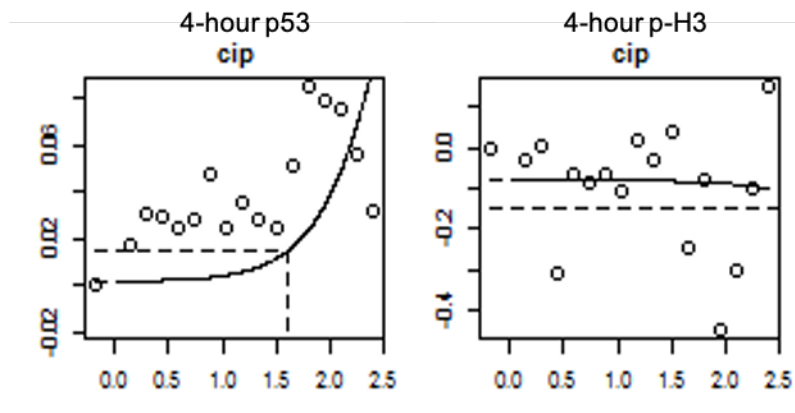
Figure 5.7. Individual dose-responses for ciprofloxacin tested in the 4-hour p53 (left panel) and 4-hour p-H3 (right panel) biomarker responses. Marginal scatter is observed in both curves which impacts the precision of the BMD estimate.

216
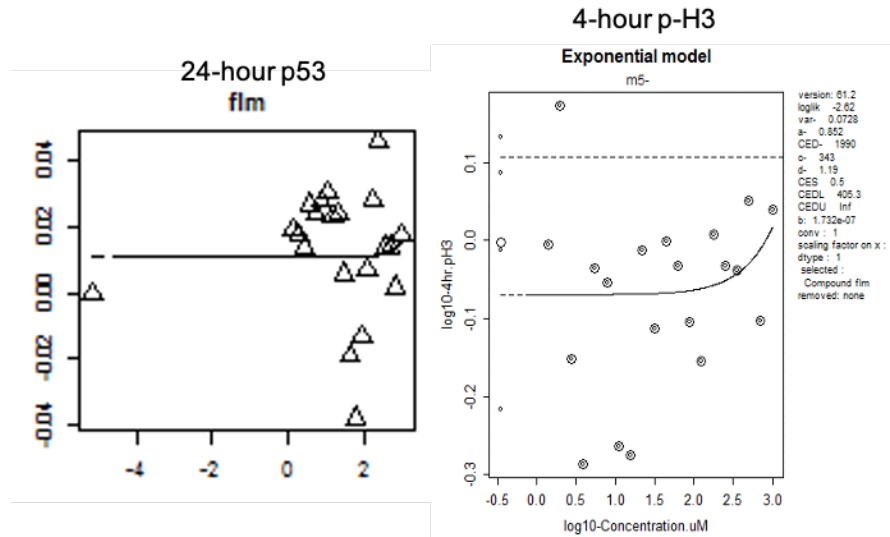
Figure 5.8. Individual dose-response curve (left panel) and BMD dose-response (right panel) for flumequine tested in the 24-hour p53 (left panel) and 4-hour p-H3 (right panel) biomarker responses with CES of 0.5. No dose-response was returned in either biomarker response. Abbreviation: flm = flumequine.
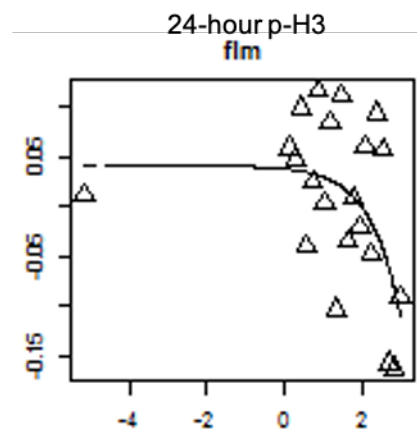
Figure 5.9. Individual dose-response curve for flumequine tested in the 24-hour p53 biomarker. Marginal scatter was observed in the dose-response curve. Abbreviation: flm = flumequine.

The dose-responses responded in the direction expected upon exposure to Topo-II Poisons. I.e., the γH2AX and p53 biomarkers were induced indicative of double strand breaks and cell cycle delay, respectively. The p-H3 biomarker became downregulated because of replication fork blockage and mitotic arrest. Dose-response curves for the 24-hour sampling time for each biomarker are shown in **Figure 5.10**, **Figure 5.11**, and **Figure 5.12** for illustration purposes to show the direction of the dose-response for the biomarker responses.

Dose-response curves for the 4-hour biomarker responses can be found in Appendix 5. Topo-II Poison 4-hour BMD Dose-Response Curves for MultiFlow Biomarker Responses.
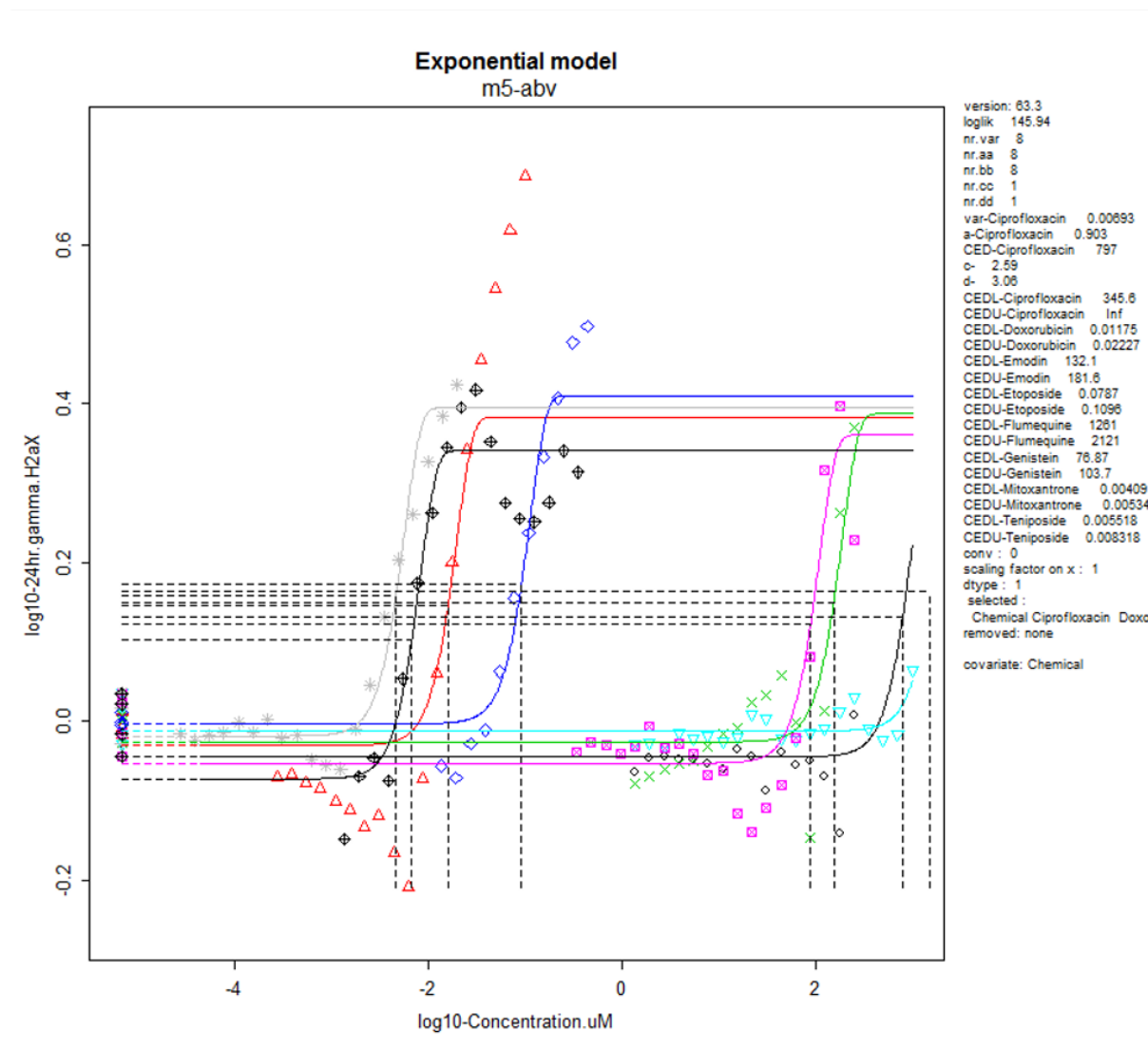
Figure 5.10. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 24-hour γH2AX biomarker response. Induction (increase) of the γH2AX response is indicative of DNA double strand breaks from Topo-II Poison exposure.
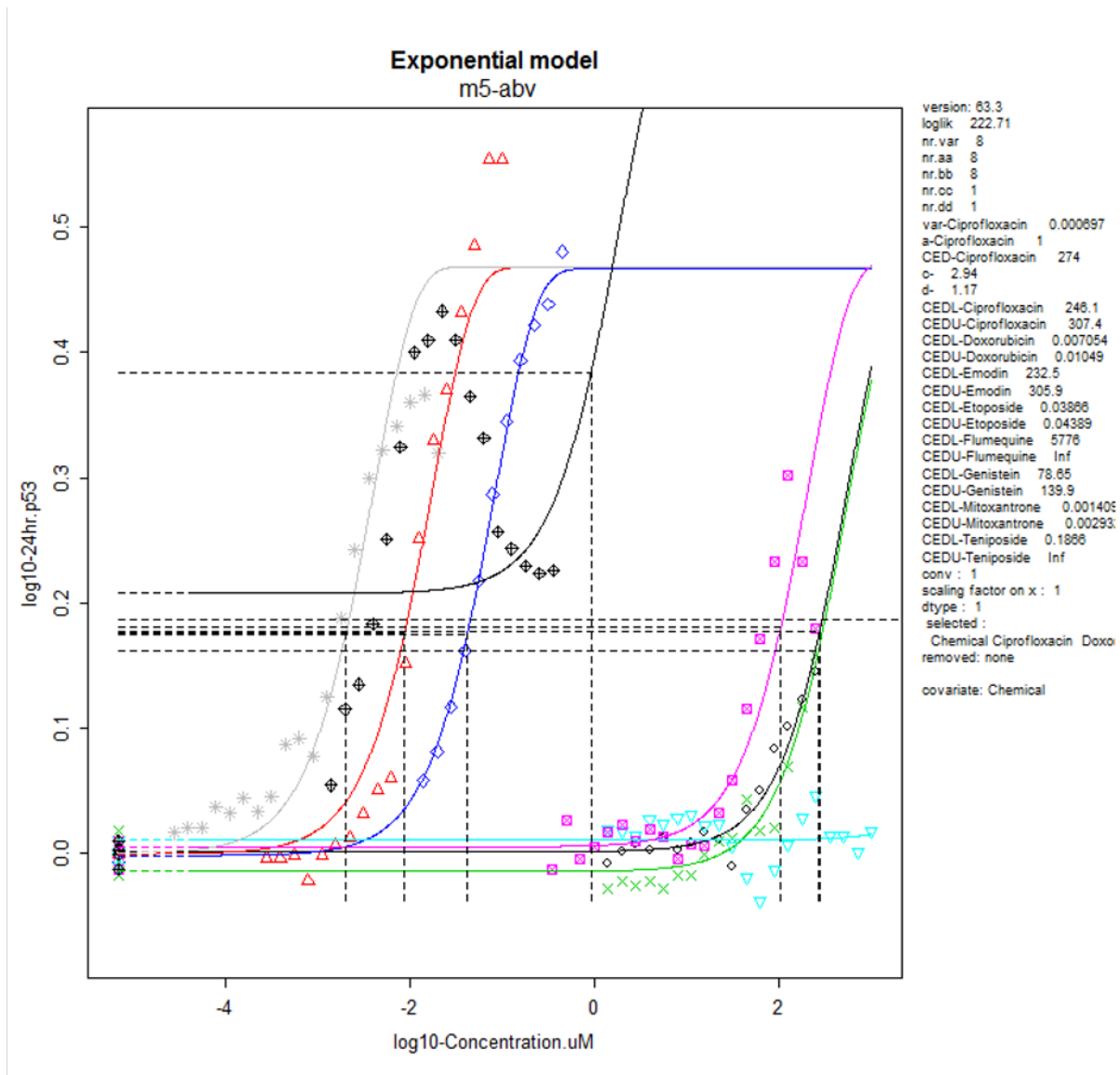
Figure 5.11. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 24-hour p53 biomarker response. Induction (increase) of the p53 response is indicative of cell cycle delay from Topo-II Poison exposure.
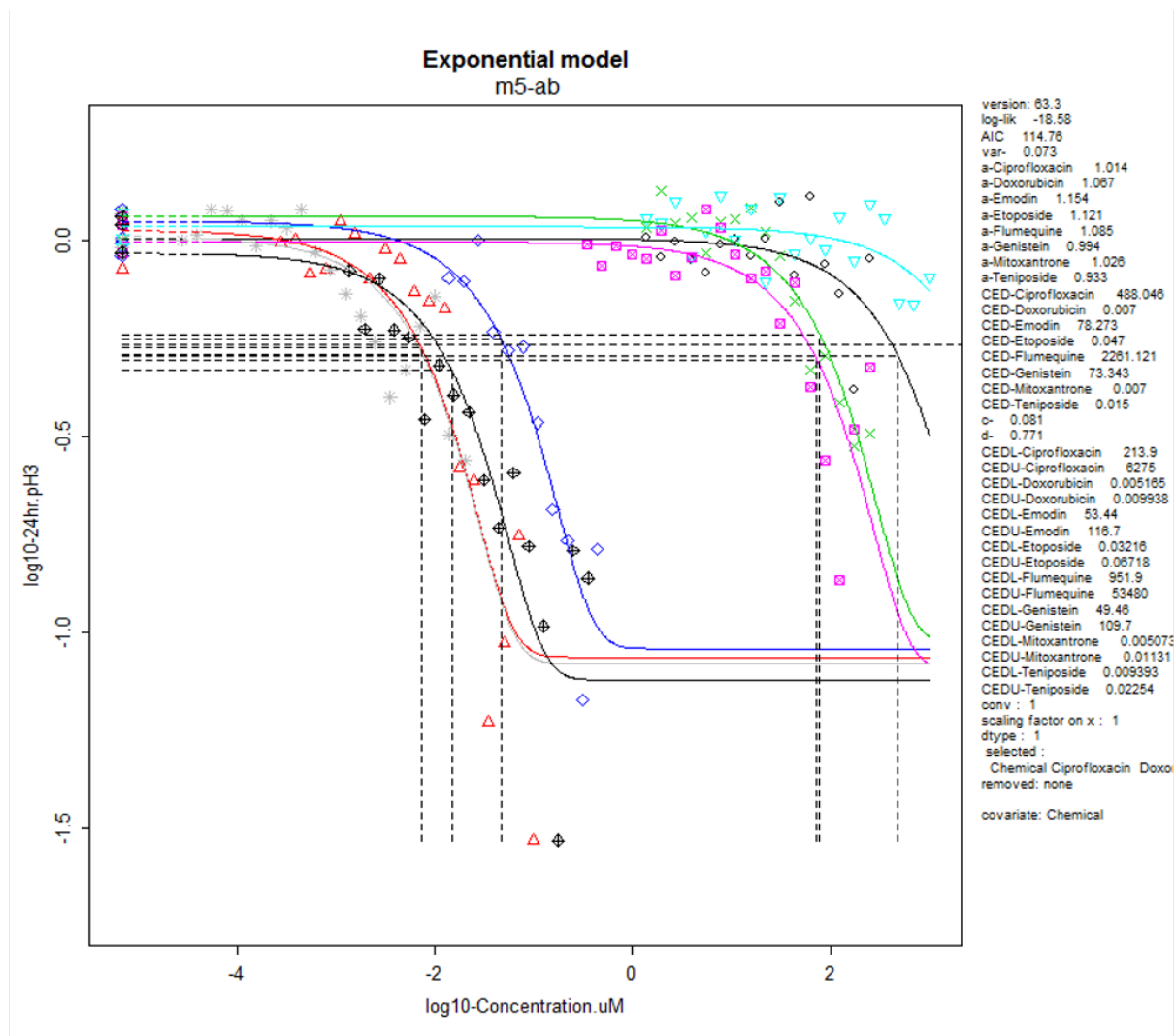
Figure 5.12. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 24-hour p-H3 biomarker response. Downregulation (decrease) of the p-H3 response is indicative of replication fork blockage and mitotic arrest from Topo-II Poison exposure.

5.3.2 Various CES Values Hypothesis Testing

The following CESs were used for the MultiFlow biomarker response BMD analyses: 4- and 24-hour γH2AX 0.1 (10%), 0.5 (50%) and 1.0 (100%); 4- and 24-hour p53 0.1 (10%), 0.2 (20%) and 0.5 (50%); and 4- and 24-hour p-H3 0.1 (10%), 0.2 (20%) and 0.5 (50%). The resulting confidence intervals were plotted in a Log scale for each level of the covariate for each CES. **Figure 5.13** shows the 24-hour γH2AX biomarker confidence interval plot, **Figure 5.14** for the 24-hour p53, and **Figure 5.15** for the 24-hour pH3. Only the confidence interval plots for the 24-hour timepoint biomarker responses are shown herein since

consistency was observed in the 4-hour timepoint (all 4-hour plots can be found in Appendix 6. Topo-II Poison 4-hour BMD Confidence Intervals for MultiFlow Biomarker Responses).

24-hour γH2AX BMD Confidence Interval Plots with Different CESs
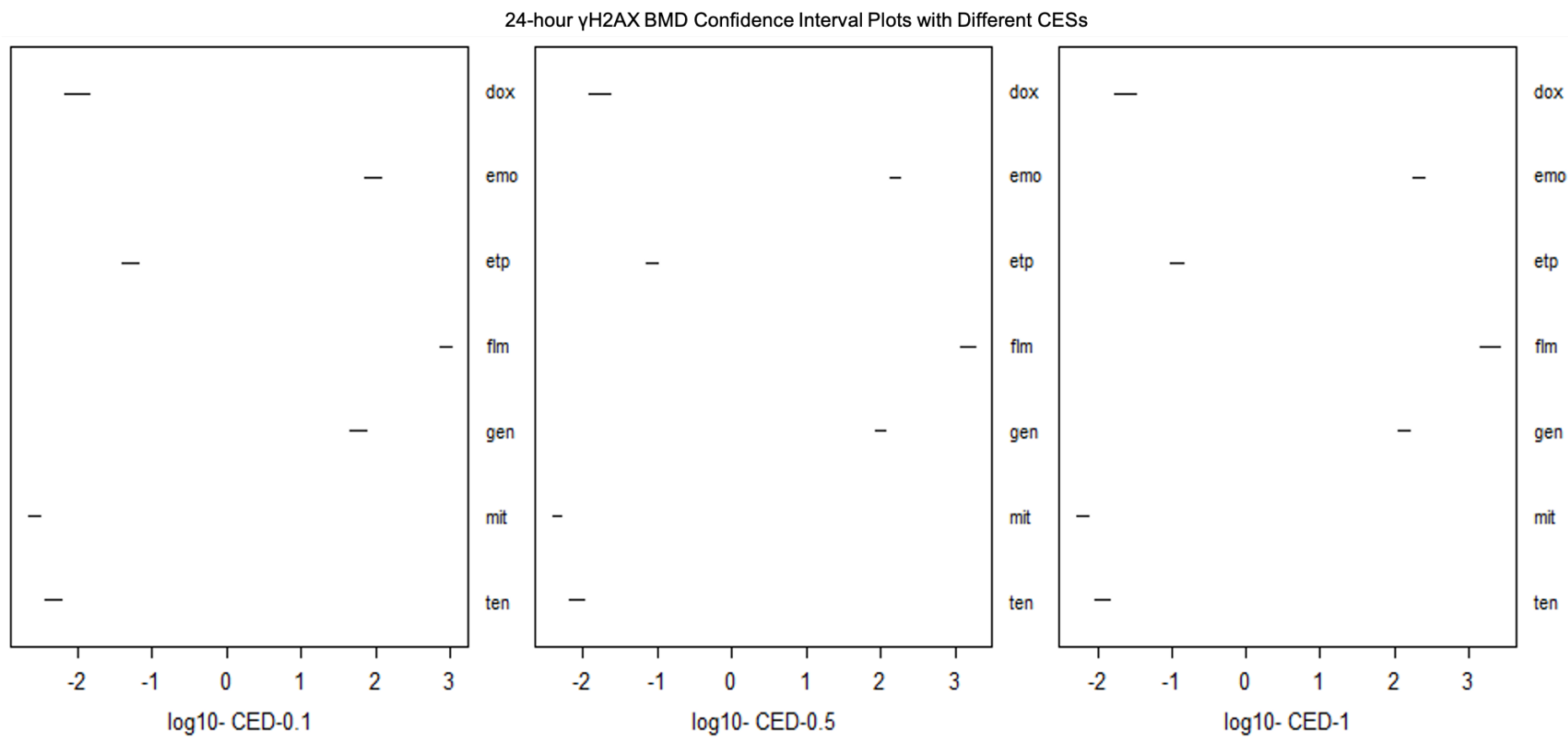
Figure 5.13. BMD Confidence Interval Plots resulting from different CES values. CED (Critical Effect Dose) denoted on the *x-axis* corresponding to the indicated CESs in percent. Panel left to right CES 0.1 (10%), 0.5 (50%) and 1.0 (100%) for the 24-hour γH2AX biomarker responses to Topo-II Poison compounds. The compound abbreviations cited on the right-hand side of each panel correspond to the compound identification information in **Table 5.1**. The potency rank order does not differ between the three different CESs.

Figure 5.14. BMD Confidence Interval Plots resulting from different CES values for the 24-hour p53 biomarker responses to Topo-II Poison compounds. CED (Critical Effect Dose) denoted on the *x-axis* corresponding to the indicated CESs in percent. Panel left to right CES 0.1 (10%), 0.2 (20%) and 0.5 (50%). Compound abbreviations on the right-hand side of each panel following the abbreviations in **Table 5.1**. The potency rank order does not differ between the three different CESs.

Figure 5.15. BMD Confidence Interval Plots resulting from different CES values for the 24-hour p-H3 biomarker responses to Topo-II Poison compounds. CED (Critical Effect Dose) denoted on the *x-axis* corresponding to the indicated CESs in percent. Panel left to right CES 0.1 (10%), 0.2 (20%) and 0.5 (50%). Compound abbreviations on the right-hand side of each panel following the abbreviations in **Table 5.1**. The potency rank order does not differ between the three different CESs.

5.3.3 BMD Confidence Interval Potency Ranks

PROAST v62.1 did not produce confidence interval plots in rank order arranged by confidence interval Log midpoint. It was deemed ideal to display the confidence interval plots with decreasing potency from left to right of the plot consistent with the confidence interval plots in the literature (Wills et al., 2016a; 2016b; 2017). Since PROAST v62.1 ordered the confidence intervals by alphabetical or numerical order, this issue was overcome by manually coding the potencies in rank order from 1 to 8 for the maximum number of compounds tested. Thus, PROAST v62.1 generated BMD confidence interval plots in potency rank ordered arranged by Log midpoint. Each biomarker response BMD confidence interval plots are presented in **Figure 5.16**, **Figure 5.17**, and **Figure 5.18**. Since the potency rank order did not differ based on the choice of CES, a CES of 0.5 (50%) was selected for all further data analysis and interpretation.

There were very small confidence intervals obtained for most compounds in the MultiFlow biomarker responses, generally in the order of approximately 0.25-0.50 orders of magnitude, which demonstrates excellent BMD precision in these analyses. The least potent Topo-II Poisons ciprofloxacin and flumequine were problematic showing some instances of disproportionately large or infinite BMDUs in the confidence interval potency plots.
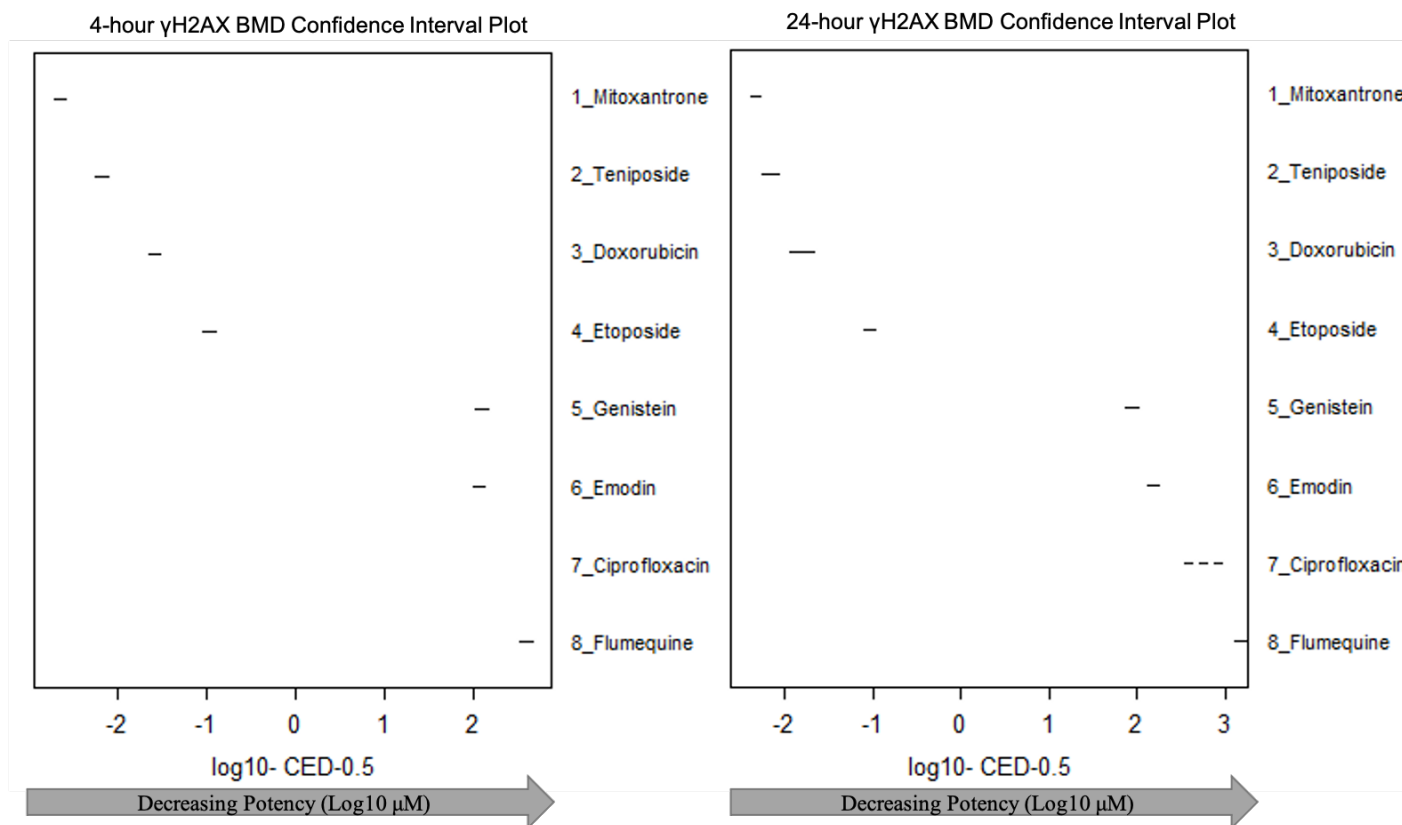
Figure 5.16. BMD confidence interval plots for the γH2AX biomarker responses. Confidence intervals are plotted in potency rank order with decreasing potency from left to right. There were very small bound confidence intervals for most compounds. No dose-response for ciprofloxacin in the 4-hour γH2AX biomarker and infinite BMDU in the 24-hour γH2AX biomarker shown as a dashed line confidence interval. Flumequine's BMDU was disproportionally larger, so plotting was restricted in the BMDU in the 24-hour γH2AX biomarker response. *x*-axis represents Log10 concentration (μM). Figure adapted from Wheeldon et al. (2020).
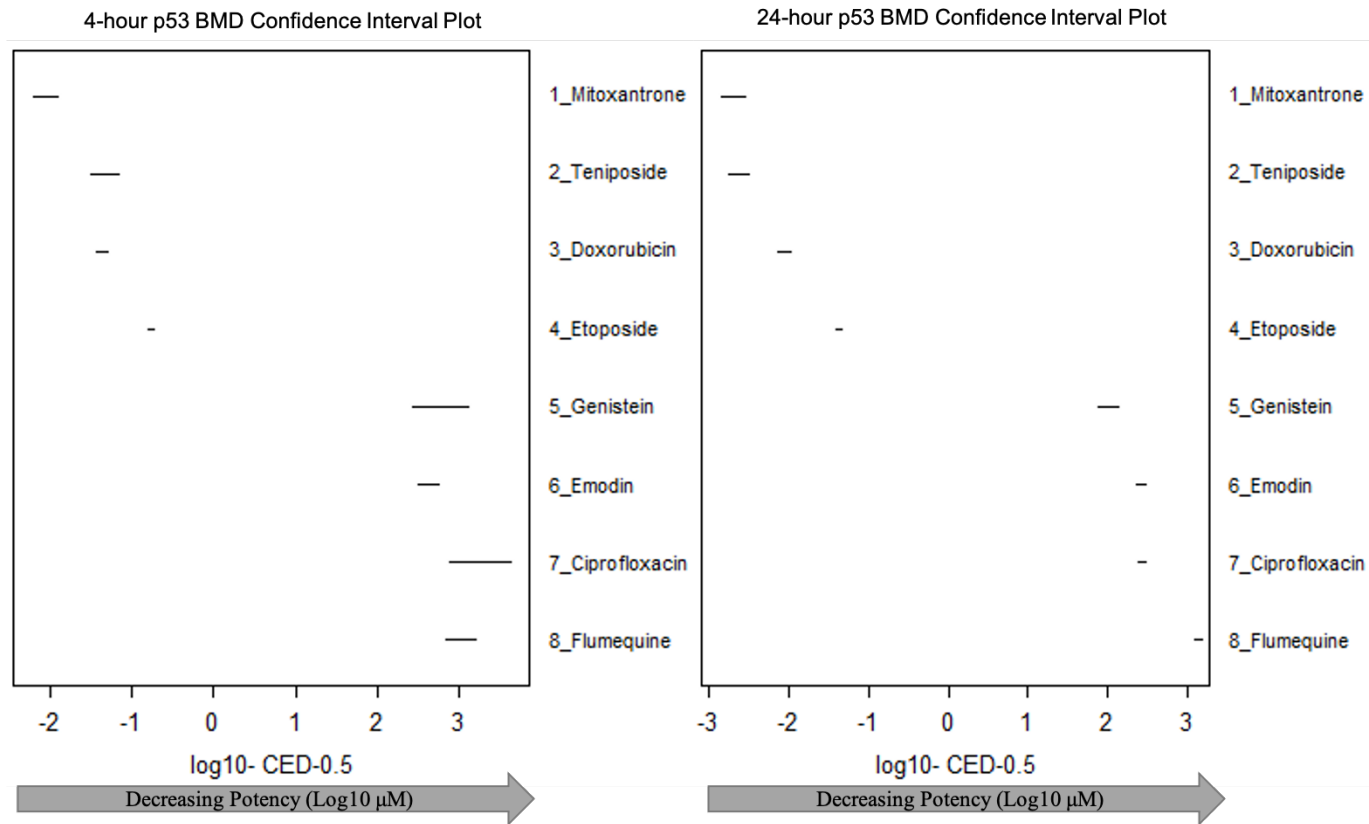
Figure 5.17. BMD confidence interval plots for the p53 biomarker responses. Confidence intervals are plotted in potency rank order with decreasing potency from left to right. *x*-axis represents Log10 concentration (μM). There were very small bound confidence intervals for most compounds. Flumequine's BMDU was disproportionally larger, so plotting was restricted in the BMDU in the 24-hour p53 biomarker response. Figure adapted from Wheeldon et al. (2020).
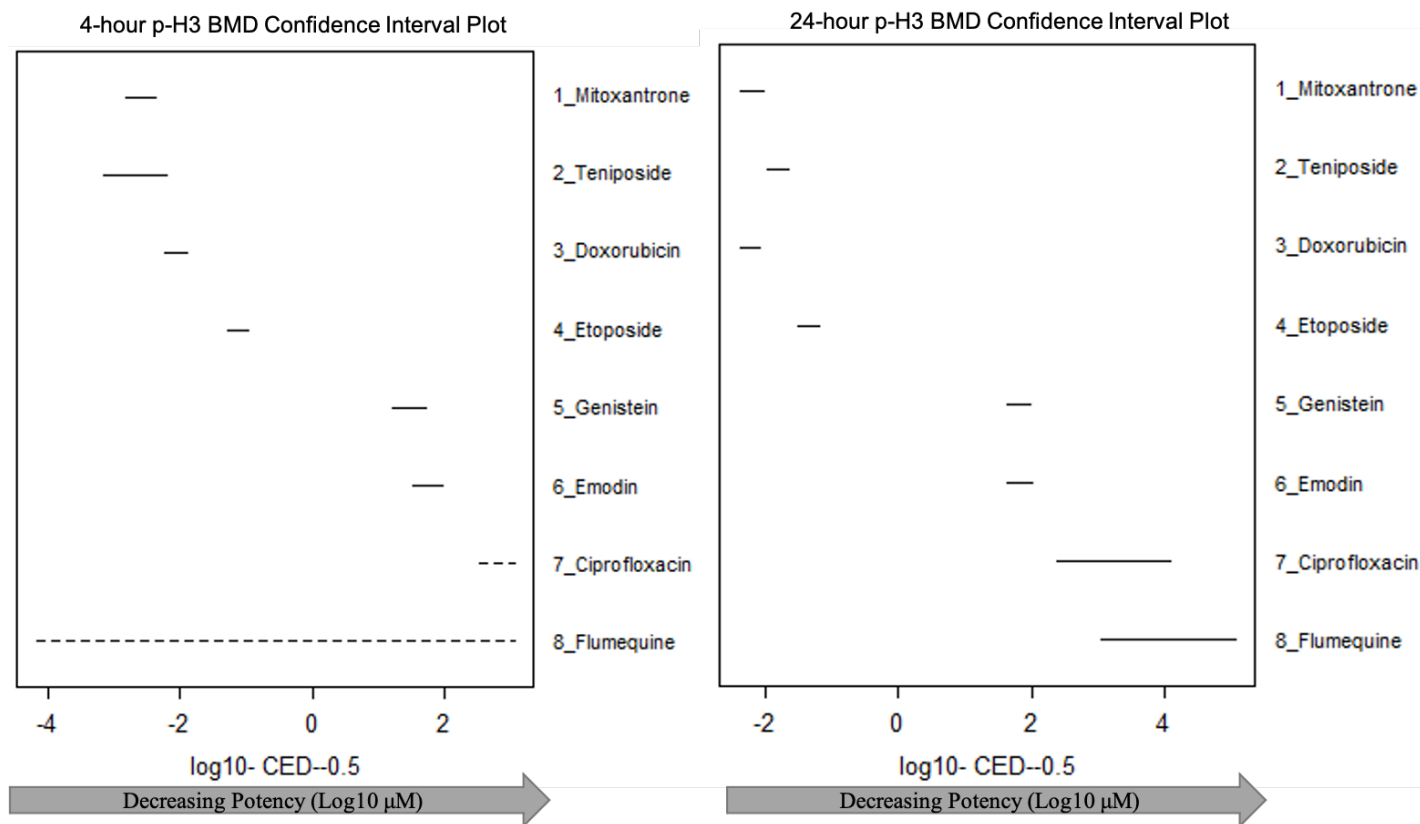
Figure 5.18. BMD confidence interval plots for the p-H3 biomarker responses. Confidence intervals are plotted in potency rank order with decreasing potency from left to right. *x*-axis represents Log10 concentration (μM). There were very small bound confidence intervals for most compounds. Unbound confidence intervals for ciprofloxacin and flumequine shown as dashed lines in the 4-hour p-H3 plot. Figure adapted from Wheeldon et al. (2020).

5.3.4 BMD Analyses ToxTracker Biomarker Responses

Exponential model 5 was selected as the best model to describe the dose-responses of the ToxTracker GFP-biomarker responses after exposure of mES cells to the Topo-II Poison compounds tested in the system. The exponential model 5 contains the full complement of dose-response model parameters consistent with other genotoxicity dose-response analyses reported in the literature (Slob and Setzer, 2014; EFSA, 2017). In general, excellent dose-response curves were returned for most compounds and ToxTracker biomarkers responsive to DNA damage. The exception was doxorubicin whose 2nd concentration above the control seemed to be an outlier (**Figure 5.19**). BMD model dose-response curves (endpoint specific CES and CES 0.5) for the Topo-II Poisons tested in ToxTracker DNA damage response biomarkers Rtkn-GFP, Btg2-GFP, and Bscl2-GFP are shown in **Figure 5.20**, **Figure 5.21**, and **Figure 5.22**, respectively.
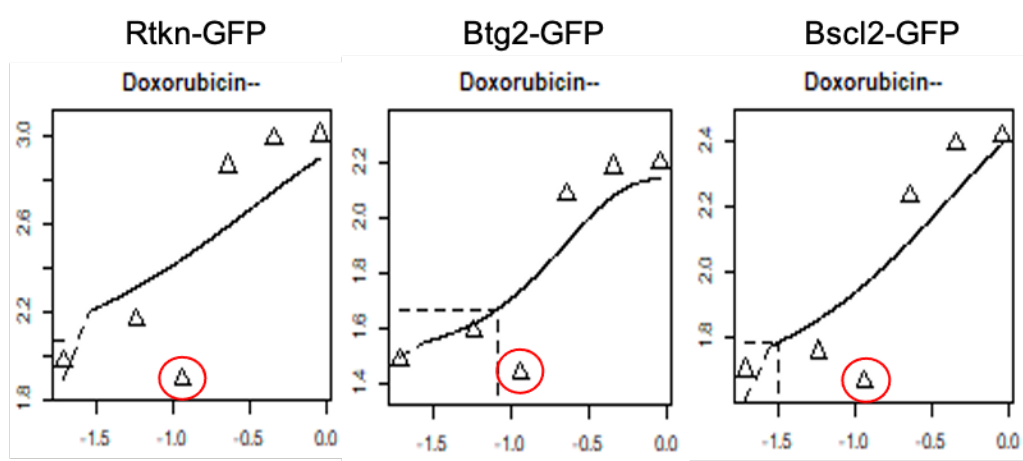
Figure 5.19. Dose-response curves for doxorubicin in the Rtkn-GFP, Btg2-GFP, and Bscl2-GFP biomarker responses showing the 2nd concentration above the control outlier circled in red.

The dose-responses responded in the direction expected upon exposure to Topo-II Poisons. I.e., the Rtkn-GFP biomarker was induced indicative of DNA double strand breaks, the Btg2-GFP biomarker was induced due to p53 activation and cell cycle delay, and the Bscl2-GFP biomarker was induced from because of ATR recruitment to checkpoint 1 inducing mitotic arrest. The direction of the dose-response curves for each biomarker is evident in **Figure 5.20**, **Figure 5.21**, and **Figure 5.22**. Individual dose-response curves for each level (compound) of the covariate for each biomarker can be found in Appendix 7.

Individual Dose-Response Curves for the MultiFlow and ToxTracker System Biomarker Responses Resulting from Topo-II Poison Exposure
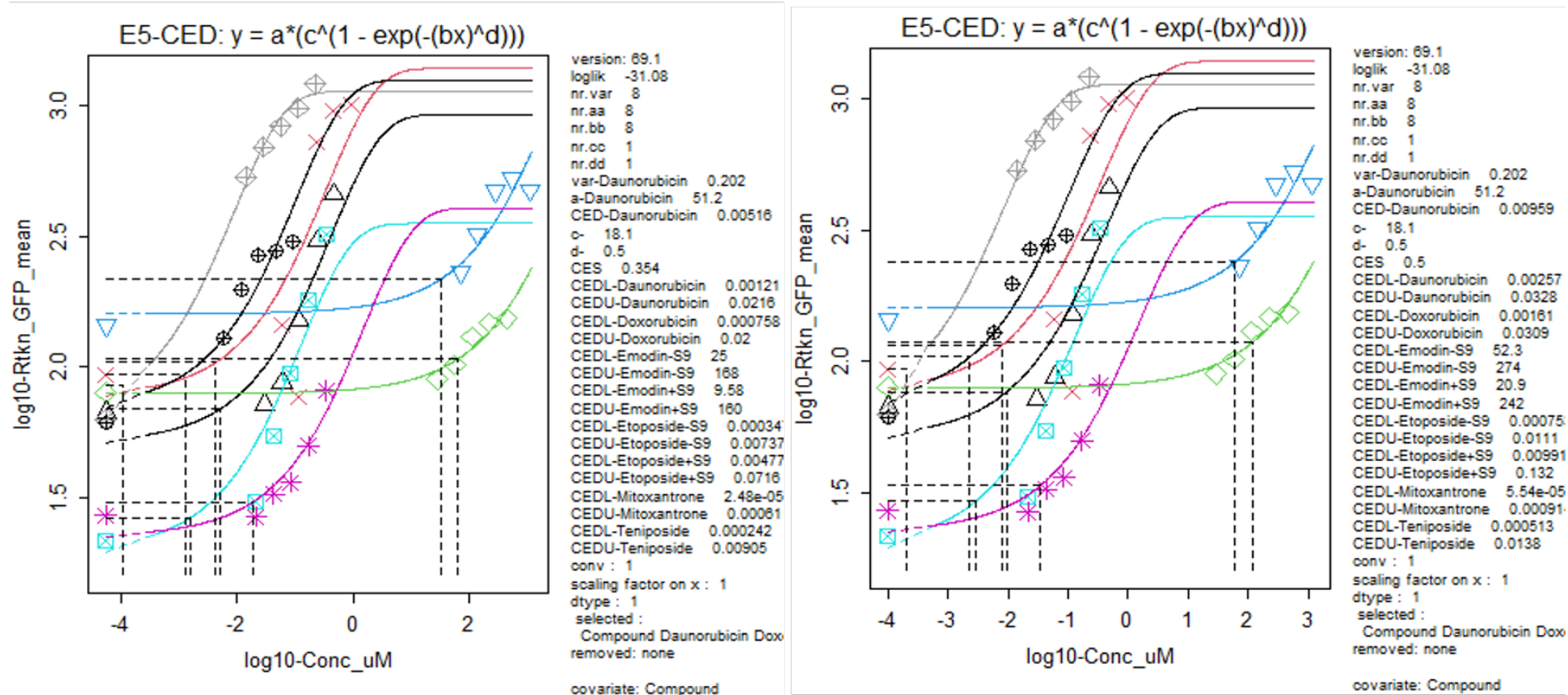
Figure 5.20. BMD dose-response model for the Topo-II Poisons exposure to mES cells for 24-hour and evaluated for Rtkn-GFP expression using the ToxTracker assay. Induction (increase) of the Rtkn-GFP response is indicative of DNA double strand breaks from Topo-II Poison exposure. Left panel endpoint specific CES of 0.354. Right panel fixed CES 0.5.
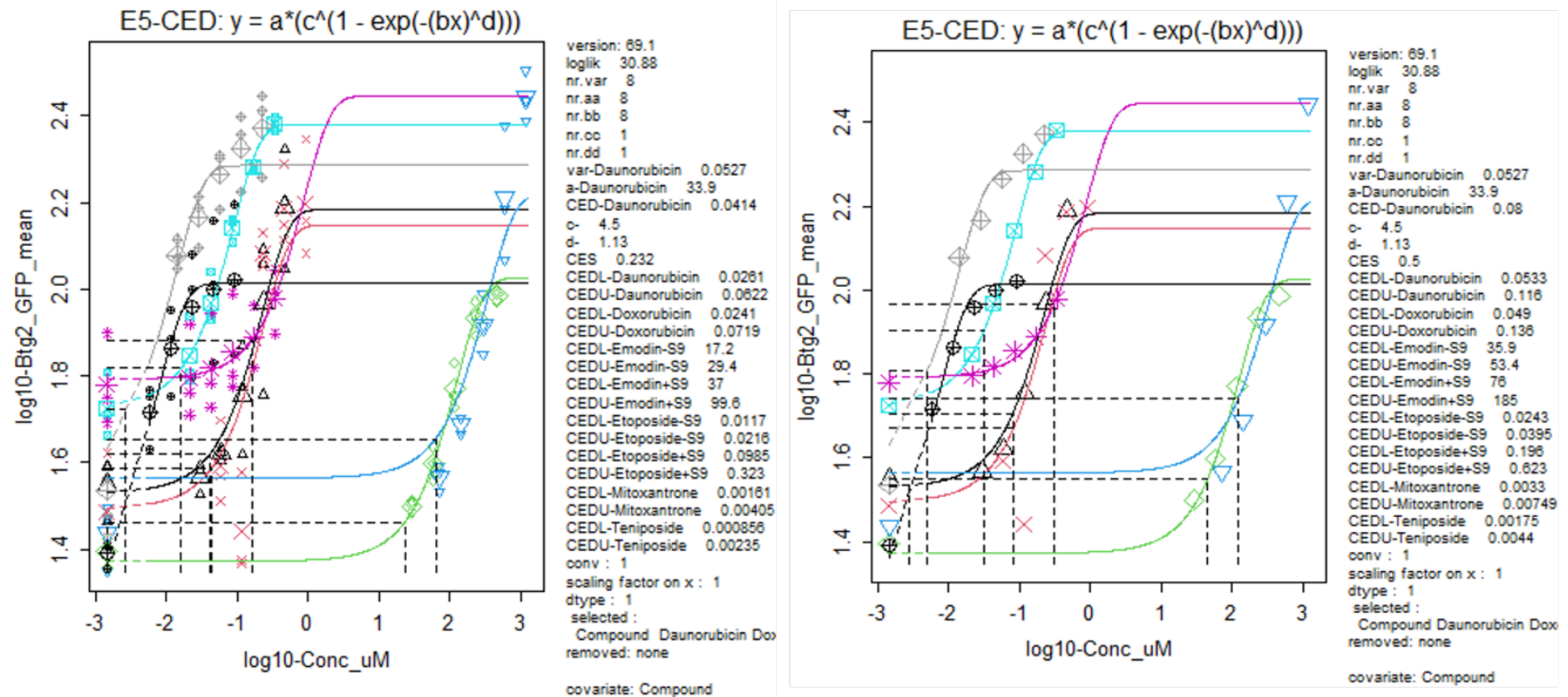
Figure 5.21. BMD dose-response model for the Topo-II Poisons exposure to mES cells for 24-hour and evaluated for Btg2-GFP expression using the ToxTracker assay. Induction (increase) of the Btg2-GFP response is indicative of p53 recruitment and cell cycle delay from Topo-II Poison exposure. Left panel endpoint specific CES of 0.232. Right panel fixed CES 0.5.
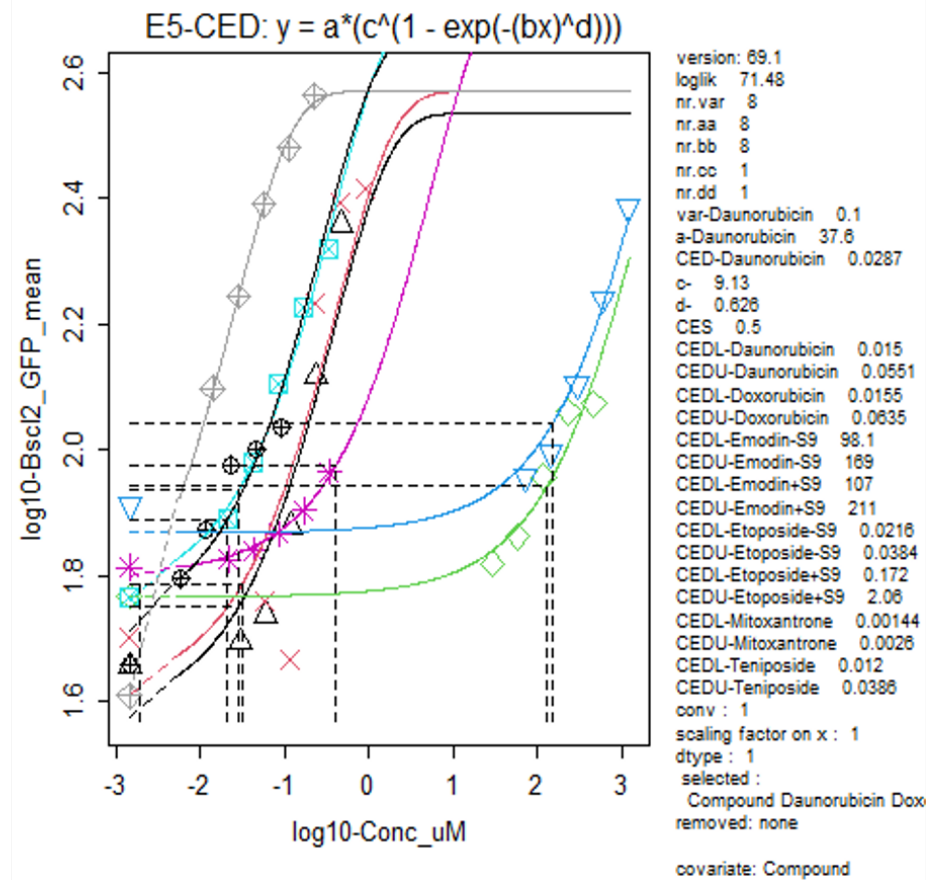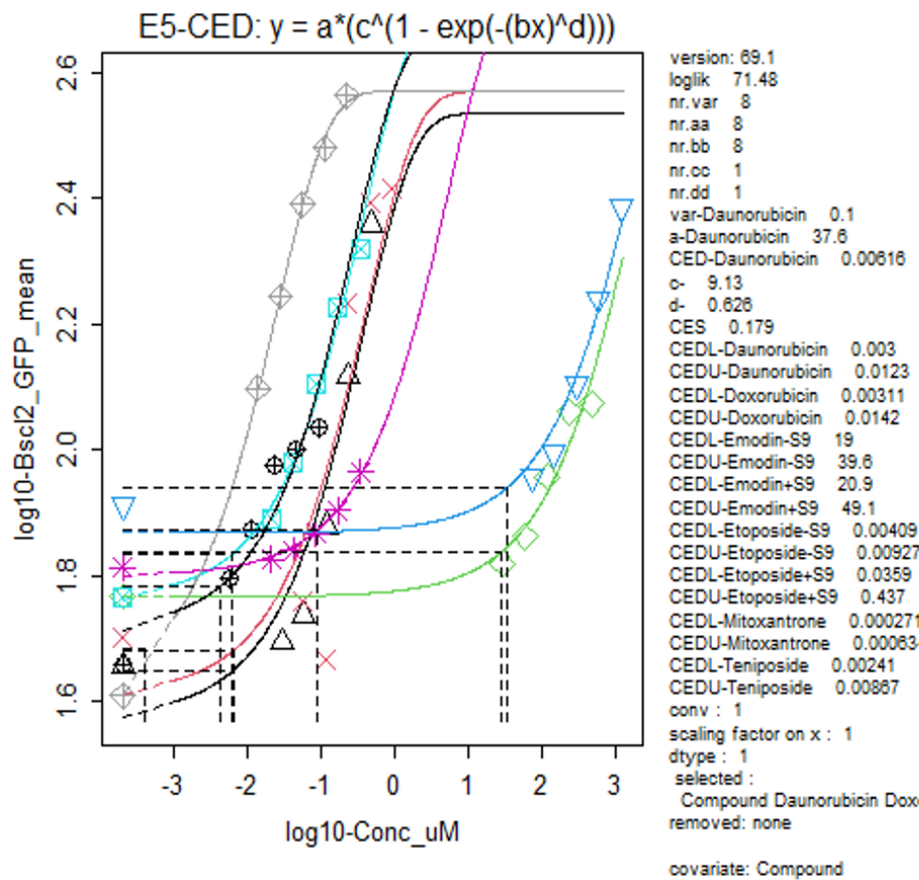
Figure 5.22. BMD dose-response model for the Topo-II Poisons exposure to mES cells for 24-hour and evaluated for Bscl2-GFP expression using the ToxTracker assay. Induction (increase) of the Bscl2-GFP response is indicative of stalled replication forks recruiting ATR for checkpoint 1 control. from Topo-II Poison exposure. Left panel endpoint specific CES of 0.179. Right panel fixed CES 0.5.

5.3.4.1 Marker Specific CES Hypothesis Testing

The marker specific CESs calculated by PROAST for the ToxTracker biomarker responses were 0.354 (35.4%) for Rtkn-GFP, 0.232 (23.3%) for Btg2-GFP, and 0.179 (17.9%) for Bscl2-GFP. BMD confidence interval plots were automatically generated by PROAST v69.1 and arranged the BMD confidence intervals by Log midpoint with decreasing potency from left to right. The BMD confidence intervals of each biomarker with a marker specific CESs was compared with those derived from a CES 0.5 (50%). **Figure 5.23** shows the comparison of BMD confidence intervals for Rtkn-GFP, **Figure 5.24** for Btg2-GFP, and **Figure 5.25** for Bscl2-GFP.
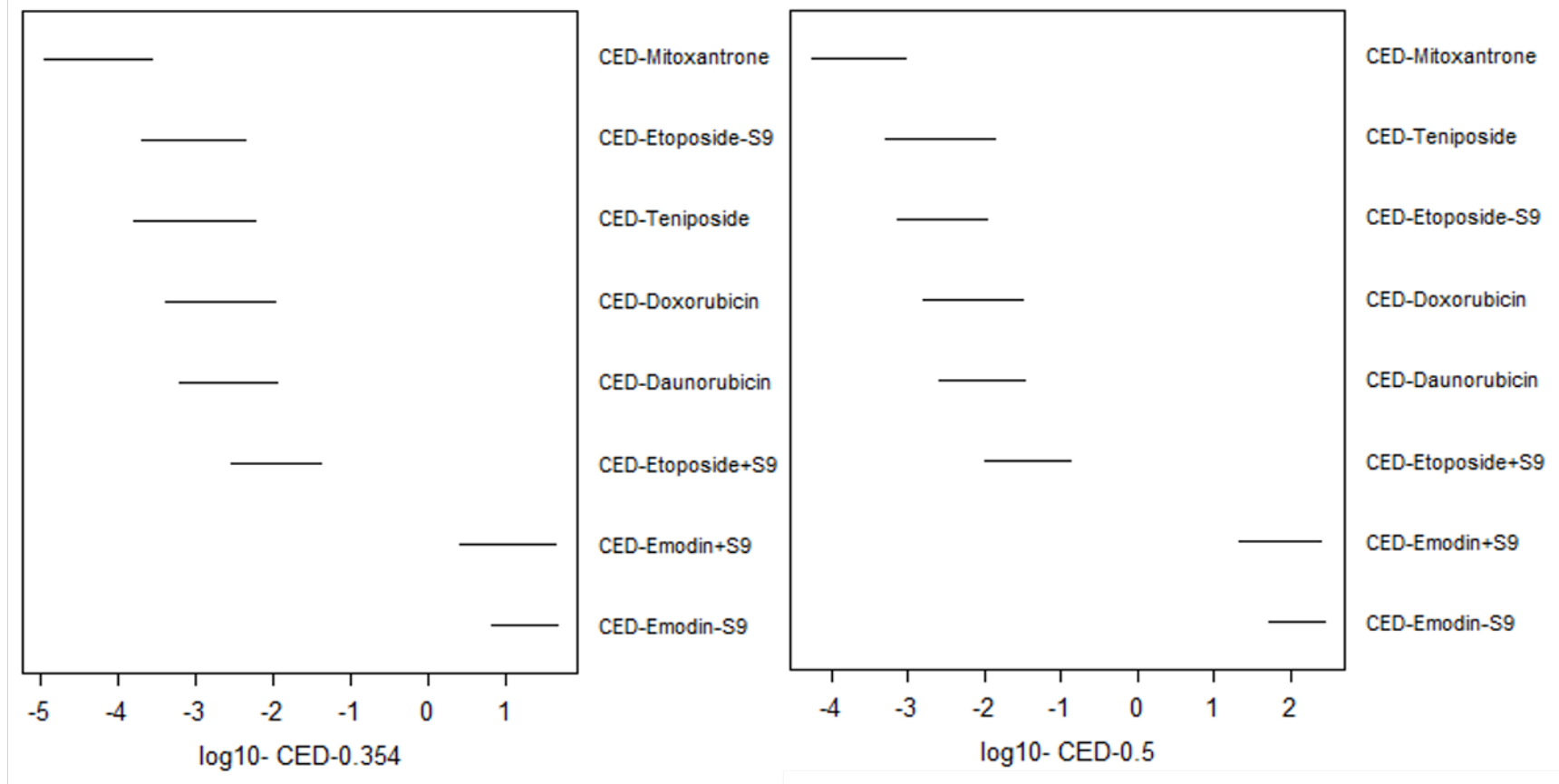
Figure 5.23. BMD confidence interval plots for the Rtkn-GFP biomarker response with a marker specific CES (left panel) versus a fixed CES of 0.5 (50%) (right panel).
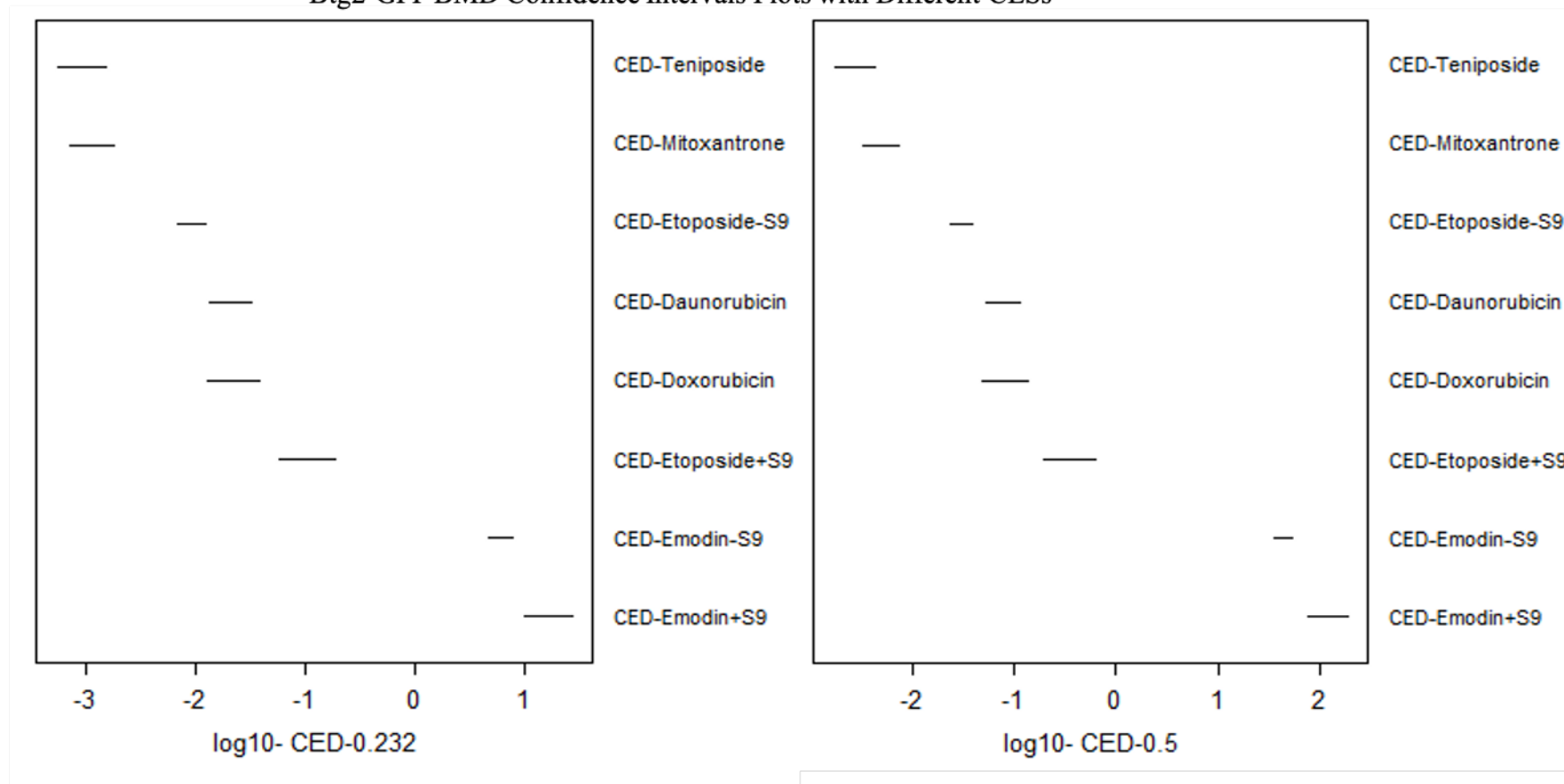
Figure 5.24. BMD confidence interval plots for the Btg2-GFP biomarker response with a marker specific CES (left panel) versus a fixed CES of 0.5 (50%) (right panel).

Figure 5.25. BMD confidence interval plots for the Bscl-2-GFP biomarker response with a marker specific CES (left panel) versus a fixed CES of 0.5 (50%) (right panel).

5.3.4.2 BMD Confidence Interval Potency Ranks

BMD confidence interval plots were generated by PROAST v69.1 arranged by decreasing potency *via* confidence interval Log midpoint for the ToxTracker biomarkers responsive to DNA damage. Each biomarker response BMD confidence interval plots are presented in **Figure 5.26**. Confidence intervals were generated using CES 0.5 (50%) for all biomarker dose-responses to be consistent with the MultiFlow endpoint BMD analyses for further data analysis and comparative results interpretation.

BMD confidence intervals were generally wider than those obtained from the MultiFlow assay's biomarker responses, with maximum confidence interval width in the order of 1-Log.
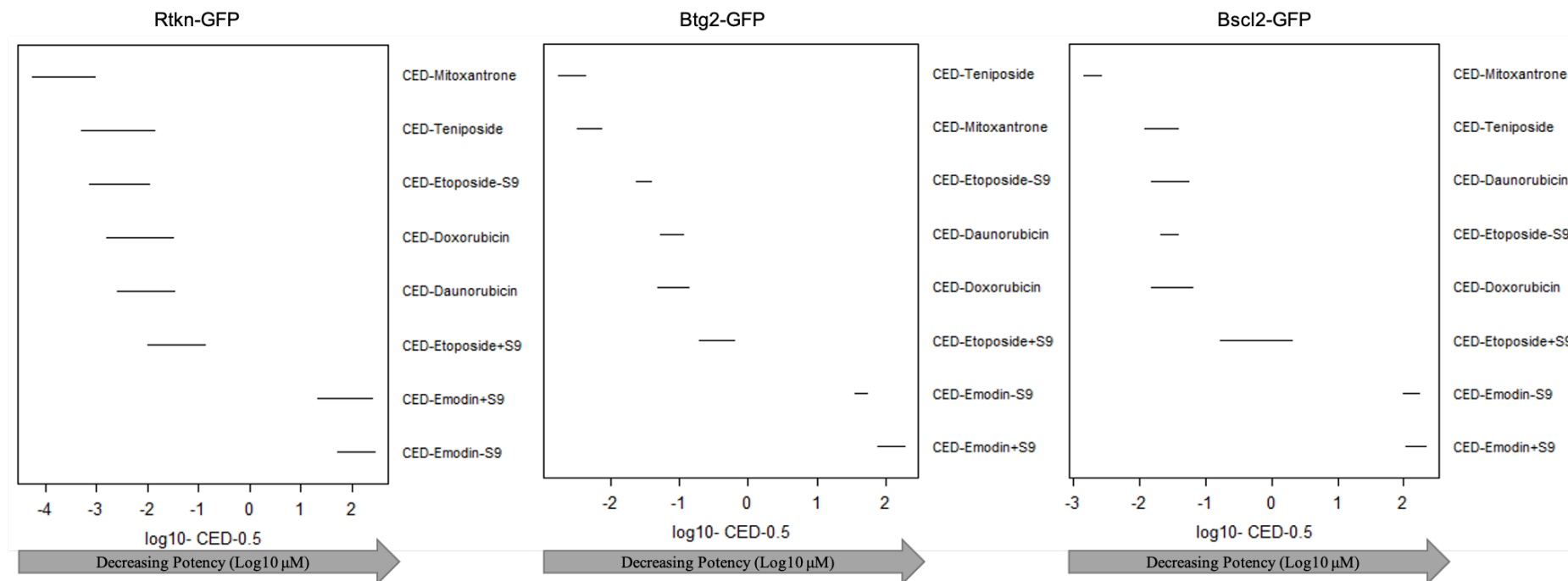
Figure 5.26. BMD confidence interval plots for the ToxTracker biomarker responses specific to DNA damage. Confidence intervals are plotted in potency rank order with decreasing potency from left to right. There were bound confidence intervals for all compound and biomarker combinations with maximum span of approximately 1 order of magnitude. *x*-axis represents Log10 concentration (µM).

5.3.5 Mechanistic Biomarker Cross-Correlations

The confidence intervals from compounds assessed for similar mechanistic responses to DNA damage in the MultiFlow and ToxTracker assays were correlated based on the *a-priori* knowledge of Topo-II Poison genotoxic mechanisms of action, including:

- DNA double strand breaks: MultiFlow γH2AX induction versus ToxTracker Rtkn-GFP expression **Figure 5.27**.
- p53 dependent cell cycle delay: MultiFlow p53 induction versus ToxTracker Btg2-GFP expression **Figure 5.28**.
- Stalled replication fork cell cycle delay/mitotic arrest: MultiFlow p-H3 downregulation versus ToxTracker Bscl2-GFP expression (**Figure 5.29**).
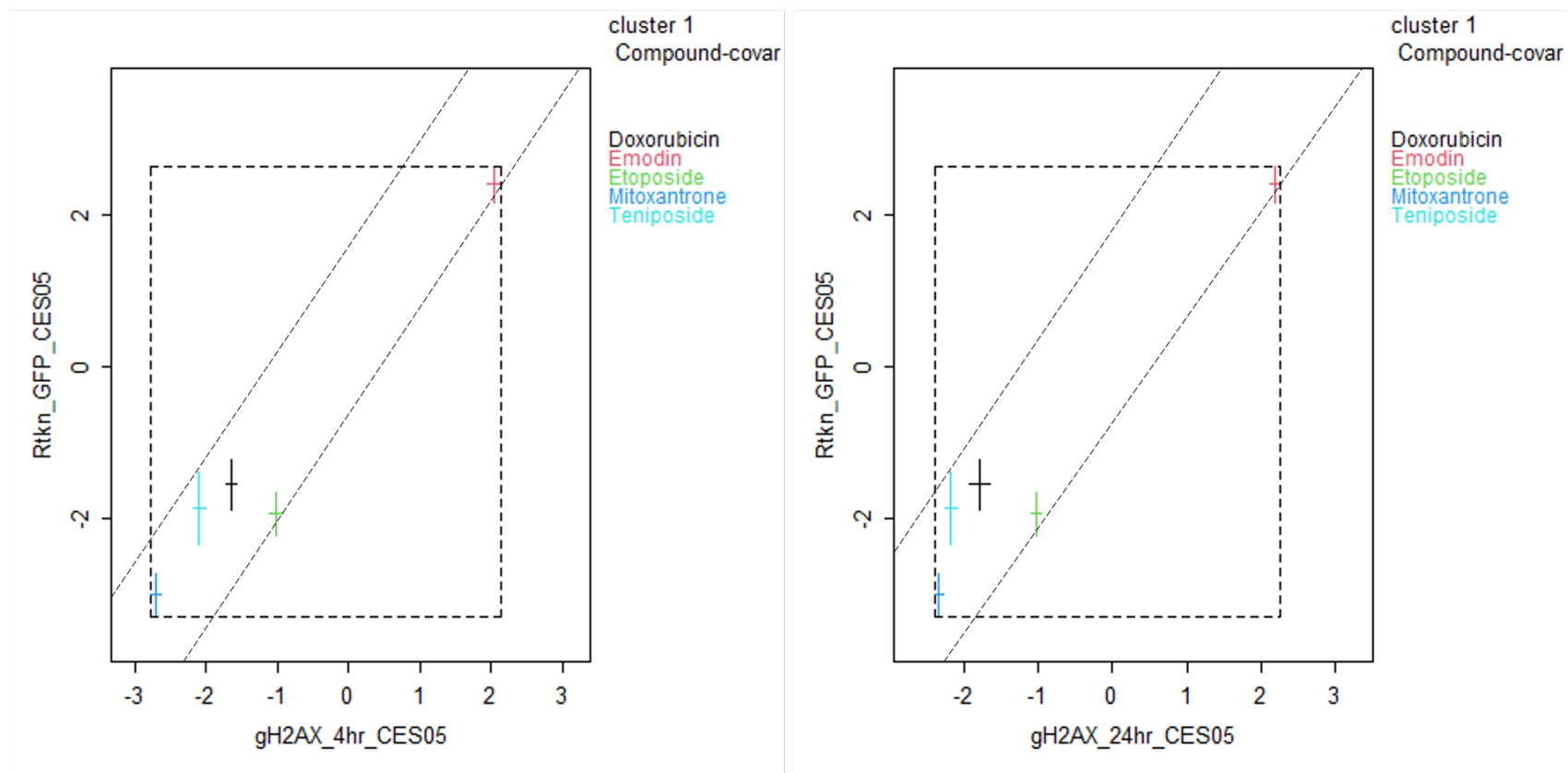
Figure 5.27. Cross-correlation plots for the same Topo-II Poisons tested in the MultiFlow and ToxTracker assay biomarkers indicative of DNA double strand breaks. The *x* and *y* axes represent Log10 concentration of compounds in μM. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs. All CIs lie within the dashed lines showing a magnitude of approximately 1-Log. Left Panel: 4-hour γH2AX versus Rtkn-GFP responses. Right Panel: 24-hour γH2AX versus Rtkn-GFP responses. Abbreviation: BMD = Benchmark Dose, CI = Confidence Interval.
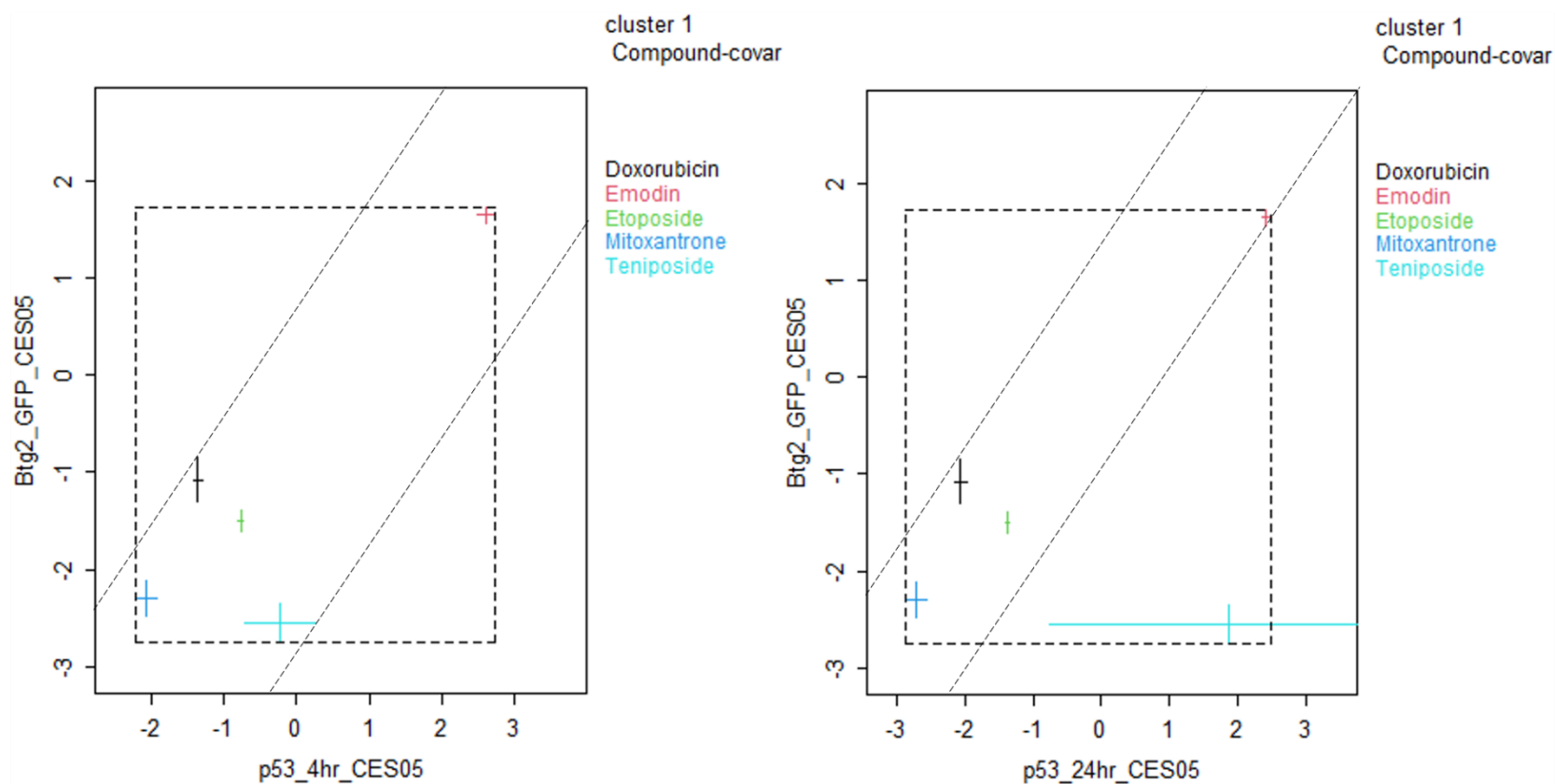
Figure 5.28. Cross-correlation plots for the same Topo-II Poisons tested in the MultiFlow and ToxTracker assay biomarkers indicative of p53 dependent cell cycle delay. The *x* and *y* axes represent Log10 concentration of compounds in μM. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs except outliers. Most CIs lie within the dashed lines except for teniposide which is an outlier in the 24-hour p53 biomarker dose-response. Correlation of between 1 and 2 orders of magnitude for the biomarkers. Left Panel: 4-hour p53 versus Btg2-GFP responses. Right Panel: 24-hour p53 versus Btg2-GFP responses. Abbreviation: BMD = Benchmark Dose, CI = Confidence Interval.
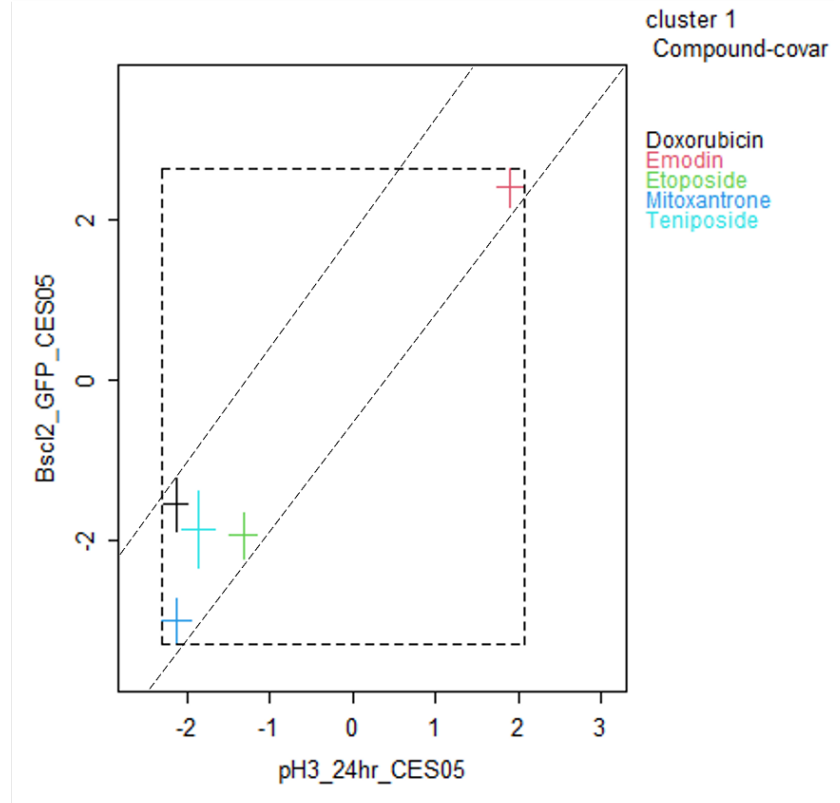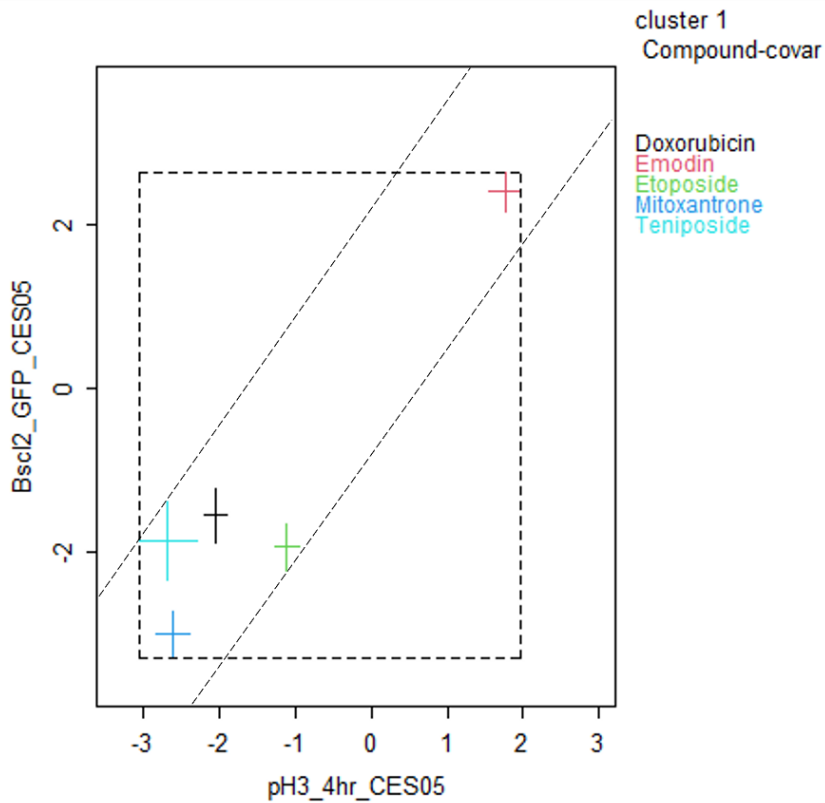
Figure 5.29. Cross-correlation plots for the same Topo-II Poisons tested in the MultiFlow and ToxTracker assay biomarkers indicative of stalled replication forks and mitotic arrest. The *x* and *y* axes represent Log10 concentration of compounds in μM. The dashed parallel lines were drawn to encompass most of the BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs. The dashed square box default PROAST output encompassing finite BMD CIs. All CIs lie within the dashed lines showing a magnitude of approximately 1-2-Log. Left Panel: 4-hour p53 versus Btg2-GFP responses. Right Panel: 24-hour p53 versus Btg2-GFP responses. Abbreviation: BMD = Benchmark Dose, CI = Confidence Interval.
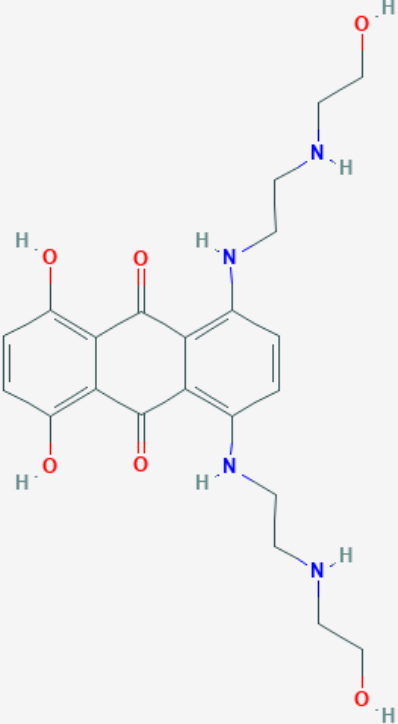
5.3.6 Derek Nexus Structural Activity Knowledge

Derek Nexus was searched to find compound structural features that may relate to genotoxic potency. The alert descriptions and structural information were found in the Derek Nexus database relating to the individual compounds. The information permitted the grouping of Topo-II Poison compounds into the following 4 different sub-classes: Anthraquinones (mitoxantrone, daunorubicin, doxorubicin); podophyllotoxins (teniposide and etoposide); quinolones (ciprofloxacin and flumequine) and an isoflavone (genistein).
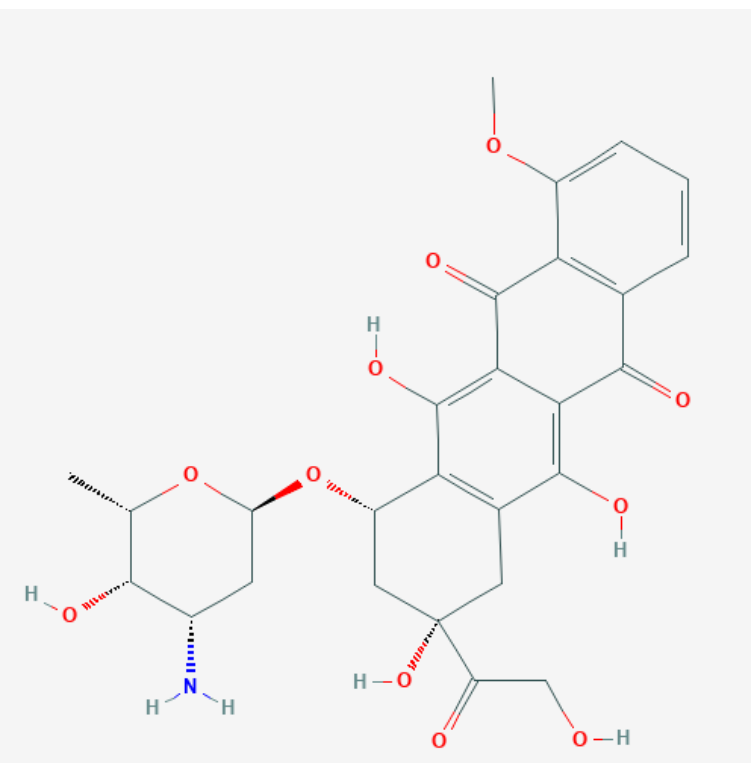
The Topo-II Poison compound sub-classes, Derek Nexus alerts and compound structures are presented in **Table 5.5**.
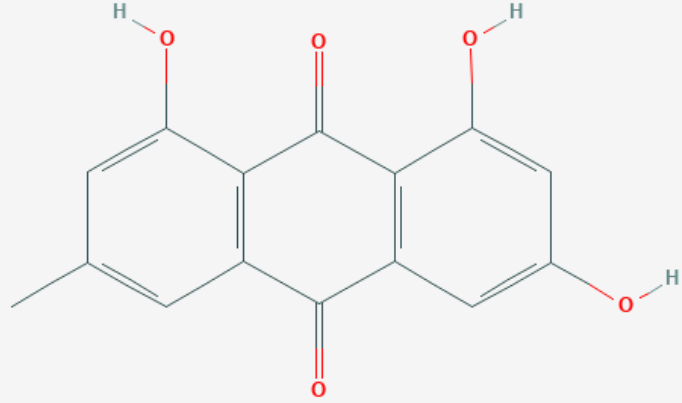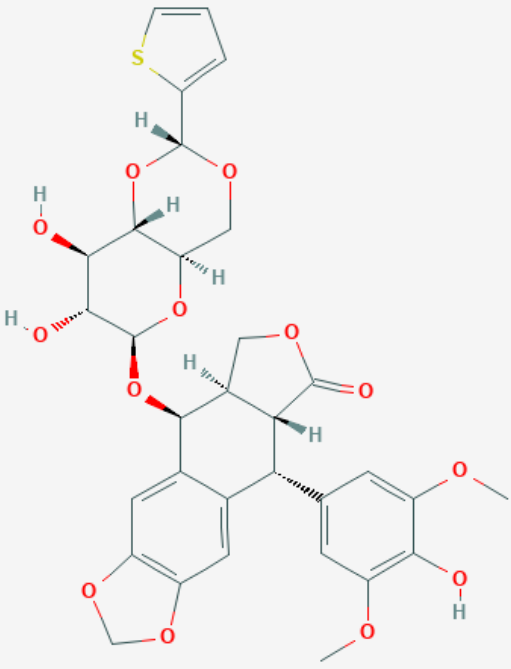
Table 5.5. Topo-II Poison Compound Subclassification and Derek Nexus Alert Descriptions
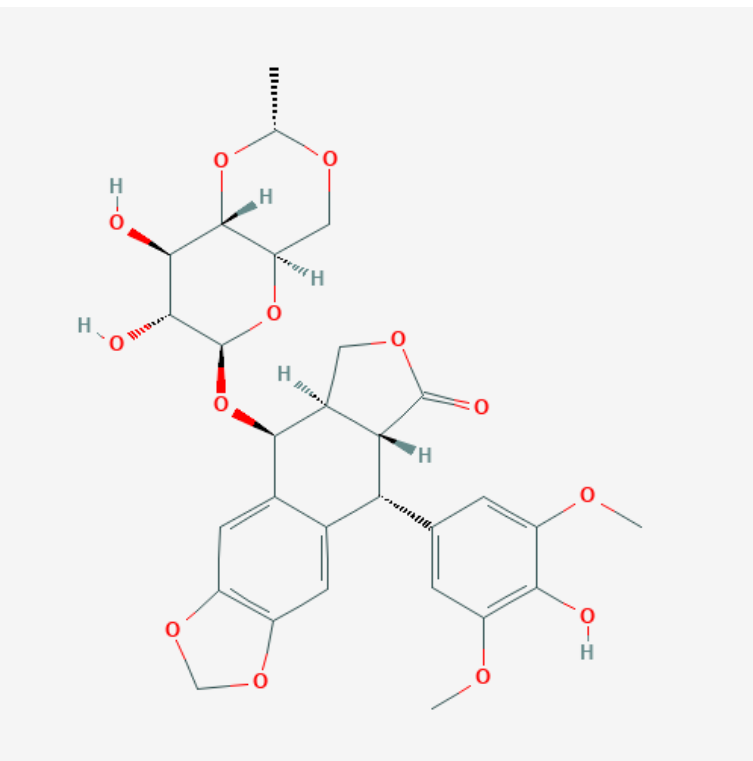
Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Anthraquinone	Mitoxantrone	Hydroxylated anthraquinone/Hydroquinone derivative; Ethanolamine	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Anthraquinone	Daunorubicin	Hydroxylated anthraquinone/Hydroquinone derivative; Anthracycline; Alkyl aldehyde or precursor.	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Anthraquinone	Doxorubicin	Hydroxylated anthraquinone and/or derivative; Anthracycline	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Anthraquinone	Emodin	Hydroxylated anthraquinone or precursor; Anthraquinone derivative	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Podophyllotoxin	Teniposide	Podophyllotoxin or analogue; Thiophene	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Podophyllotoxin	Etoposide	Podophyllotoxin or analogue	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Quinolone	Ciprofloxacin	Halogenated quinolone-3-carboxylic acid	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Quinolone	Flumequine	Quinolone-3-carboxylic acid	

Compound Sub-Class	Topo-II Poison Compound	Derek Nexus Alert Description	Compound Structure
Isoflavone	Genistein	4',7-Dihydroxyisoflavone or derivative; Enol ether	

5.3.7 Topo-II Poison Potency Fluorescent Anisotropy Assay

The summary statistics for the same Topo-II Poison compounds studied in the fluorescent anisotropy assay as the MultiFlow and ToxTracker assays reported in this chapter were exported from the publication by Shapiro and Austin (2014). The summary statistics (mean and standard deviation) as they appear in Shapiro and Austin (2014) are replicated in **Table 5.6**.

Table 5.6. Inhibitory potencies (IC$_{50}$s) versus human Topo-IIa and β by 7 Topo-II Poisons common in the Shapiro and Austin (2014) publication to those studies in the MultiFlow and ToxTracker assays.

Topo-II Poison Compound	Topo-IIα IC$_{50}$ (μM)	Topo-IIβ IC$_{50}$ (μM)
Daunorubicin	1.6 ± 0.2	1.1 ± 0.2
Doxorubicin	1.5 ± 0.2	1.1 ± 0.3
Etoposide	8.1 ± 3.0	15 ± 6.9
Teniposide	1.4 ± 0.3	3.7 ± 1.4
Mitoxantrone	1.0 ± 0.5	0.5 ± 0.2
Genistein	4.1 ± 1.0	3.5 ± 1.0
Ciprofloxacin	110 ± 25	111 ± 15

Means ± standard deviations of 3 replicates.

The summary statistics were Log10 transformed to approximate to a normal distribution and the Log10 values of the standard deviations were used to calculate standard errors. Log10 means and standard errors were plotted in order of Log10 midpoint (mean) (**Figure 5.30**).
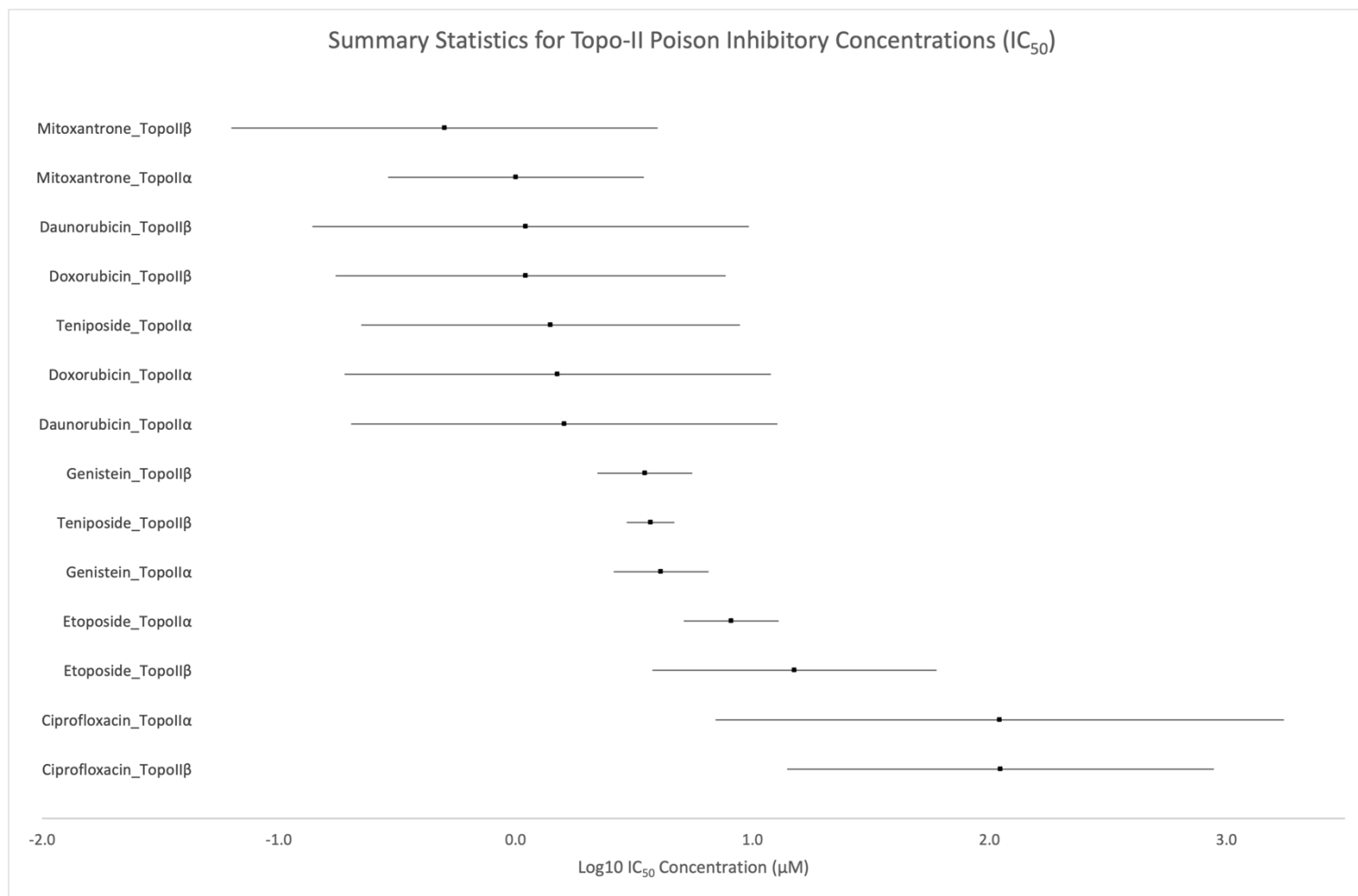
Figure 5.30. Log10 mean inhibitory (IC$_{50}$s) potencies and standard errors plotted for the matching 7 Topo-II Poison compounds reported in (Shapiro and Austin, 2014). Log10 means shown with the black square datapoints. Uncapped standard errors plotted in the horizontal direction. *x*-axis Log10 IC$_{50}$ concentration in µM units. Compounds and Topo-II enzyme isomer plotted on the *y*-axis.

5.4 Discussion

5.4.1 Topo-II Poison BMD Potency and Assay Mechanistic Information

In general, the BMD dose response analyses results reported in this results chapter demonstrated that the MultiFlow and ToxTracker assays can generate high quality potency information for classes of compounds of interest. In this instance, the Topo-II compounds were tested due to their unique genotoxic mechanisms of action. The BMD dose-responses were high quality with very small confidence intervals (approximately 0.25-0.50 orders of magnitude) for most compounds MultiFlow assay biomarker responses – with ciprofloxacin and flumequine being the exceptions. Ciprofloxacin and flumequine yielded a combination of no dose-response or varying degrees of scatter in the BMD analyses of the MultiFlow biomarker dose-response as shown in **Figure 5.6**, **Figure 5.7**, **Figure 5.8**, and **Figure 5.9**. Instances where ciprofloxacin and flumequine dose-responses exhibited scatter resulted in wider confidence intervals compared with other compounds with better dose-response relationships, which relates to the underlying uncertainty in the dose-response datasets for the same compounds. It is suggested that the uncertainty may be reduced to some extent with repeat experiments, however these compounds demonstrated the lowest potency among all the tested Topo-II Poisons in these experiments, which also relates to the efficacy of the MultiFlow tests system with low sensitivity to low potent compounds.

The ToxTracker assays GFP-biomarker dose-responses also responded well to BMD modelling with good quality small confidence intervals (approximately 1 order of magnitude) being returned for all Topo-II Poison compounds across all DNA damage response GFP-biomarkers. Although the confidence intervals were larger than those obtained for the same compounds in the MultiFlow biomarker responses, 1-Log is still considered good quality and consistent with other good quality BMD confidence intervals reported in the literature (Soeteman-Hernández et al., 2015; 2016). The most plausible explanation for the comparative difference in confidence interval width between the two systems is the lower number of concentration levels tested in the ToxTracker system compared to the MultiFlow assay – where only 5 concentrations levels of each Topo-II Poison compound were included for exposure to the mES cells in the ToxTracker assay, compared to the range of 11-20 Topo-II Poison compound concentration levels exposed to the TK6 cells in the MultiFlow assay. Increasing the number of concentration levels results in more precise BMD estimates (Wills et al., 2016b).

Exposure to Topo-II Poisons results in DNA double strand breaks, stabilisation of the Topo-II-DNA cleavage complex resulting in replication fork collapse, cell cycle delay, and mitotic arrest (Berger, 1998; Lynch et al., 2003; Deweese and Osheroff, 2008). In the same respect, the MultiFlow and ToxTracker assays responded as expected regarding the Topo-II Poison mechanism of action. Increases in DNA double strand breaks was associated with γH2AX induction in the MultiFlow assay, and Rtkn-GFP activation in the ToxTracker assay. Nuclear p53 translocation was associated with an increase in p53 signal in the MultiFlow assay and accompanied with Btg2-GFP induction in the ToxTracker assay. Btg2 is a component of the p53-dependent DNA damage response and is involved in regulation of the G1/S cell cycle checkpoint. A normally dividing cell would undergo mitosis as the cell cycle progresses, with accompanied phosphorylation of histone H3 in late G2 phase (Hans and Dimitrov, 2001). With a cell cycle that has been delayed due to Topo-II Poison activity, histone H3 phosphorylation would be expected to be downregulated due to mitotic arrest. This has been demonstrated in the MultiFlow assay where decreases in p-H3 signals were evident after exposure to Topo-II Poisons. The Topo-II Poison compounds form a unique response in the MultiFlow assay that does not align with the typical mode of action signals that the assay was developed for. For example, classical clastogens typically result in strong induction signals in the γH2AX and p53 endpoints, with either no response or modest reduction in p-H3 responses. Aneugens respond with dramatic increases in the p-H3 and p53 biomarkers with no response in the γH2AX biomarker (Bryce et al., 2016). Thus, a compound mode of action call is made based on these prototypical MultiFlow biomarker signal responses. In the case of Topo-II Poisons, strong inductions of γH2AXs and p53 biomarkers corresponding to clastogen mode of action is accompanied with a strong downregulation of p-H3 – which is a departure from the prototypical signals used to dichotomise compound mode of action in the MultiFlow assay. In the ToxTracker system, exposure to Topo-II Poison compounds resulted in an increase in Bscl2-GFP expression which follows ATR recruitment to DNA lesions resultant from replication stress (Hendriks et al., 2015). In the case of Topo-II Poison mechanism of action, replication fork collapse is the plausible explanation for the increase in Bscl2-GFP expression observed in these experiments. The BMD confidence intervals plots reported in this chapter (sections 5.3.3 and 5.3.4.2) and published in Wheeldon et al. (2020) were the first instances where Topo-II Poison genotoxicity was related to potency based on mechanistic biomarker responses for a specific compound class mechanism of action quantitative risk assessment. The cross-

correlation plots show that compound potencies are comparable across mechanistic biomarkers responsive to the same types of DNA damage.

5.4.2 Impact of CES Choice

The choice of CES was demonstrated to be inconsequential to the BMD confidence interval potency rank orders obtained in the MultiFlow and ToxTracker datasets BMD analyses. This finding is consistent with other researchers that have studied varying values of CES in other test systems, namely the micronucleus endpoint (Bemis et al., 2016). The main reason for testing different CES values in the MultiFlow assay was as a proof-of-concept to rule out the influence of CES on the BMD potency plot rank order.

Since there was limited prior published experience of BMD modelling the ToxTracker biomarker responses, it was deemed imperative to also determine the influence of CES on the resulting BMD potency rank order in said test system. The newer method of calculating an endpoint (or marker) specific CES based on within-group variance (Slob, 2017) was selected for the ToxTracker BMD dose-response analysis. An arbitrary CES of 0.5 (50%) was also selected for BMD modelling of the ToxTracker biomarker dose-responses consistent with the MultiFlow assay. Again, the choice of a marker specific CES or 0.5 (50%) was inconsequential on the BMD confidence interval potency rank order. Thus, for all further analyses, a CES of 0.5 (50%) was kept constant across all biomarker BMD response analyses.

5.4.3 Compound Structural Information Supports BMD Potency Ranks

Toxicologists often rely on *in silico* system results to augment their expert judgement. Derek Nexus is one of the gold-standard *in silico* systems used in industry utilising a strong knowledge based on compound structural alerts that relate to toxicity (Barber et al., 2017). The Derek Nexus structural alerts presented in **Table 5.5** were given that relate to toxicity, especially for inducing chromosome damage which was outlined in the Derek Nexus "reasoning summary" results. The alert descriptions permitted the grouping of the Topo-II compounds into 4 sub-classes: 1) Anthraquinones (mitoxantrone, daunorubicin, doxorubicin); podophyllotoxins (teniposide and etoposide); quinolones (ciprofloxacin and flumequine) and an isoflavone (genistein). This was particularly useful to provide an unbiased classification of compounds, which would be valuable to a non-chemist.

Based on the same application of compound structure and genotoxic mechanism information to account for potency as Wills et al. (2016a) did for the benzimidazole compounds, the Derek Nexus alerts and MultiFlow and ToxTracker mechanistic BMD potency ranks was applied to the Topo-II Poison compounds assessed in these experiments. Confidence interval rank orders in **Figure 5.16**, **Figure 5.17**, **Figure 5.18**, and **Figure 5.26** were scrutinised together with the Derek Nexus structural alerts and a literature search for supporting information. In general, the potent chemotherapeutic compound anthraquinone mitoxantrone is the most potent Topo-II Poison compound tested, and the quinolone antibiotic compounds ciprofloxacin and flumequine were the least potent.

Of the podophyllotoxin derivatives, teniposide was consistently more potent than etoposide in all MultiFlow biomarkers, and the Btg2-GFP and Bscl2-GFP ToxTracker biomarker responses. Both compounds were equipotent for DNA double strand breaks measured by Rtkn-GFP activation in the ToxTracker biomarker. Teniposide and etoposide contain four planar fused rings with fundamental Topo-II binding activity dependent on a 4-hydroxy group ring on the chiral centre (Long, 1992). Evidence suggests that both teniposide and etoposide are non-intercalative in nature based on experiments that show both compounds have a weak affinity for DNA (Ross et al., 1984; Chow et al., 1988). The difference in structure between teniposide and etoposide is the presence of a thiophene ring (**Figure 5.31**), which was also characterised in the Derek Nexus structural alerts. Compounds with sulfhydryl reactive groups such as those present in teniposide have been shown to be potent Topo-II Poisons that appear to act by covalently adducting to the Topo-II enzyme (Wang et al., 2001). The thiophene ring's hydrophobic reaction with the Topo-II enzyme (Chow et al., 1988) would account for the contribution in orders of magnitude greater potency observed for teniposide compared to etoposide.
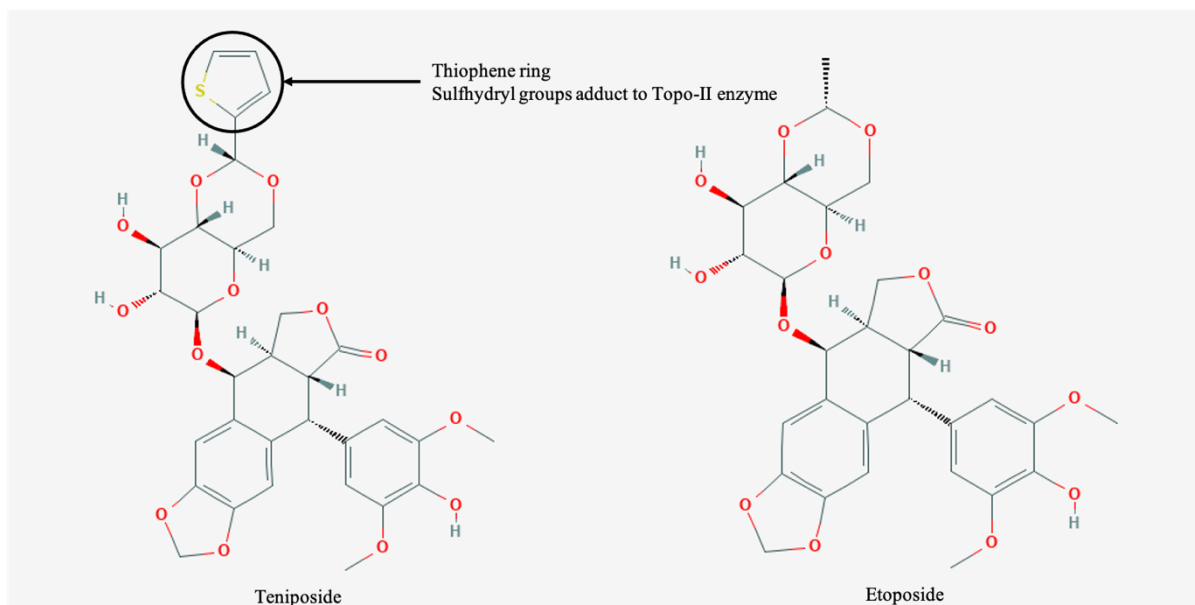
Figure 5.31. Compound structures for the podophyllotoxins teniposide (left) and etoposide (right) that both contain four planar fused rings and differ by a thiophene ring in teniposide which is outlined with a labelled circle in the figure.

With respect to the anthraquinone derivatives mitoxantrone, daunorubicin, doxorubicin and emodin: the presence of hydroxylated planar DNA-intercalative rings (Awasthi et al., 2019) makes the compounds highly DNA reactive (Ribaudo et al., 2019; Cárdenas and Nogueira, 2020). The BMD confidence intervals for daunorubicin and doxorubicin overlapped in the BMD analyses of all three ToxTracker DNA damage biomarker responses, indicating that both compounds are equipotent in these systems. It is suggested that this is because the two molecules are near identical in structure (**Figure 5.32**). Mitoxantrone was the most potent compound in most of the biomarker responses but had overlapping BMD confidence intervals with teniposide in the 24-hour p35 and 4-hour p-H3, Rtkn-GFP and Btg2-GFP biomarker responses, and doxorubicin in the 24-hour p53 biomarker response. This suggests that mitoxantrone is equipotent to teniposide and doxorubicin in the same biomarkers. The repeating ethanolamine moieties in mitoxantrone's structure (**Figure 5.33**) compared to the other anthraquinone derivatives may account for the higher potency observed for mitoxantrone due to possible strengthened reaction with the Topo-II enzyme. This hypothesis can be supported by examining the other anthraquinone emodin, which displayed several orders of magnitude lower potency in the MultiFlow and ToxTracker biomarker responses, yet emodin's structure is closely related to mitoxantrone

(**Figure 5.33**) but lacks the ethanolamine moieties and is less elaborate in structure than the other anthraquinones daunorubicin and doxorubicin.
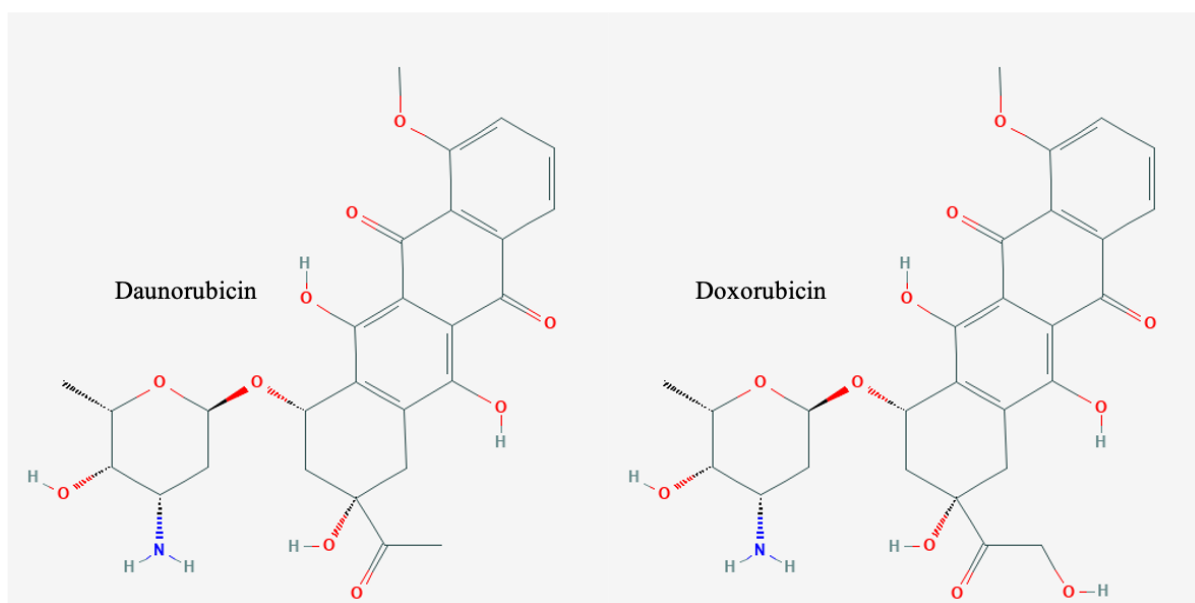
Figure 5.32. Compound structures for daunorubicin and doxorubicin which are nearly identical molecules.
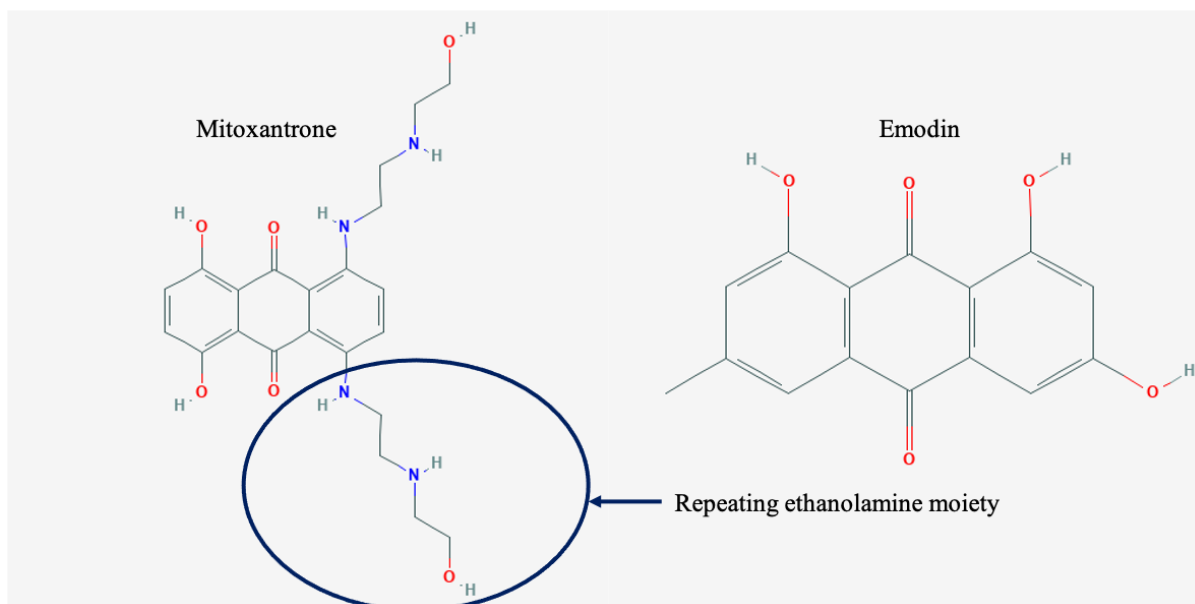
Figure 5.33. Compound structures for mitoxantrone (left) and emodin (right) which are closely related anthraquinone molecules that differ significantly by the addition of repeating ethanolamine moieties in the case of mitoxantrone which is outlined with a labelled circle in the figure.

The isoflavone compound genistein was equipotent to emodin in the tested MultiFlow biomarker responses, with overlapping confidence intervals in most timepoint/biomarker combinations. The two structures are closely related (**Figure 5.34**) which likely accounts for their similar observed genotoxic potencies in the DNA damage response biomarkers.
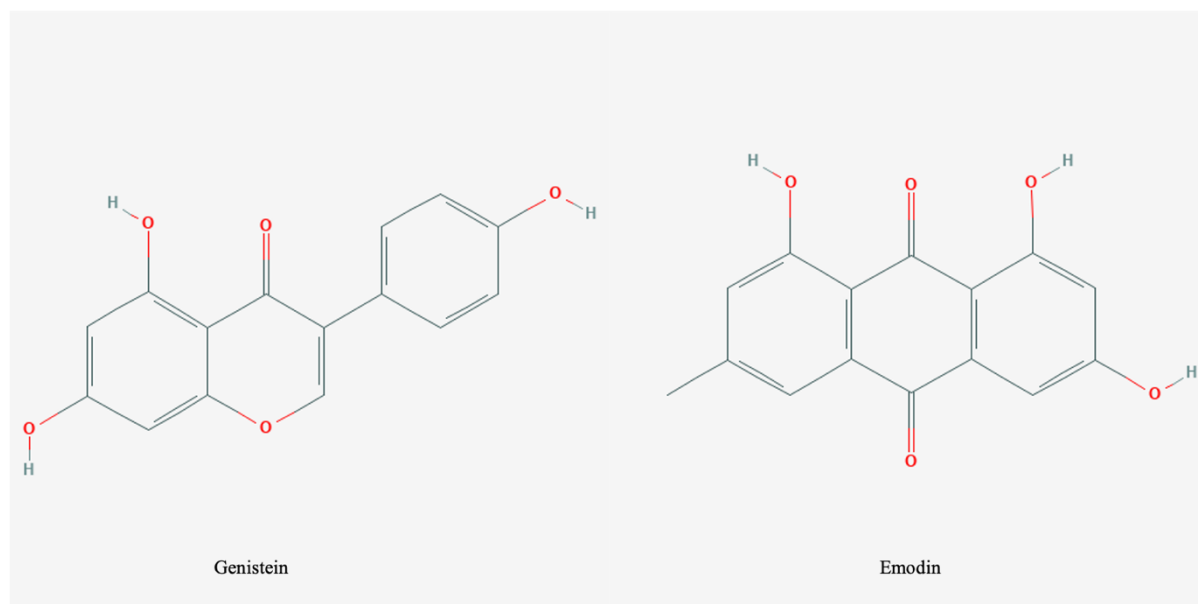
Figure 5.34. Compound structures for genistein (left) and emodin (right) which have closely related structures.

The least potent compounds in all tested biomarkers were the synthetic antimicrobial fluoroquinolones ciprofloxacin and flumequine (**Figure 5.35**). The clinical efficacy of the antimicrobial fluoroquinolones is attributed to targeting of bacterial Topo-II's, Gyrase, and Topoisomerase IV through the formation of a water-metal (Mg^{2+}) ion "bridge" anchor between the fluoroquinolone drug and the gyrase/topoisomerase IV enzyme's serine and acidic residues (Aldred et al., 2013; Aldred et al., 2014a; 2014b; 2016). On the other hand, *in silico* molecular docking studies show that fluoroquinolones form hydrogen bonds and have good binding affinity for the active site of Topo-II (Jadhav and Karuppayil, 2017; Sayin et al., 2018). However, the data reported in this results chapter was derived from *in vitro* cellular systems – so extrapolation is questionable from limited *in silico* studies to the potency in an *in vitro* system. Human Topo-II enzymes are deficient in the serine and acidic residues present in bacterial gyrase/topoisomerase IV enzymes; hence, they are unable to interact with the fluoroquinolones via the critical "bridge" anchor mechanism (Aldred et al.,

2013). The lack of specificity for human Topo II enzymes can explain the difference in potency observed between the fluoroquinolones and the other classes of Topo II poisons.
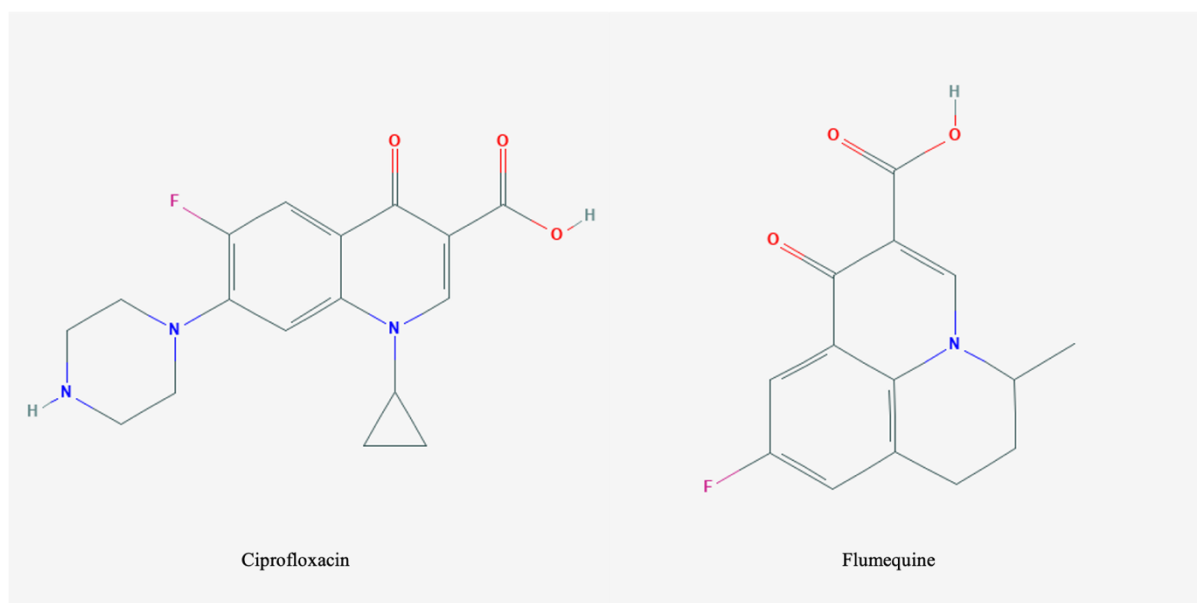
Figure 5.35. Compound structures for the synthetic antimicrobial fluoroquinolones ciprofloxacin (left) and flumequine (right).

5.4.4 Comparing Potencies with the Biochemical Assay

There is overall concordance between the potency rank order obtained from the MultiFlow and ToxTracker biomarkers and the fluorescence anisotropy-based assay, with mitoxantrone being the most potent, and the fluoroquinolones ciprofloxacin and flumequine being the least potent. There was a considerable amount of overlapping error observed in the fluorescence anisotropy-based assay compared to the *in vitro* genotoxicity tests systems evaluated in this results chapter. Interestingly, the observed potencies from the fluorescence anisotropy assay match with the MultiFlow p-H3 endpoint regarding the rank order of doxorubicin in relation to mitoxantrone and teniposide, in comparison to the assays evaluating DNA double-strand break repair foci, γH2AX, and genotoxicity, p53. Genistein's potency in the fluorescence anisotropy assay overlaps that of teniposide. This is highly dissimilar to the potencies of genistein observed in all the tested MultiFlow endpoints where potency was several orders of magnitude different. The observed differences in genistein's potency could be accounted for by the major differences in assay systems, i.e., extracellular vs. cellular. In addition to Topo-II poisoning, genistein concurrently inhibits tyrosine kinase activity, with consequential G2/M arrest (Markovits et al., 1989; Spinozzi et al., 1994; Yan et

al., 2012). The effects of this activity would not be present in an extracellular system. It is proposed that further mechanistic investigation and understanding of the *in vitro* toxicokinetic characteristics of the assays could provide more understanding of the compound's response.

5.5 Concluding Remarks: *In Vitro* Multiplexed Assays Provide Compound Class Specific Mechanistic and Potency Information

The analyses of the MultiFlow and ToxTracker responses to DNA damage presented in this chapter demonstrate the utility of the assay's biomarker dose-responses beyond genotoxic hazard characterisation. This chapter presents novel findings relating to dose-response potency and mechanistic information for the Topo-II Poison class of compounds with a unique mechanism of action. The use of compound structure activity related information provides evidence to substantiate genotoxic potency in a robust and detailed genotoxic risk assessment for the compounds of interest. In fact, coupling the compound structure activity information with mechanistic property information derived from the multiplexed *in vitro* assays adds significant information to support the genotoxic risk from exposure to the compounds under study. Additionally, this could be considered in the framework of read across, where toxicological information from a characterised molecule is applied to a molecule with similar structure with lacking toxicological data (Schultz et al., 2015; Ball et al., 2016). When sections of the data from this chapter was presented in a poster presentation at the Genetic Toxicology Association annual meeting in Delaware (USA) in 2019, scientists from Lhasa Ltd became interested in the application of their Derek Nexus structural alert database to quantitative BMD potency information. This was the first instance where quantitative potency information had been applied to compound genotoxicity information derived from the Derek Nexus database. Derek Nexus compound structural alerts have been used elsewhere in the field of toxicology for *in silico* read across predictions for quantitative measures of skin sensitisation potential (Chilton et al., 2018; Macmillan and Chilton, 2019). The conversations with scientists from Lhasa Ltd at the GTA annual meeting fuelled further hypothesis development of how potency information and structural alerts can be used in the context of genotoxicity risk assessment, and findings are presented in the next chapter of this thesis.

Chapter 6. Using the Topoisomerase-II Adverse Outcome Pathway to Develop a Quantitative Structural Activity Relationship Model for Predicting Genotoxic Potency.

6.1 Introduction

Previous results chapters in this thesis demonstrated the generalisation of comparative BMD potency analysis to the Litron MultiFlow and Toxys ToxTracker multiplexed *in vitro* genotoxicity biomarkers, as well as using the same biomarkers to perform in depth mechanistic and potency analyses for the Topoisomerase II (Topo-II) Poison class of compounds. Compound potency ranks were supported with Topo-II Poison qualitative compound structural information from the Derek Nexus database. Genetic Toxicologists have recently adopted the Adverse Outcome Pathway (AOP) process to describe the biological pathway leading to an adverse genotoxic outcome. It is hypothesised that Derek Nexus database alerts and the *in vitro* MultiFlow and ToxTracker biomarkers correspond to Key Events (KEs) associated with the Topo-II Poison AOP. The aim of this results chapter is to test this hypothesis by mapping the Topo-II Poison AOP KEs to relevant Derek Nexus alerts and *in vitro* biomarker responses that can substantiate KEs. Further hypotheses were developed whereby the following questions were asked and tested in this results chapter[10]:

a) Do compound structural properties correlate with potency relative to activating different Topo-II Poison AOP KEs?

b) Can the potency of a nearest neighbour be predicted based on structural activity relationships (SAR)?

If the hypotheses hold true in this proof-of-concept study, it will be possible to develop Quantitative (Q)SAR models to predict compound potency for Topo-II KEs within the same AOP.

---

[10] The data presented in this results chapter was published in "Benchmark Dose Analysis of DNA Damage Biomarker Responses Provides Compound Potency and Adverse Outcome Pathway Information for the Topoisomerase II Inhibitor Class of Compounds" (Wheeldon et al., 2020) and some passages in this chapter have been quoted verbatim from the same article. A copy of the research article is available in Appendix 3. The data presentation was also shortlisted for the prestigious Lhasa Ltd Richard Williams Memorial Award. A copy of the abstract is available in Appendix 8.

6.1.1 Adverse Outcome Pathways in Genetic Toxicology

In May 2017 at the Health and Environmental Sciences Institute (HESI) Genetic Toxicology Technical Committee (GTTC) workshop in Delaware (USA), the current understanding of various chromosome damaging modes of action (MOAs) were evaluated. The workshop participants discussed the application of Adverse Outcome Pathways (AOPs) to genetic toxicology MOAs which could establish standardised documented protocols that describe the established genotoxic pathways relative to chromosome damage (Sasaki et al., 2020). The outcome of the workshop was that AOPs can identify testing and data gaps that would be integral to a compound risk assessment.

Sasaki et al. (2020) describe that the concepts of MOA and AOPs are similar; however, MOAs are specific to the description of adverse outcomes (AOs) for individual chemicals or chemical groups whereas AOPs can be applied to a variety of agents. The definition of an AOP from OECD (2018) is "a sequence of events commencing with initial interaction(s) of a stressor with a biomolecule within an organism that causes a perturbation in its biology (i.e., molecular initiating event, MIE), which can progress through a dependent series of intermediate Key Events (KEs) and culminate in an Adverse Outcome (AO) considered relevant to risk assessment or regulatory decision making". The AOP framework is used in other disciplines of toxicology and has been successfully used to advance the development of novel assays in the Integrated Assessment and Testing Approach (IATA) (Tollefsen et al., 2014) such as pivotal *in vitro* skin sensitisation assay example (Patlewicz et al., 2014). As such, AOPs can significantly contribute to the advancement of modern testing approaches for regulatory submission purposes (Tollefsen et al., 2014). The Organisation for Economic Cooperation and Development (OECD) developed the AOP Knowledge Base (KB) web platform (https://aopwiki.org/) which stores, and shares peer-reviewed user developed AOPs. Some fundamental concepts described by OECD (2018) that are applicable to AOPs include:

- An AOP is specific to the MIE rather than being chemical-specific.
- KEs relative to the AOP should be directly or indirectly measurable and essential to the progression of the toxicological response.
- Connections between the KEs are termed Key Event Relationships (KERs), which can be described quantitatively (e.g., dose-response relationships), semi-quantitatively, or qualitatively depending on the depth of available knowledge.

- One key objective of the AOP process is to support predictive toxicology based on inferences from the KEs.

A generalised AOP network was shown in the introduction chapter (**Figure 1.15**). The key concepts are defined as the following consistent with OECD (2018) and applied to genotoxic pathways as outlined by Sasaki et al. (2020). **Figure 6.1** illustrates an example AOP structure.

- MIE: Molecular Initiating Event. The initial point of chemical interaction on the molecular level which is a specialised Key Event that starts the AOP.
- KE: Key Event. A measurable and essential change in biological state that leads to a specific toxic outcome.
- AO: Adverse Outcome. In the context of genotoxicity, this is the end of the pathway with an adverse genotoxic event (cellular) compared to an organism level response. The AO is a specialised KE of regulatory importance.
- KER: Key Event Relationship. A relationship that connects the KEs identifying a direct link between an upstream and downstream KE.
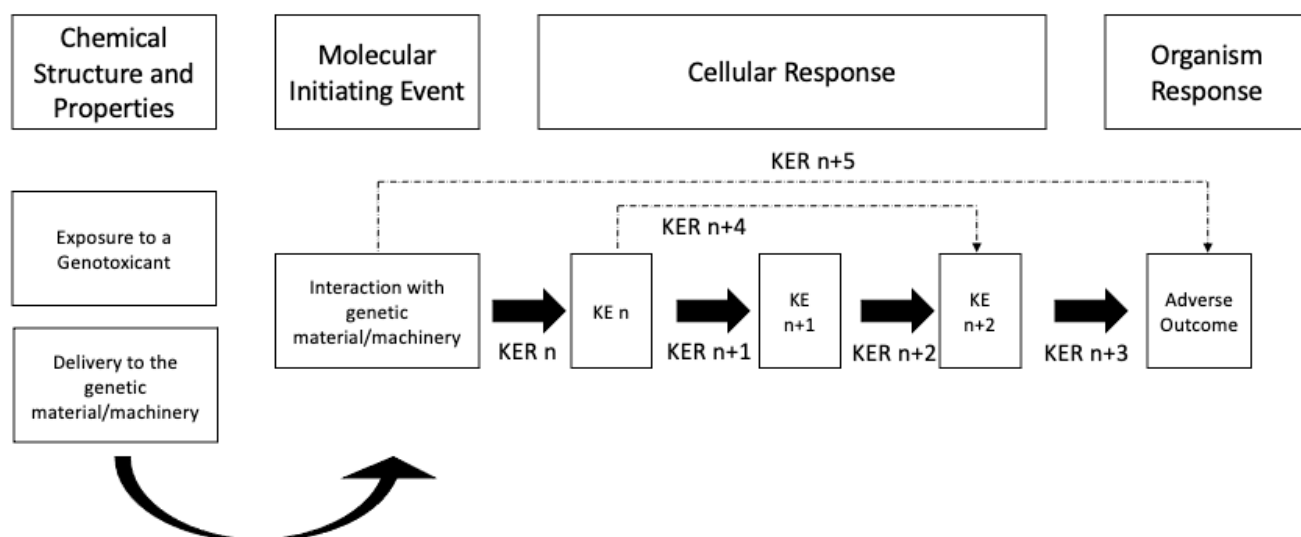
Figure 6.1. Schematic of an AOP network that begins with a stressor (in this case a genotoxicant) being exposed to an organism where it is delivered to and interacts with the genetic material or machinery, known as the molecular initiating event. A series of cellular responses follow that are known as Key Events (KEs) which are essential to the progression of the AOP. KEs are connected by Key Event Relationships (KERs) that describe the relationship between the KEs and can be sequential (KER_{n+1} linking KE_n to KE_{n+1}) or indirect

(KER_{n+4} linking KE_n to KE_{n+2}). The KEs on the cellular level promote the eventual adverse outcome on the organism level, which could be aneuploidy, or mutagenicity and chromosome breaks depending on the mode of action of the exposed stressor.

There are limited genotoxicity AOPs published in the scientific literature. One of the earliest published cases was "alkylation of DNA in male premeiotic germ cells leading to heritable mutations" (Yauk et al., 2015) which describes the biological plausibility and empirical evidence supporting germ cell mutations and subsequent mutations in the off spring of exposed males to compounds capable of alkylating DNA. The AOP consists of the MIE in which male premeiotic germ cell's DNA is alkylated; KE1 in which there is insufficient of incorrect DNA repair in male premeiotic germ cells; KE2 where there is an increase in mutations in male premeiotic germ cells; and the AO where the offspring inherits an increase in mutations. The published AOP includes a detailed account of the biological plausibility of each of the MIE, KEs and AO, detail on how each is measured or detected, and a weight of evidence of the relationships between the KEs (KERs). One example is the relationship between the MIE and KE1, where alkylation of DNA in male premeiotic germ cells leads directly to insufficient or incorrect DNA repair. This KER is based on overwhelming of the DNA repair capacity, specifically the DNA repair mechanism involving O^6-methylguanine-DNA methyltransferase (MGMT) primarily responsible for removing alkyl groups from alkylated DNA (Grombacher et al., 1998; Kaina and Christmann, 2002). DNA repair is dogma (Lindahl and Wood, 1999; Wood et al., 2001), and a detailed weight of evidence account is provided by Yauk et al. (2015) which highlights the empirical evidence supporting the KER. The AOP by Yauk et al. (2015) includes an account of how each KE can be measured or detected, including highlighting any OECD accepted assays such as the transgenic rodent mutation reporter mouse models (OECD TG 488) which can include an evaluation of the exposure effects to the sperm in the vas deferens or caudal epididymis. The AOP also highlights alternative assays than those accepted by OECD, including applying single-molecule PCR to identify tandem repeat mutations in sperm. Overall, this AOP is a very detailed account of the pathway leading to heritable mutations in offspring from males exposed to high doses of alkylating agents. The AOP was published according to the OECD guidelines including an assessment of the overall weight of evidence supporting the AOP's features and highlighting any knowledge or data gaps associated with measuring or identifying the KEs.

The HESI GTTC workshop manuscript by Sasaki et al. (2020) includes less developed AOPs for genotoxic outcomes compared to the "alkylation of DNA in male premeiotic germ cells leading to heritable mutations" AOP published by Yauk et al. (2015). The intent of the Sasaki et al. (2020) manuscript was to summarise the AOP frameworks for a select set of genotoxic outcomes including two proposed AOPs associated with aneugenicity, and three proposed AOPs associated with clastogenicity or mutagenicity *via* non-DNA reactive and indirect mechanisms.

With respect to aneugenicity, the proposed AOPs include "binding to the catalytic domain of AURKs leading to aneuploidy induction" and "binding to tubulin leading to aneuploidy and polyploidy induction". The former AOP background involves the control of chromosome segregation relying on cyclin-dependent kinases (CDKs), polo-like kinases (PLKs), aurora kinases (AURKs), and NimA-related NEK kinases, which are essential to the mitotic process by regulation of the centrosome cycle, formation of the mitotic spindle, and sister chromatid attachment (Bavetsias and Linardopoulos, 2015). Central to the AOP is selective inhibition of AURKs causing monopolar spindles and cell cycle arrest (Girdler et al., 2006; Carpinelli and Moll, 2008). The proposed AOP includes the stressor chemical binding to the catalytic domain of aurora kinases (the MIE), followed by catalytic inhibition of aurora kinases (KE1), cytokinesis inhibition and increase in binucleated cells (KE2), increase of multipolar spindles in 2nd mitotic division (KE3), and finally aneuploidy (the AO). The latter AOP related to aneuploidy includes stressors binding and perturbation of tubulin, which includes a pathway leading to aneuploidy which is widely accepted in the genetic toxicology community. The AOP network (**Figure 6.2**) proficiently outlines the steps from a chemical binding to tubulin (the MIE), to abnormal depolymerisation/stabilisation of microtubules (KE1), mitotic abnormalities (KE2), activation of the spindle assembly checkpoint (KE3), leading to either mitotic slippage (KE4a) or incorrect chromosome alignment and segregation (KE4b). The adverse outcome associated with KE4a is polyploidy, whereas KE4b leads to aneuploidy. The AOP network makes it easier to visualise the different adverse outcome mechanisms (**Figure 6.2**).
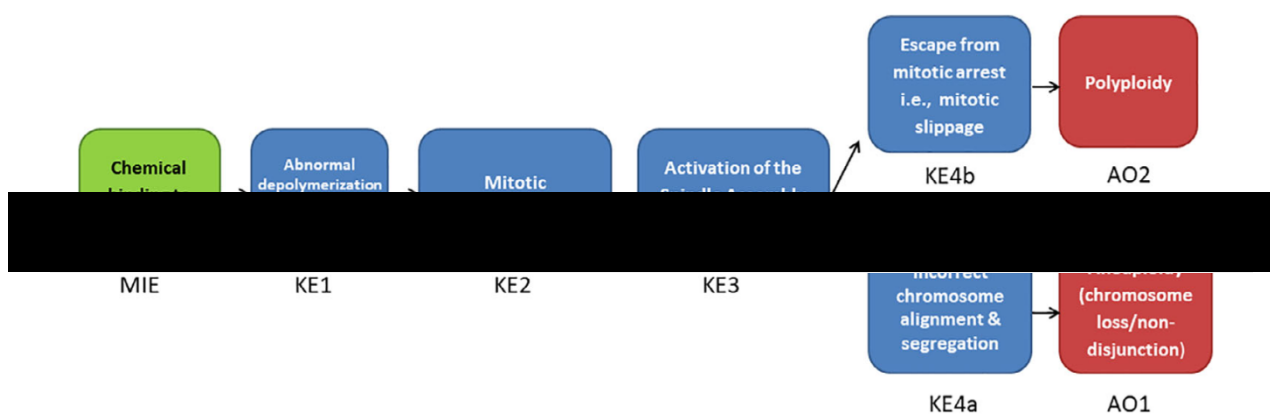
Figure 6.2. Proposed AOP for "binding to tubulin leading to aneuploidy and polyploidy induction". The AOP network proficiently shows the two distinctly different KE4a and b that can lead to different adverse outcomes. Figure taken from Sasaki et al. (2020).

The aneugenicity AOPs are relatively well developed in the publication by Sasaki et al. (2020) due to the interest in the targets for the treatment of tumours. The proposed AOPs related to clastogenicity, and mutagenicity include the non-DNA reactive mechanisms including inhibition of Topo-II or of DNA synthesis, and indirect DNA damage associated with oxidative stress mechanisms. These three AOPs are less well developed compared to the aneugenicity AOPs, however the draft schemes are outlined in the same publication of which it is intended that the AOPs will be further developed by HESI GTTC workgroups. It is also acknowledged that the AOPs share common KEs and AOs which should expedite the development of the AOPs where information can be shared across the common KEs. Once the AOPs are developed, they are intended to be used to understand the MIE and develop testing strategies that can realise the KEs that lead to clastogenic or mutagenic outcomes.

6.1.2 The Topoisomerase II Poison AOP

A particular AOP of interest in scope of this thesis is the Topo-II Poison AOP since the same class of compounds were studied for BMD derived potency measurements in the previous results chapter of this thesis. The proposed HESI GTTC AOP for clastogenicity and mutagenicity associated with Topo-II poisoning is illustrated in **Figure 6.3**:
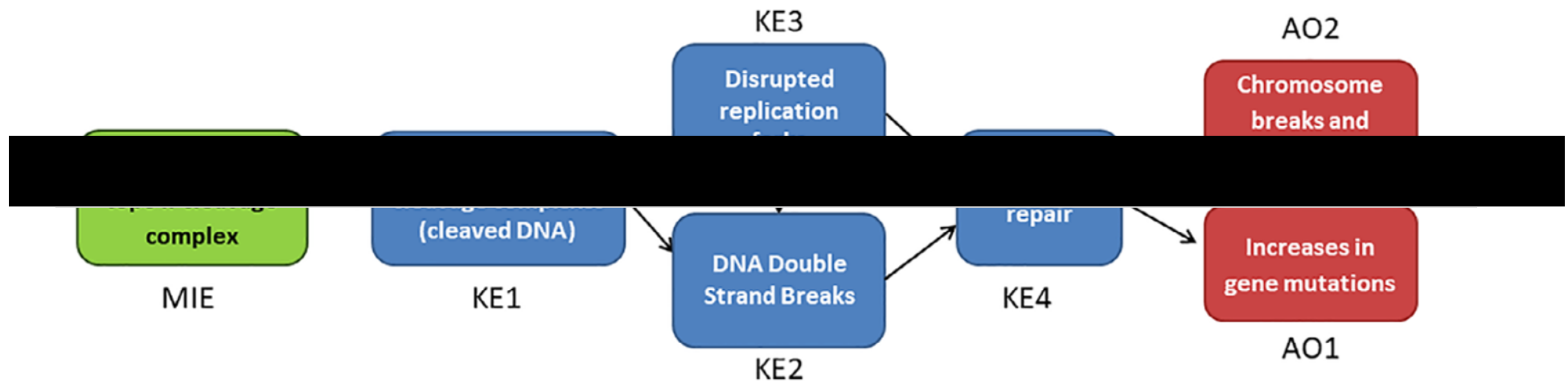
Figure 6.3. Proposed AOP for stressors binding to the DNA-Topo-II cleavage complex leading to stabilisation of the cleavage complex, disrupted replication forks and DNA double strand breaks, leading to inadequate repair and chromosome breaks and rearrangements or an increase in gene mutations. Figure taken from Sasaki et al. (2020).

It is not the purpose of this results chapter to provide the detailed Topo-II Poison AOP (which is still under preparation by the HESI GTTC workgroup), however some of the main features of the AOP are warranted for further discussion. As explained in detail in the previous results chapter, the Topo-II enzyme is involved in the essential molecular process of resolving topological constraints in the genetic material (McClendon and Osheroff, 2007; Berger, 1998; Deweese and Osheroff, 2008). A Topo-II Poison compound irreversibly binds to the DNA-Topo-II cleavage complex and therefore increases in cleavage complexes ensue and inhibit the ability of the enzyme to relegate cleaved DNA molecules (Lynch et al., 2003; McClendon and Osheroff, 2007). The MIE is specific to the Topo-II Poison stressors that bind to the DNA-Topo-II cleavage complex (Wilstermann and Osheroff, 2003); however, Sasaki et al. (2020) states that an assay to identify and quantify a DNA-protein covalently bound complex such as the *in vivo* complex of enzymes assay does not directly measure binding of the agent to DNA-Topo-II and more work would be required to strengthen the biological plausibility of the MIE. The subsequent key events are outlined as: KE1 stabilisation of the cleavage complex; KE2 DNA double strand breaks; KE3 disrupted replication forks; KE4 inadequate DNA repair; AO1 increases in gene mutations; and AO2 chromosome breaks and rearrangements. It is acknowledged that stabilisation of the cleavage complex can lead directly to DNA double strand breaks (KE1 to KE2), or DNA double strand breaks can also be caused by disrupted replication forks (KE3 to KE2).

Aside from the HESI GTTC AOP workgroup, an AOP for exposure to Topo-II Poisons has been published specific to describing the link between Topo-II inhibition and the development of infant leukaemia in which rearrangements of the mixed-lineage leukaemia (MLL) gene produce abnormal fusion proteins in infant B-cell acute lymphoblastic leukaemia (Pendleton et al., 2014; Pelkonen et al., 2017). The AOP is published in the online AOP Wiki (https://aopwiki.org/aops/202). The researchers proposed an AOP which starts with the formation of MLL gene rearrangements by Topo-II Poisoning, leading to fusion genes and ultimately acute leukaemia by global genetic dysregulation (Pelkonen et al., 2017). The AOP begins with the same MIE and KE1 as the HESI GTTC Topo-II Poison AOP in which the stressor binds and stabilises the DNA-Topo-II cleavage complex. However, the Pelkonen et al. (2017) AOP focuses on the MIE occurring *in utero* and progresses with the hallmark KE which includes Topo-II Poisons targeting the MLL breakpoint cluster region (Aplan et al., 1996) yielding DNA double strand breaks that can cleave the MLL gene (Stanulla et al., 1997). The MLL gene can be rearranged with other genes if DNA double stand breaks are not

repaired by DNA repair mechanisms, resulting in the formation of fusion genes (Meyer et al., 2009). The AOP authors (Pelkonen et al., 2017) support the plausibility of the AOP KEs by describing various *in vitro and in vivo* systems that can assess fusion gene products through immunofluorescence techniques. The authors conclude by outlining some of the knowledge gaps in the AOP which includes lack of understanding on the specific embryonic target cell that is susceptible to Topo-II Poison insult leading to infant leukaemia, and the role of epigenetic features involved in influencing the initiation and progression of the disease. The main difference between the HESI GTTC Topo-II Poison AOP and the "inhibitor binding to topoisomerase II leading to infant leukaemia" AOP is the latter's culmination with adverse event visible on the organism level, whereas the genotoxic AOP proposed by HESI GTTC stops at the level where the genotoxic mode of action (aneugenicity, chromosome damage, mutagenicity) is evident.

6.1.3 *In Vitro* Genotoxicity Biomarkers and the Topoisomerase II Poison AOP

Most of the assays summarised by Sasaki et al. (2020) are well characterised assays which can be used to identify and measure the AOPs KEs. For example, stabilisation of the cleavage complex can be measured by identifying trapped complexes using gel electrophoresis techniques such as TARDIS (trapped in agarose DNA immunostaining) (Lynch et al., 2003). KE2 associated DNA double strand breaks can be measured by assessing γH2AX foci. Disrupted replication forks associated with KE3 induce the replication stress response ataxia telangiectasia mutated protein (ATM) and Rad3-relayed protein (ATR). Biochemical phosphorylation assays can be used to measure the ATR dependent phosphorylation of checkpoint kinase 1 (Chk1) which is a specific marker of replication stress. Sasaki et al. (2020) highlights that an indirect measure of inhibition of replication can be obtained from the ToxTracker Bscl2-GFP marker which relies on the activation of ATR and Chk1. Identification and measurement of chromosome breaks/rearrangement and increased gene mutations adverse outcomes include well characterised OECD accepted assays such as chromosome aberration, micronucleus, and Ames bacterial reverse mutation assay.

In the previous results chapter, it was demonstrated that the *in vitro* MultiFlow biomarkers (p53, γH2AX and p-H3) and ToxTracker (Rtkn-GFP, Bscl2-GFP, and Btg2-GFP) biomarkers respond to the Topo-II Poison compounds mechanisms of action. Whilst Sasaki et al. (2020) stated that the ToxTracker Bscl2-GFP offers an indirect measurement of inhibition

of replication, it is hypothesised that the other ToxTracker biomarkers and the MultiFlow biomarkers could offer information relative to supporting the KEs in the Topo-II Poison AOP.

6.1.4 Quantitative Structural Activity Relationships in Genetic Toxicology

Characterising a compounds structure and structural properties is central to compound development. A genetic toxicologist frequently uses *in silico* tools to augment their expert judgement and has been proposed by Dearfield et al. (2017) as crucial to the genetic toxicology risk-assessment. In the previous results chapter, a qualitative dataset consisting of compound structural features was used to scrutinise the potency patterns observed from the Topo-II Poisons studied in the multiplexed *in vitro* genotoxicity biomarker responses. An expert rule-based system, Lhasa's Derek Nexus, was used to derive the compound structural information. Derek Nexus is an expert knowledge system that includes a series of structural alerts arranged by endpoints that builds predictive models for toxicity risk (Marchant et al., 2008). Derek Nexus utilises a strong knowledge base of primarily Ames bacterial mutagenicity data to augment the user's expert judgment (Barber et al., 2017); however, also includes expert-rule based information on DNA damage like chromosome aberrations and micronuclei. In Derek Nexus, a prediction is derived by comparing structural alerts of target compounds (known as toxicophores) with compounds that are coded as structural patterns in its knowledge base (Honma, 2020). The transparency of the Derek Nexus prediction is attributed to the referenced related literature which the genetic toxicologist can use to add to the risk assessment weight of evidence. Derek Nexus has regulatory toxicology applications and is widely used by industry scientists to predict a compound's DNA reactivity per ICHM7 (ICHM7R1, 2013). Whilst the ICHM7 approach is commonplace to genetic toxicologists, the Derek Nexus database includes other alerts not restricted to DNA reactivity. Upon presentation of the same Topo-II Poison *in vitro* genotoxic potency dataset and conclusions to the scientists at Lhasa, it was hypothesised whether the alerts within the Derek Nexus database could relate to the Topo-II Poison AOP Key Events and whether the *in vitro* MultiFlow and ToxTracker biomarkers could be assigned (mapped) to the KEs within the Derek Nexus database. This forms the primary hypothesis relative to this results chapter.

The majority of the predictive (Q)SAR models focus on the ability to predict the mutagenicity outcome in the Ames test (Benigni and Bossa, 2018) by using different statistical or expert-rule based evaluation techniques. The two main expert rule-based

systems include Derek (already briefly introduced), and ToxTree. The ToxTree mutagenicity predictions use a decision tree implementing the Benigni/Bossa rules (Benigni, 2008) based on the identification of structural alerts known to exert carcinogenic/mutagenic properties (Contrera, 2013), mainly through direct DNA interaction. An advantage of the ToxTree predictions is that the software is freely available and easy to navigate, and that the decision tree is transparent to aid in interpretation of the mutagenicity prediction result. Like Derek Nexus, ToxTree provides literature references for the mentioned structural alerts indicative of mutagenicity (Honma, 2020). Some disadvantages of ToxTree include its dichotomous result (mutagenic or not) versus Derek Nexus results that are provided as 'inactive, equivocal, plausible' which could be overturned by expert judgment with supporting evidence (Foster et al., 2020).

The other (Q)SAR systems are based on statistical evaluation. Some examples of statistical based (Q)SAR software include Lazar (lazy structure-activity relationships), Sarah Nexus, and Leadscope. Lazar (https://lazar.in-silico.ch/predict) uses fragments of the query compound to compare them to the training set of over 10,000 compounds from 2 different database sources (Kazius/Bursi and the benchmark dataset) (Maunz et al., 2013). A series of statistical algorithms including k-nearest neighbours and multi-linear regression analyses to compare the results of the applicability domain (number of neighbours, similarity of neighbours) and assign a probability of mutagenicity to the query compound. An advantage of the Lazar approach is its ease of use in a freely available web-based application. However, the system lacks transparency and is irreproducible due to its proprietary algorithms and is poorly validated with low consideration of biological mechanisms (Maunz et al., 2013). Sarah Nexus was developed by the same developers as Derek Nexus (Lhasa Ltd.) and uses compound structural fingerprint information together with a self-organising hypothetical network (SOHN) to produce an activity hypothesis for test compounds based on compound substructures that relate mutagenic activity in the training set (Hanser et al., 2014). Predictions are based on similarity scores of test and query compounds, and the result is transparent with percent similarity and probability scores reported in the user interface. The outcome of the analysis is either positive, negative, equivocal, or out of domain (Honma, 2020) for bacterial mutagenicity predictions. The Leadscope statistical (Q)SAR model for mutagenicity prediction is based on models that are constructed from publicly available Ames mutagenicity data (Roberts et al., 2000). The compound structural descriptors include 27,000 structural features related to common structures present is small molecules. Partial

logistic regression models are used to predict Ames mutagenicity outcome based on the selected descriptors (independent variables) and a database of Ames mutagenicity data (responding variables) (Honma, 2020). Similarly, to Sarah Nexus, the Leadscope outcome includes probability scores related to the test query compound and structures in the training set which can augment an experts review of the model's predicted outcome. The drawback to these three statistical models is that the associated training set complex compound structural information is not easy to calculate by the non-chemist or non-computer scientist, which would usually involve using complex modelling systems to evaluate the compound structural features.

The previous results chapter used qualitative information such as the presence of hydroxylated anthraquinone rings, sulfhydryl groups, and planar rings to support the *in vitro* genotoxic potency estimates of the Topo-II Poisons. It is acknowledged that Quantitative Structural Activity Relationship (QSAR) models include the construction of predictive models as a function of more in-depth structural and molecular information that is more comprehensive than qualitative measures of compound structure used to account for potency in the previous results chapter. However, the genotoxicity (Q)SAR models described in the last few paragraphs relate structural information to the outcome in the bacterial mutagenicity assay, or apical endpoints like chromosome damage and micronucleus. Since genotoxicity is broader than just mutagenicity – and as the novel biomarkers replace the chromosomal damage and micronucleus apical endpoints – there is the urgent need for (Q)SAR models to evolve to predict the outcome in the same novel assays. Another critique relates to the dichotomous outcome of the previously mentioned (Q)SAR models that predicts (with a level of uncertainty) whether a compound is genotoxic or not, which is consistent with the hazard characterisation paradigm and ignores the dose-response and potency of the compound. Additionally, (Q)SAR models that predict assay outcomes relative to an AOP KE would be ideal. With respect to the Topo-II Poisons focussed on in the previous and this results chapter, it would be beneficial if the QSAR model could predict potency of a query compound for the Topo-II KE assays. Thus, after testing the primary hypothesis of whether the Derek Nexus alerts map to the Topo-II Poison AOP KEs, a secondary hypothesis of whether *in vitro* genotoxic potency for the same assays mapped to the KEs can be predicted using compound structural information. The following two specific questions evolved and are aimed to be answered in the next sections of this results chapter:

a) Do compound structural properties correlate with potency, relative to activating different Topo-II Poison AOP KEs?

b) Can the potency of a nearest neighbour be predicted based on structural activity relationships (SAR)?

6.2 Materials and Methods

6.2.1 Datasets

The following datasets constitute the raw data obtained to test the hypotheses in this results chapter.

6.2.1.1 Derek Nexus Datasets

The scientists at Lhasa Ltd performed an initial preliminary mapping of their Derek Nexus database into an attempted AOP framework. This was achieved by exporting alert key words from the database relating to a query compound and assigning an appropriate Key Event name. The query compounds were the same Topo-II Poison compounds studied in the previous results chapter and reiterated in **Table 6.1**. For example, all Topo-II Poison compounds fired the alert for carcinogenicity which was renamed "malignant neoplasm" and denoted a tissue response (tissue). Within each alert the supporting knowledge-based literature key words were used to ascertain whether they correspond to molecular interactions (molecular) or cellular outcomes (cell). Supporting assays were also ascertained from the knowledge-based literature key words. Thus, the alerts and key words from the preliminary mapping were provided as raw data in a network Clustify Summary Statistics File (.cys) where data entries were classed as the following level IDs:

- Compound
- Molecular
- Cell
- Organism
- Tissue
- Assay

Table 6.1. Topo-II Poison Compounds. Abbreviations and Identities.

Compound	Compound Abbreviation	CAS No.
Ciprofloxacin	Cip	85721-33-1
Daunorubicin	Dan	20830-81-3
Doxorubicin	Dox	23241-92-8
Emodin	Emo	518-82-1
Etoposide	Etp	33419-42-0
Flumequine	Flm	42835-25-6
Genistein	Gen	446-72-0
Mitoxantrone	Mit	65271-80-9
Teniposide	Ten	29767-20-2

An exported text file of the Lhasa preliminary level IDs is provided in Appendix 9. Topo-II Poison Derek Nexus Network Mapping by Lhasa.

6.2.1.2 Topo-II Poison Compound Molecular Descriptors

One of the hypotheses being tested is whether compound structural information can be used to build a predictive model for Topo-II compound potency for a KE biomarker. As briefly mentioned in the introduction chapter, most statistical QSAR methodologies use detailed compound structural information to build the QSAR model training sets. Such information is known as compound descriptors which are usually represented as 1-dimensional (1D), 2-dimensional (2D) or 3-dimensional (3D) representations of chemical structures (Cortés-Ciriano et al., 2020). In general, 1D descriptors are the simplest and

represent information calculated from the molecular formular including count and types of atoms and the molecular weight (Chandrasekaran et al., 2018). 2D descriptors are more complex than 1D descriptors and describe molecules based on their topological features (Nettles et al., 2006) including information on the size, shape, and electron distribution of the molecule (Chandrasekaran et al., 2018). 3D descriptors provide a richer profile of a compounds molecular description including information on tautomers, charge distribution, and atomic alignment (Nettles et al., 2006). Examples of 3D properties include information from quantum mechanics calculations, but also include descriptors calculated from the polar and nonpolar surface area of the molecule (Chandrasekaran et al., 2018).

As cheminformatics becomes more advanced, so do the available models to calculate molecular descriptors. However, one does not need to be a computer scientist to derive compound descriptor information. For example, a freely available web-based application 'Mordred' can calculate over 1800 2D and 3D molecular descriptors relative to the query compound (Moriwaki et al., 2018). This provides a challenge, where one is faced with an overwhelming amount of molecular descriptor information. The question becomes which descriptor(s) to select to include in testing the hypothesis presented in this chapter?

A literature search was performed to narrow down descriptors that are commonly used in QSAR models. A common 3D descriptor, the n-octanol-water partition coefficient K_{ow} (or LogP) has been successfully applied in numerous studies to predict the acute toxicity and cytotoxic potency of query compounds (Halle et al., 1991; Saito et al., 1993; Sun et al., 2020; Mombelli and Pandard, 2021). Of note is the inverse relationship between LogP and acute toxicity to aquatic species predicted by the OECD QSAR toolbox which has REACH regulatory acceptance (Mombelli and Pandard, 2021). The octanol water coefficient (LogP) value is the ratio of concentrations of a compound in a mixture of two immiscible solvents at equilibrium and can be experimentally derived or computed (Johanson, 2010). Thus, partition coefficients are a measure of a solute's hydrophobicity and are used in drug design as a proxy for a compound's membrane permeability (Speight, 2018). One study demonstrated that LogP is an important descriptor in the genotoxic potency in E.Coli (Mersch-Sundermann et al., 1994), whilst another demonstrated the relationship between LogP and the genotoxic potential of quinolone antibiotics in the SOS bioassay (Hu et al., 2007). A simple 1D descriptor used widely in QSAR analysis is the compound's molecular weight (MW). A successful predictive model which uses MW is the Derek Nexus quantitative skin

sensitisation (EC3) potency predictor (Macmillan and Chilton, 2019) where the training set was built using Local Lymph Node Assay quantitative data, and compound MW. The skin sensitisation Molecular Initiating Event (MIE) is the interaction of the compound with skin peptides/proteins (Maxwell et al., 2014) and most skin sensitisers are electrophilic (Roberts and Aptula, 2014). This draws parallels with genotoxicity where most DNA reactive compounds are electrophilic, and specifically where the Topo-II Poison AOP MIE also includes interaction with a protein (Topo-II enzyme) to form the DNA-Topo-II enzyme cleavage complex. Another 3D descriptor used in cheminformatics is the calculated polar surface area (PSA), which is the ratio of a compound's surface belonging to polar atoms and is a descriptor that correlates well with passive transport through membranes used in drug development (Ertl et al., 2000). Like the application of LogP, PSA has also been demonstrated to be an important predictor of bacterial mutagenicity (Votano et al., 2004). Therefore, based on some examples where these molecular descriptors have been used to develop QSAR models, it was decided to us the same molecular descriptors for the Topo-II Poison compound analyses presented in this results chapter: molecular weight, LogP, and polar surface area.

The ChEMBL database (https://www.ebi.ac.uk/chembl/) was used to search for compound molecular descriptor data specifically molecular weight, LogP and polar surface area for the 9 Topo-II Poison compounds in **Table 6.1**.

6.2.1.3 Topo-II Poison BMD Potency Metrics

Genotoxic potency has been consistently presented as a range based on the width of the calculated BMD confidence intervals. Therefore 3 potency metrics are provided from the analysis (BMDL, BMD and BMDU). The values of the BMDL, BMD, and BMDU (or CEDL, CED, and CEDU) were obtained from the same analyses presented in previous results chapter for the following MultiFlow and ToxTracker biomarker responses:

- 4- and 24-hour γH2AX
- 4- and 24-hour p53
- 4- and 24-hour p-H3
- Rtkn-GFP
- Bscl2-GFP
- Btg2-GFP

Separate databases for each biomarker were curated in Excel which included the BMD metrics values, molecular weight, LogP and polar surface area and their Log transformed values. All databases were transformed into a .csv text file for further statistical analyses.

6.2.2 AOP Mapping

The Derek Nexus network Clustify Summary Statistics File (.cys) was used to edit the mapping of Derek Nexus alerts according to the proposed HESI GTTC Topo-II Poison AOP presented in the introduction section of this chapter. Editing was performed using the Cytoscape v3.8.0 software programme (https://cytoscape.org/) which is an open-source bioinformatics platform for the creation of complex networks. The Level IDs were renamed to the following to be consistent with the AOP framework terminology:

- Compound → Stressor
- Molecular → MIE
- Cell → KE
- Organism → AO
- Tissue → AO
- Assay → Test

Links that represent the KERs were made between the different KEs by creating a shared node in Cytoscape by formatting the data entries as either 'source' or 'target' nodes according to the alerts in the Derek Nexus database export results. MultiFlow and ToxTracker biomarkers were added intuitively to the respective KE that they align to. The Lhasa proposed Cytoscape neural network was too complex and not optimised for illustration of the Derek Nexus alerts mapped to the Topo-II Poison AOP. Furthermore, their proposed MIEs of 'DNA nucleobase reaction with electrophile' and 'protein reaction with electrophile' were combined with the 'Topoisomerase II poisoning' MIE due to the reasoning that the molecular events relate to binding to the DNA-Topo-II cleavage complex.

6.2.3 QSAR Model Development: K-Nearest Neighbour Analysis

Various statistical modelling techniques have been reported in the literature for building predictive QSAR models to for compound biological action. The techniques include decision trees (Contrera, 2013; Macmillan and Chilton, 2019), logistic regression modelling and

artificial neural networks (Saito et al., 1993) and classification techniques such as nearest neighbour analysis (Mattioni et al., 2003). There is extensive debate as to which model performs the best; however, for this proof-of-concept study in this results chapter it was decided to use the K-Nearest Neighbours algorithm since the aim was to determine if compound structural information can be used to identify compound genotoxic potential relative to its nearest neighbour.

The K-Nearest Neighbours algorithm represents a simple method for performing classification tasks (discreet variables) or regression analysis (continuous variables). (Hackeling, 2017) describes that the basis of the K-Nearest Neighbours method is described in its title, where neighbours are representations of training instances within a metric space. A metric space can be a graph where distances between variables are defined. The distances between neighbours in the metric space can be calculated using a metric function. The hyperparameter 'k' represents how many neighbours can be used in the estimation. Since the variables included in this results chapter are continuous data type, the K-Nearest Neighbours modelling performed will be regression analysis. A classical K-Nearest Neighbours regression example provided by Hackeling (2017) is whether a person's height and sex can be used to predict their weight. Applying K-Nearest Neighbours regression analysis to the hypotheses presented in this results chapter includes using the algorithm to predict if compound molecular descriptors can be used to predict genotoxic potency. To this end, the BMD metrics and molecular descriptors were used as variables for K-Nearest Neighbours regression analysis.

K-Nearest Neighbours analysis was performed using SPSS v26. The biomarker specific database files (6.2.1.3) were imported into SPSS as text files. Nearest neighbour analysis was selected from the 'Analyze' function, and the variables were assigned to the following instances:

- Compound → Target
- BMD metrics (LogBMDL, LogBMD, LogBMDU) → Features
- Log Molecular Weight[11] → Features
- LogP → Features

---

[11] Molecular Weight values were Log transformed

- Log Polar Surface Area[12] → Features

K-Nearest Neighbours analysis was performed iteratively to test the performance of the following combination of molecular descriptors relative to BMD potency: LogP vs. Log Polar Surface Area; Log Polar Surface Area vs. Log Molecular Weight; Log Molecular Weight vs. LogP. K-Nearest Neighbours analysis was also performed using only the molecular descriptors in the same combination without including the BMD potency metrics. K-Nearest Neighbours distance outputs were visually compared. Number of nearest neighbours (value of k) was set as 3 (rounded $\sqrt{n}$ where maximum n was 8). Distance computation was based on the Euclidean distance (more detailed information on Euclidean distance is provided in the general methods chapter 2.5 of this thesis). Dataset partitions included the default setting of 70:20 training to holdout percent ratios. V-fold cross-validation was not performed since the value of n was relatively small. All other settings were held at default. Predictor space graphical outputs were exported for K-Nearest Neighbours distance evaluation.

6.3 Results

6.3.1 Derek Nexus Database Mapping to the Topo-II Poison AOP

Combining the 'DNA nucleobase reaction with electrophile' and 'protein reaction with electrophile' with the 'Topoisomerase II poisoning' MIE significantly simplified the network. Additionally, some alerts are indicative of oxidative stress related DNA damage mechanisms, namely "uncoupling of oxidative phosphorylation", "radical species generation", and "exogenous oxidant increase". These three KEs were combined into a single KE named "ROS generation" and mapped to the "Topoisomerase II Poisoning" and "DNA double-strand break (DSB)" KEs. The complexity of the network with these 3 KEs added can be found in Appendix 10 (Appendix 10. Complex Topo-Poison Derek Nexus Mapped AOP Network). The formatted level IDs aligning the AOP network to the AOP framework are summarised in **Table *6.2***. The final Derek Nexus alert and key word database mapped to the Topo-II Poison AOP network and is shown in **Figure 6.4**.

---

[12] Polar Surface Area values were Log transformed

Table 6.2. Level ID Labels for the Derek Nexus Alerts and Key Words Mapped to the Topo-II Poison AOP

Level ID	Name
Ciprofloxacin	STRESSOR
Daunorubicin	STRESSOR
Doxorubicin	STRESSOR
Emodin	STRESSOR
Etoposide	STRESSOR
Flumequine	STRESSOR
Genistein	STRESSOR
Mitoxantrone	STRESSOR
Teniposide	STRESSOR
Topoisomerase II poisoning	MIE
Non-covalent interaction with DNA	MIE
Estrogen receptor (ER) binding	MIE
Tubulin binding	MIE
Gyrase inhibition	MIE
Mitotic dysregulation	KE

Level ID	Name
Tumour suppressor protein p53 signalling pathway dysregulation	KE
ROS Generation	KE
DNA damage repair	KE
DNA double-strand break (DSB)	KE
Micronucleus formation	KE
MultiFlow 4- and 24-hour p-H3	TEST
MultiFlow 4-hour and 24-hour p53	TEST
MultiFlow 4- and 24-hour gamma-H2AX	TEST
Transgenic Rodent Study	TEST
Rodent Carcinogenicity Study	TEST
Mouse Lymphoma Assay	TEST
In Vivo Uds Assay	TEST
In Vivo Micronucleus Test	TEST
In Vivo Chromosome Aberration Test	TEST
In Vitro Uds Assay	TEST
In Vitro Micronucleus Test	TEST

Level ID	Name
In Vitro Hgprt Assay	TEST
In Vitro Comet Assay	TEST
In Vitro Chromosome Aberration Test	TEST
Ames Test	TEST
ToxTracker Rtkn-GFP	TEST
ToxTracker Btg2-GFP	TEST
Mitochondrial dysfunction	AO
Inherited DNA mutation	AO
Chromosome structural damage	AO
Malignant neoplasm	AO
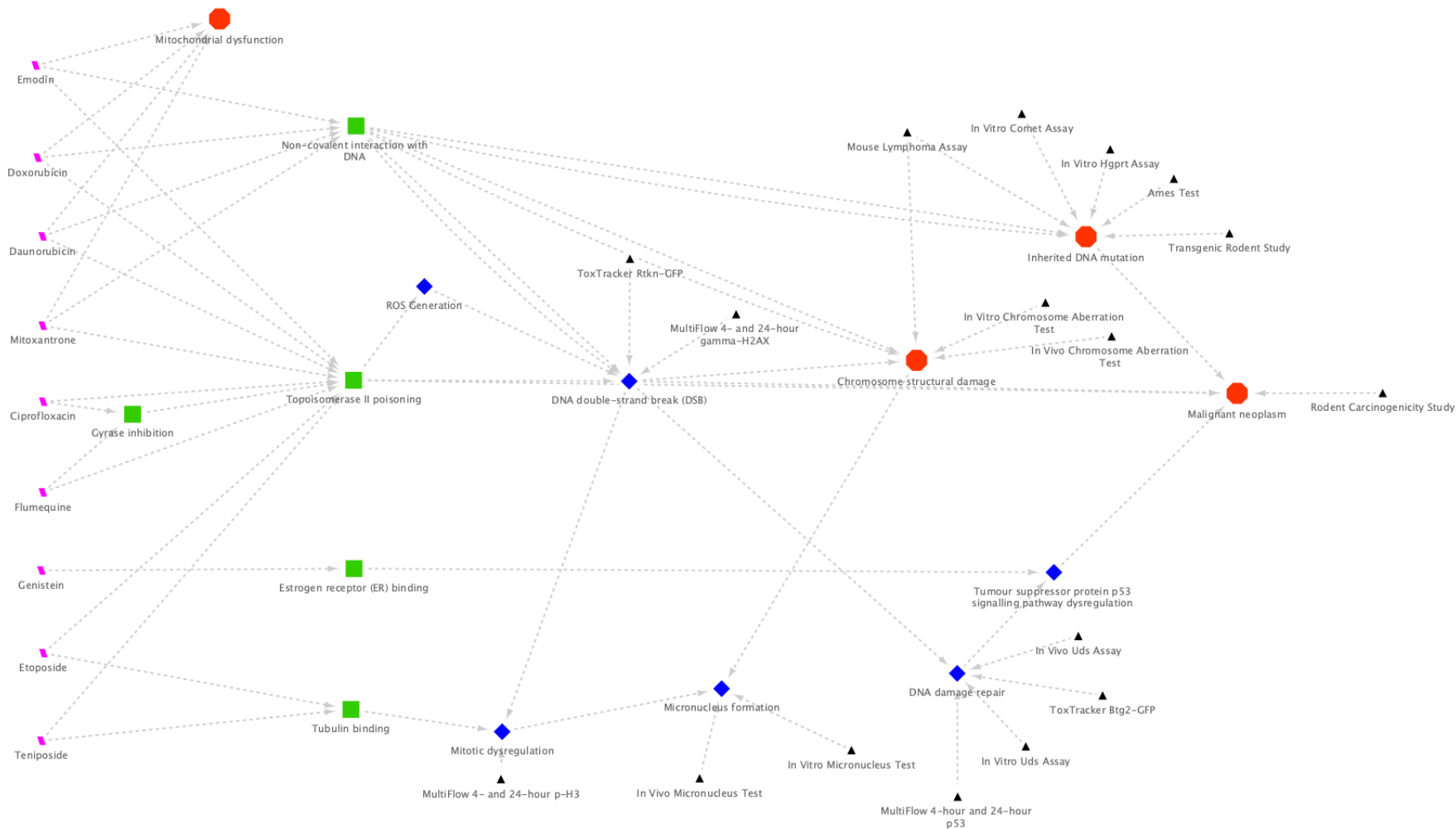
Figure 6.4. Derek Nexus Alerts and Key Words Mapped as a Topo-II Poison AOP Network. Key: Pink parallelogram = STRESSOR; green square = MIE; blue diamond = KE; red octagon = AO; black triangle = TEST.

6.3.2 *In Vitro* MultiFlow and ToxTracker Biomarkers Mapped to the Topo-II Poison AOP

The MultiFlow and ToxTracker biomarkers were intuitively mapped to the KEs corresponding to the types of DNA damage they respond to according to **Table *6.3***. The ToxTracker Bscl2-GFP biomarker indicative of replication fork stress did not correspond to any alerts or key words in the Derek Nexus database.

Table 6.3. MultiFlow and ToxTracker Biomarkers Corresponding to the Derek Nexus Mapped KEs

Derek Nexus AOP Network Node KE	Mapped MultiFlow and ToxTracker Biomarker Responses
DNA double-strand breaks (DSB)	4- and 24-hour γH2AX, Rtkn-GFP
Tumour suppressor protein p53 signalling pathway dysregulation	4- and 24-hour p53, Btg2-GFP
Mitotic dysregulation	4- and 24-hour p-H3

6.3.3 Topo-II Compound Molecular Descriptors

The following table (**Table *6.4***) summarises the molecular descriptor data obtained for each compound from the ChEMBL database. All values were Log transformed for further statistical analyses to approximate to the normal distribution.

Table 6.4. Topo-II Poison Compound Molecular Descriptors

Compound	Molecular Weight g.mol^{-1}	LogP	Polar Surface Area (Å)
Ciprofloxacin	331.35	-0.85	74.57
Daunorubicin	527.53	1.34	185.84

Compound	Molecular Weight g.mol^{-1}	LogP	Polar Surface Area (Å)
Doxorubicin	543.53	0.53	206.07
Emodin	270.24	3.82	94.83
Etoposide	588.56	1.16	160.83
Flumequine	261.25	2.42	59.3
Genistein	270.24	3.08	90.9
Mitoxantrone	444.49	0.75	163.18
Teniposide	656.66	2.78	160.83

6.3.4 K-Nearest Neighbours Analysis

K-Nearest Neighbours analysis followed a tiered approach summarised in the K-Nearest Neighbours analysis flow chart (**Figure 6.5**). The results of the K-Nearest Neighbours analyses are presented in the following subsections.
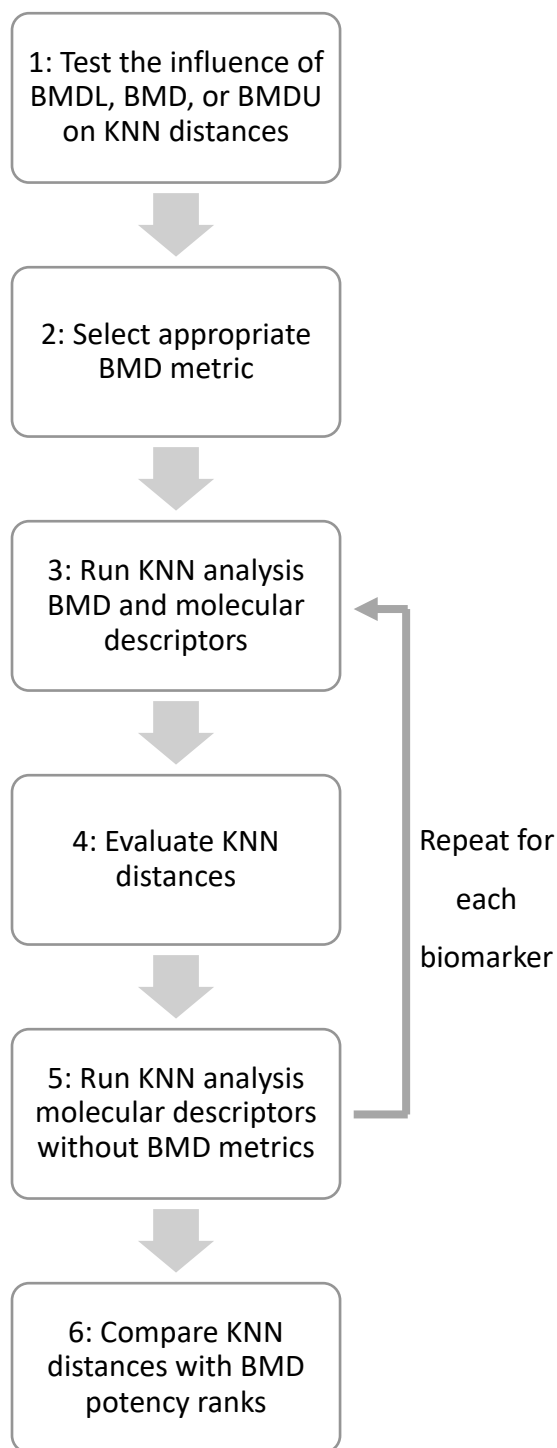
Figure 6.5. K-Nearest Neighbours analysis flow chart

6.3.4.1 Choice of BMD Metric

A fundamental issue needed to be resolved relating to the influence of using either a BMDL, BMD, or BMDU value to build the K-Nearest Neighbours QSAR model. For potency differences to be statistically significantly determined, one should compare the

relative overlap/non-overlap of BMD confidence intervals as advocated in the scientific literature and other chapters of this thesis. Unfortunately, all BMDL, BMD and BMDU values could not be incorporated into the K-Nearest Neighbours analysis matrix space with the molecular descriptors that are fundamental to defining the QSAR nearest-neighbours. A maximum of 3 axes could be included in the SPSS K-Nearest Neighbours analysis matrix space and this would leave no space for plotting molecular descriptors if all 3 BMD metrics were included. To overcome this obstacle, the ToxTracker Rtkn-GFP and Bsc2-GFP biomarker response BMD metrics were selected for iterative K-Nearest Neighbours analysis in the following order to determine whether using BMDL, BMD, or BMDU introduced bias or changed the outcome:

1. BMDL
    a. LogP vs. Log Polar Surface Area
    b. Log Polar Surface Area vs. Log Molecular Weight
    c. Log Molecular Weight vs. LogP
2. BMD
    a. LogP vs. Log Polar Surface Area
    b. Log Polar Surface Area vs. Log Molecular Weight
    c. Log Molecular Weight vs. LogP
3. BMDU
    a. LogP vs. Log Polar Surface Area
    b. Log Polar Surface Area vs. Log Molecular Weight
    c. Log Molecular Weight vs. LogP

The Rtkn-GFP and Bscl2-GFP biomarkers were selected since the resulting Topo-II Poison BMD confidence intervals were the widest or most variable, respectively, compared to all other ToxTracker and MultiFlow BMD confidence intervals for the same compounds. The wider BMD confidence intervals in the Rtkn-GFP biomarker responses corresponds to greater levels of uncertainty in the BMD analyses for this biomarker which was suggested may have an influence on the K-Nearest Neighbours analyses. Conversely, comparably smaller confidence intervals correspond to lower uncertainty, and it was suggested may have less influence on the K-Nearest Neighbours analyses. Variable width confidence intervals within the Bscl2-GFP biomarker BMD analyses correspond to the variation in the precision of the BMD estimates across the compound covariate studies in the same biomarker

response. It was also deemed necessary to evaluate the impact of this variation on the K-Nearest Neighbours analysis. **Figure 6.6**, **Figure 6.7**, and **Figure 6.8** report the K-Nearest Neighbours analysis results for the Rtkn-GFP biomarker responses, and **Figure 6.9**, **Figure 6.10**, and **Figure 6.11** for the Bscl2-GFP biomarker responses.
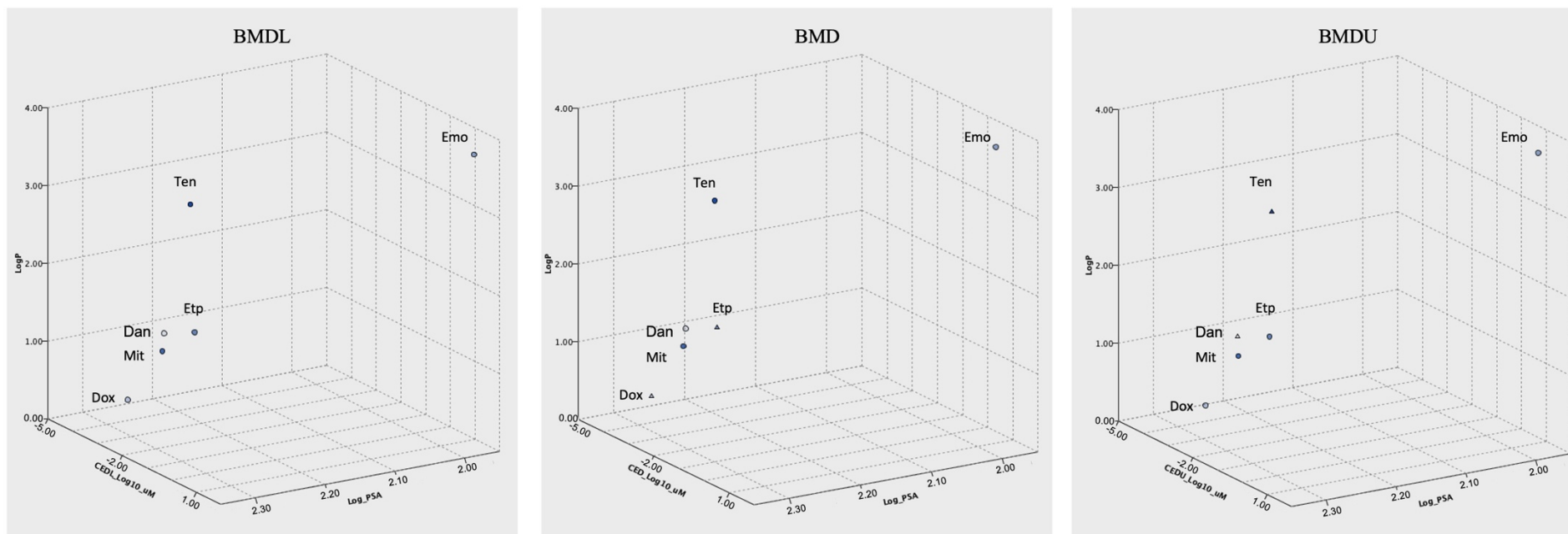
Figure 6.6. ToxTracker Rtkn-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogP and LogPSA Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

Figure 6.7. ToxTracker Rtkn-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogPSA and LogMW Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

Figure 6.8. ToxTracker Rtkn-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogMW and LogP Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

BMDL, BMD, or BMDU vs. LogP and LogPSA Molecular Descriptors KNN Matrix Space Plots: Bscl2-GFP Biomarker Responses
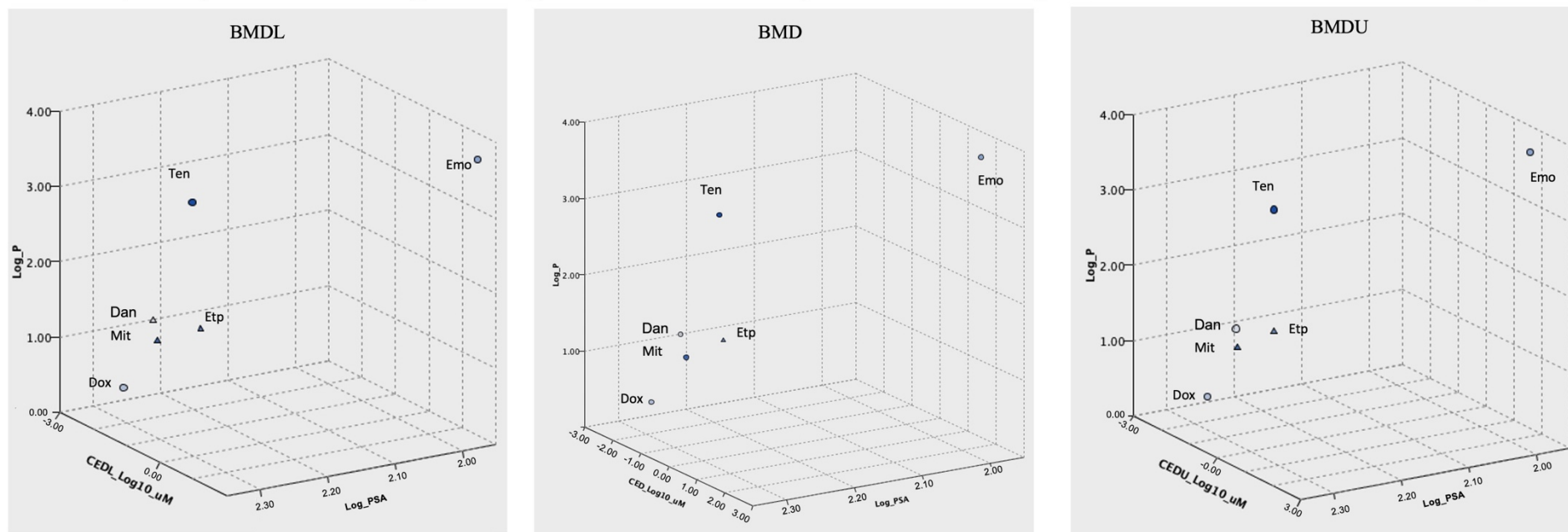
Figure 6.9. ToxTracker Bscl2-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogP and LogPSA Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

Figure 6.10. ToxTracker Bscl2-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogPSA and LogMW Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

Figure 6.11. ToxTracker Bscl2-GFP biomarker response K-Nearest Neighbours Matrix Space Plots evaluating the influence of BML, BMD, or BMDU values on the K-Nearest Neighbours distance metrics for the LogMW and LogP Topo-II Poison compound molecular descriptors. K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CEDL, CED, and CEDU as PROAST notations correspond to BMDL, BMD, and BMDU, respectively. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2 K-Nearest Neighbours Analysis Compound Molecular Descriptor Combinations with BMD Potency

The BMD values were selected to proceed with K-Nearest Neighbours analyses. A detailed discussion of the justification for the choice of using the BMD point estimate metric for K-Nearest Neighbours analysis is included in the discussion section 6.4.3.1. The following subsections present the results of the K-Nearest Neighbours analyses within each MultiFlow and ToxTracker biomarker response. Only the 24-hour responses for the MultiFlow assay are displayed in the following subsections as a proof of concept.

6.3.4.2.1 24-hour γH2AX
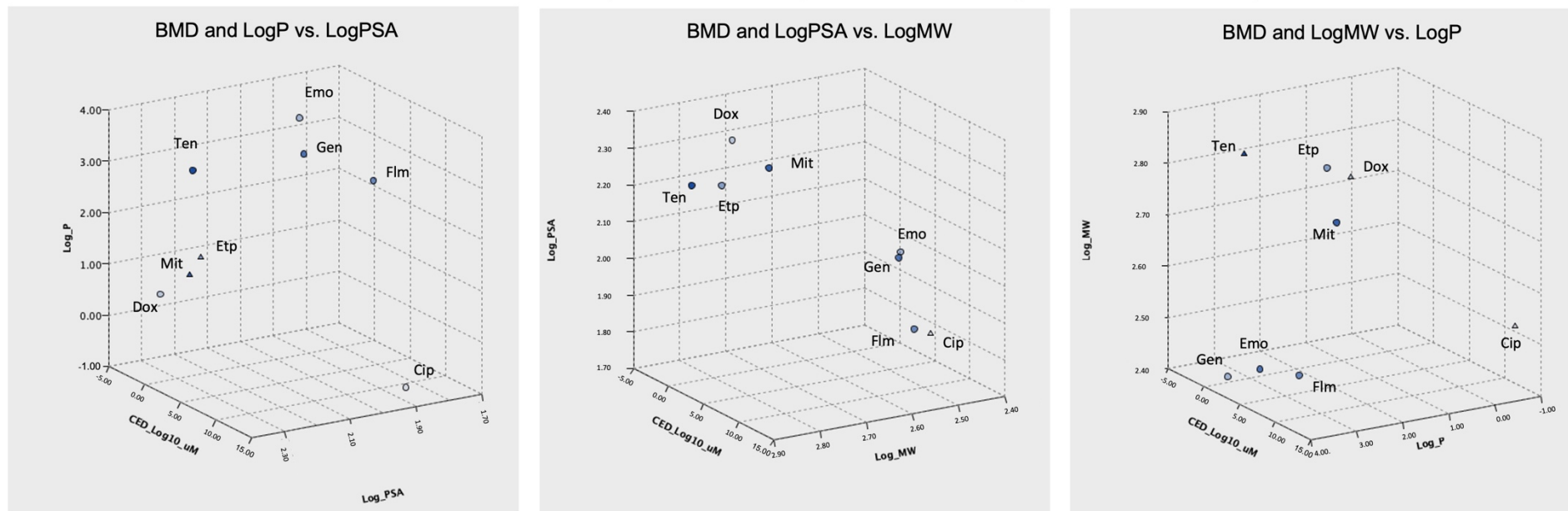
Figure 6.12. MultiFlow 24-hour γH2AX biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2.2 24-hour p53
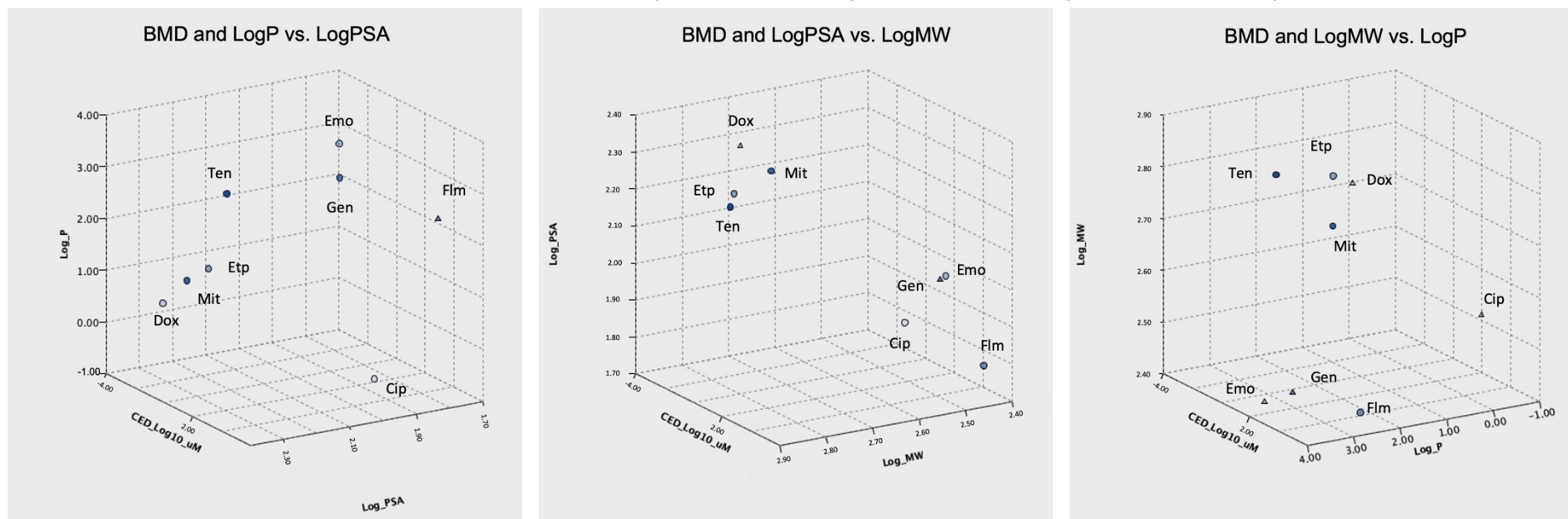
Figure 6.13. MultiFlow 24-hour p53 biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2.3 24-hour pH3
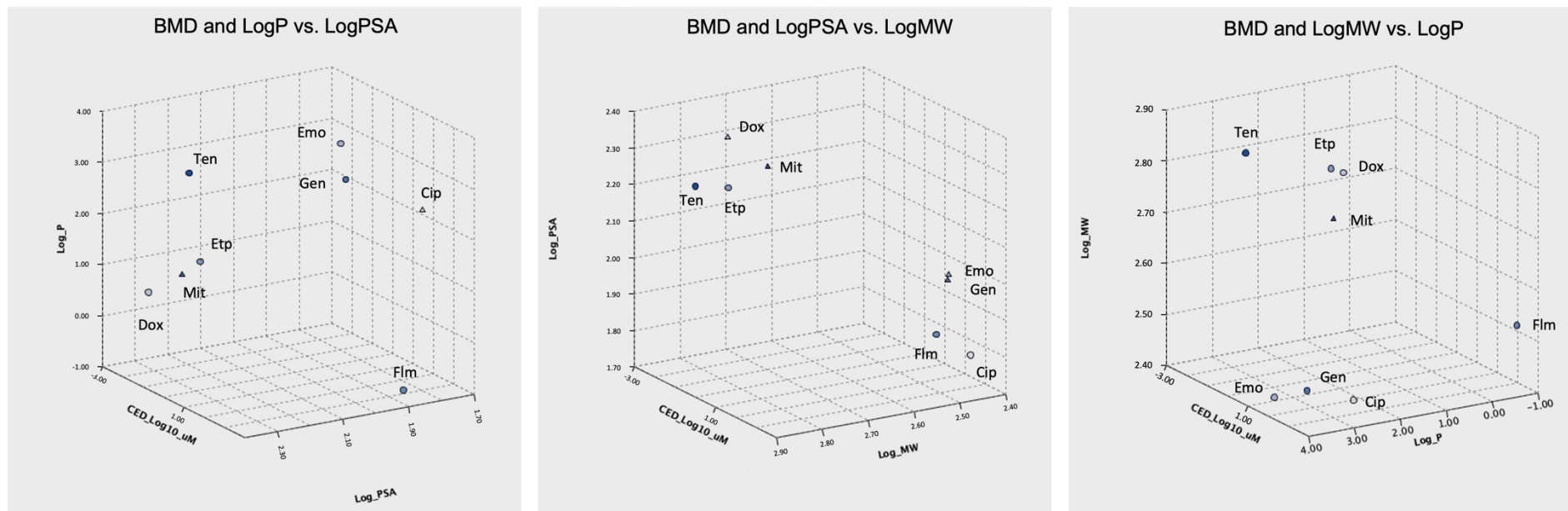
Figure 6.14. MultiFlow 24-hour p-H3 biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2.4 Rtkn-GFP
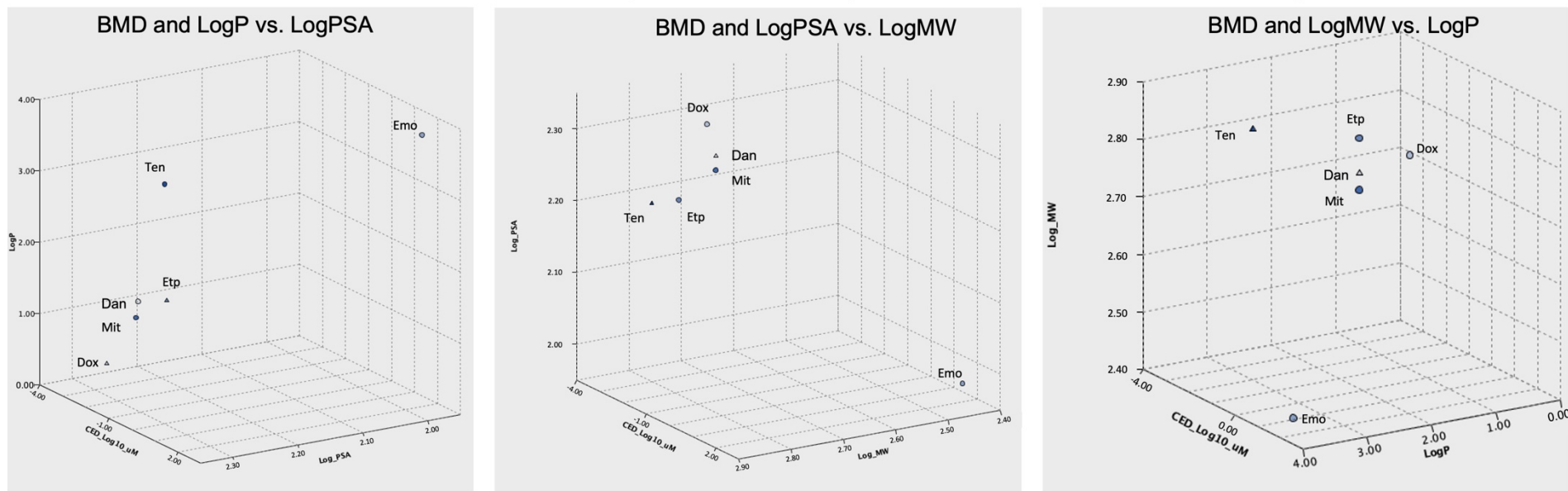
Figure 6.15. ToxTracker Rtkn-GFP biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2.5 Bscl2-GFP

Figure 6.16. ToxTracker Rtkn-GFP biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.2.6 Btg2-GFP
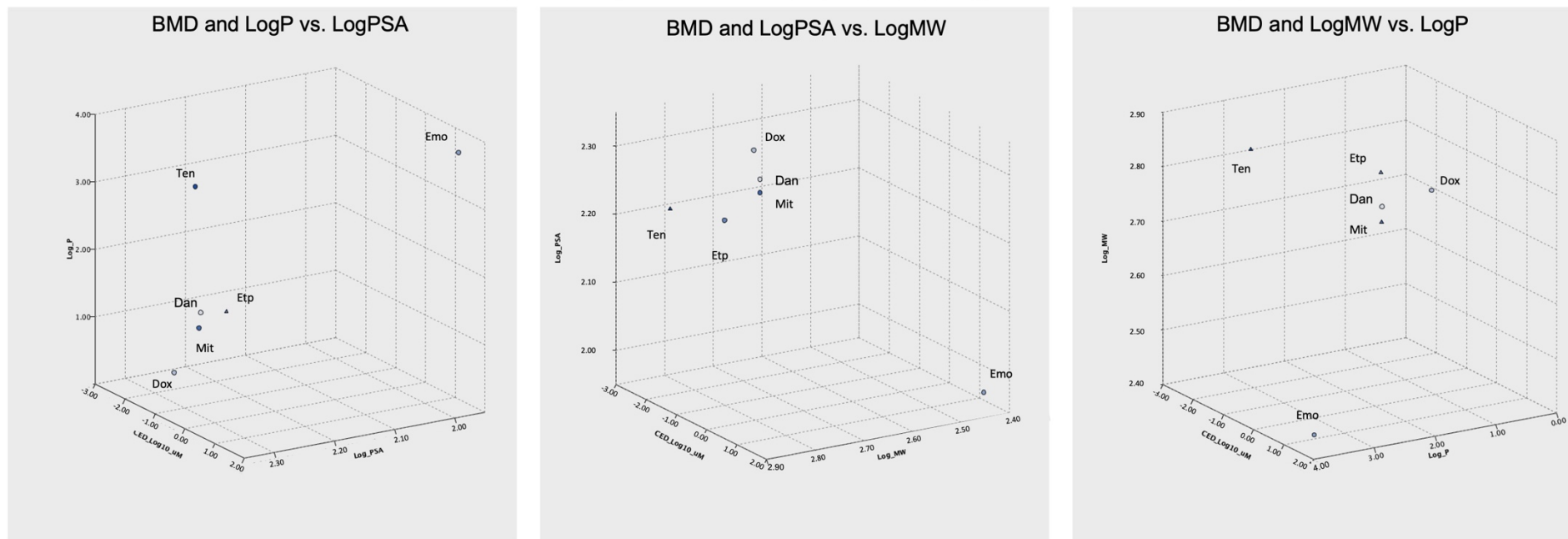
Figure 6.17. ToxTracker Btg2-GFP biomarker response K-Nearest Neighbours 3-Dimensional Matrix Space Plots built with BMD values for the Topo-II Poison compounds and plotted in relative space with LogP and LogPSA values (left panel), LogPSA and LogMW values (centre panel) and LogMW and LogP values (right panel). K-Nearest Neighbours distances are visualised in 3-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. CED PROAST notation correspond to BMD. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.3.4.3 K-Nearest Neighbours Analysis Compound Molecular Descriptors without BMD Potency

K-Nearest Neighbours analysis was repeated without BMD metrics to validate the impact of only molecular descriptor variables on K-Nearest Neighbours distances. This is an important step to assess whether molecular descriptor K-Nearest Neighbours distances are predictive of compound genotoxic potency without the K-Nearest Neighbours model factoring in the BMD metrics in the training sets. Thus, K-Nearest Neighbours analyses were performed using the following combination of compound molecular descriptors: LogP vs. Log Polar Surface Area; Log Polar Surface Area vs. Log Molecular Weight; and Log Molecular Weight vs. LogP. Results are shown in **Figure 6.18**.
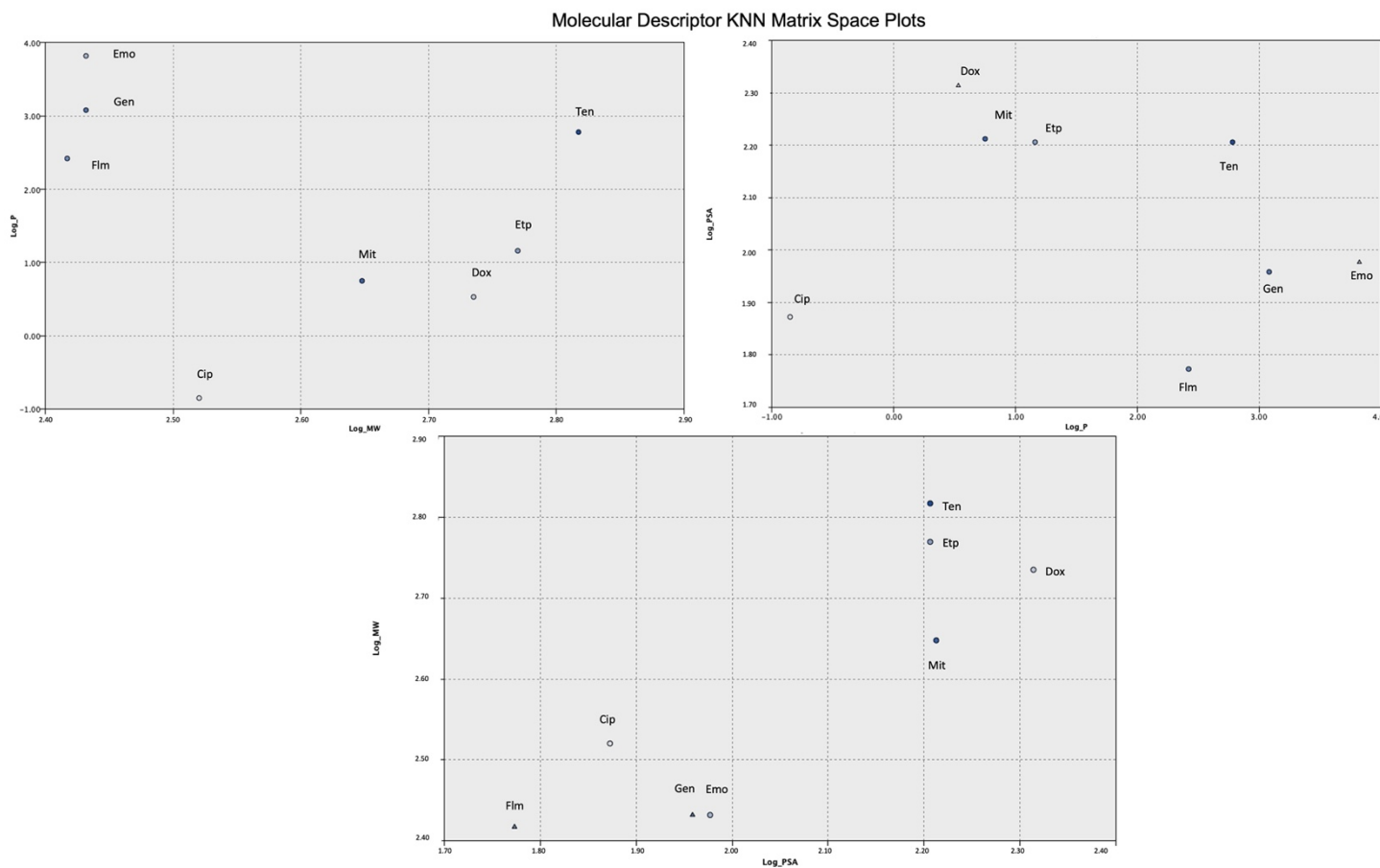
Figure 6.18. K-Nearest Neighbours 2-Dimensional Matrix Space Plots built with molecular descriptors LogP, LogPSA, and LogMW values for the Topo-II Poison compounds. K-Nearest Neighbours distances are visualised in 2-dimensional matrix space plots where the distances between nearest neighbours can be visualised by order of magnitude in the Log scales. Data points are labelled by compound abbreviations per **Table 6.1**. Abbreviations: KNN = K-Nearest Neighbours; PSA = Polar Surface Area; MW = Molecular Weight.

6.4 Discussion

6.4.1 Derek Nexus Database Topo-II Poison AOP Mapping

In the case of the 9 Topo-II Poisons included in this study, the Derek Nexus knowledge-based database contains relevant key words that describe the molecular and cellular perturbations that lead to chromosome damage, inherited mutations, and a potential malignant neoplasm. However, as Derek Nexus is currently utilised, the user must read the report in detail to interpret the linkages between the compounds molecular and cellular interactions to understand the potential adverse outcomes. The Derek Nexus output is also currently only utilised to determine what assays describe the adverse outcome with little focus on the intermittent steps. By performing this Derek Nexus database mapping to the AOP, equal weighting is designated to the description of the cellular and molecular events that lead to the adverse outcomes relative to the outcomes themselves. Additionally, the mapping aligns the molecular and cellular Derek Nexus database key words and alerts to the AOP framework. To this end, the mapping revealed that the Derek Nexus database adequately describes the following Topo-II Poison AOP events (AOP can be visualised in **Figure 6.3**):

1. MIEs:
    1. The main Topo-II Poison MIE where the compound binds to the DNA-Topo-II cleavage complex. This was characterised by key words in the database which mention "Topoisomerase II Poison" or "Topoisomerase II Inhibition". 8 of the 9 compounds contained these key words in the Derek Nexus database. However, Genistein did not contain any key words to fire the "Topoisomerase II Poisoning" alert despite this property of the compound being well described in the scientific literature. This attests to one of the deficiencies of the database which would need further refinement. The Derek Nexus AOP network in **Figure 6.4** currently has this MIE described as "Topoisomerase II Poisoning" which would need to be corrected to "Binding to the DNA-Topo-II Complex" when this AOP network is integrated into Lhasa's software framework since "Topoisomerase II Poisoning" is not an adequate description for an MIE as these special KEs need to be able to be identifiable and measurable. There is a current link available in the AOP Wiki for this MIE (https://aopwiki.org/events/1252) which provides a starting point for the available test to identify and measure binding to a DNA-Topo-II Cleavage

Complex such as the *in vivo* complex assay (Rodríguez-Cortez and Menéndez, 2020). Further development of this MIE will occur as the HESI GTTC workgroup experts convene to develop this AOP.

2. Another important MIE shown in **Figure 6.4** is the "non-covalent interaction with DNA" MIE which identifies compounds that are DNA intercalators such as the anthraquinone derivatives with highly DNA reactive hydroxylated planar rings (Ribaudo et al., 2019). This MIE accounts for the relatively higher genotoxic potency of these compounds compared to non-intercalative compounds. There is no current description of this MIE in the AOP Wiki. The most promising method for evaluation DNA intercalation comes from (Snyder et al., 2004) who developed an *in silico* 3-dimentional computational DNA-docking algorithm to predict DNA intercalative potential which shows concordance with the Chinese hamster V-79 cell-based bleomycin amplification system, which is a cell-based assay to predict frameshift mutagenicity of intercalative agents (Snyder and Diehl, 2000, Snyder and Strekowski, 1999).

3. Two compounds included key words alluding to "gyrase inhibition" which formed a designated MIE in this AOP network. The same alert corresponded to the quinolone antibiotics ciprofloxacin and flumequine that target DNA gyrase (Evans-Roberts et al., 2016) in bacterial cells where the same compounds exert their pharmacological properties.

4. Two additional MIEs were distinguished based on key words present in the Derek Nexus database, including "estrogen receptor binding" and "tubulin binding". Although both MIEs do not align with the HESI GTTC AOP, these may be important off-target effects for consideration in compound development.

2. KEs:

1. As previously mentioned, some alerts are indicative of oxidative stress related DNA damage mechanisms, namely "uncoupling of oxidative phosphorylation", "radical species generation", and "exogenous oxidant increase". These three KEs were combined into a single KE named "ROS generation". It is suggested that this KE could be linked with the oxidative stress AOP in the AOP Wiki and which would provide users of the Derek Nexus mapped AOP additional tools to assess oxidative stress related damage potential for Topo-II Poison compound exposure.

2. The main Topo-II Poison AOP KE of "DNA double-strand breaks" emulates in this Derek Nexus mapped AOP with the "DNA double-strand breaks (DSB)" KE. This is one of the fundamental KEs from the Topo-II Poison AOP. There is an existing KE in the AOP Wiki for "DNA Double Strand Breaks" (https://aopwiki.org/events/1461) which could be linked to the Derek Nexus AOP network. This KE aligns with the "DNA double strand breaks" KE2 from the HESI GTTC Topo-II Poison AOP **Figure 6.3**.

3. The "DNA double-strand breaks (DSB)" KE links to the "DNA damage repair" KE in the Derek Nexus AOP network. This KE indirectly links with the "inadequate repair" KE4 from the HESI GTTC Topo-II Poison AOP **Figure 6.3**, since DNA damage that is not adequately repaired can lead to chromosome structural damage or inherited mutations. The DNA damage response is indirectly quantifiable in biological test systems by measuring p53 expression since it is well accepted that endogenous and exogenous DNA damage is largely reverted by the action of elaborate DNA-repair mechanisms, largely involving the p53 protein (Lee et al., 1995, Liu and Kulesz-Martin, 2001). Hence, "DNA damage response" can be a defined KE in this AOP network.

4. The "tumour suppressor protein p53 signalling pathway dysregulation" KE is related to the "DNA damage response" KE through the translocation of p53 to the nucleus upon DNA damage. Genistein contained key words matching the "estrogen receptor (ER) binding" MIE in the Derek Nexus database. There are several publications that report on the coordinated link between ER inhibition and p53 dysregulation (Konduri et al., 2010, Hurd et al., 1997, Dinda et al., 2002), especially in ER+ breast cancers (Bailey et al., 2012). The progression of cells through the cell cycle is blocked by p53 protein translocation to the nucleus in the G1 phase of replication (Levine et al., 1991). Mutations in the p53 protein fail to block the progression of the cell cycle or fail to bind to the p53 tumour-suppressor gene targets which have been demonstrated to be the single most common genetic alterations observed in human cancers (Levine et al., 1991). This KE links to the "malignant neoplasm" AO. This KE can be indirectly measured through the upregulation of p53 protein expression coordinated with ER binding activity.

5. An additional KE "mitotic dysregulation" was identified associated with the "tubulin binding" MIE and "DNA double-strand breaks" KE. With respect to the

latter, this "mitotic dysregulation" KE could be indicative cell cycle delay and mitotic arrest following exposure to Topo-II Poisons. "Micronucleus formation" was separated out as a distinct KE since the associated assays can be used to distinguish micronucleus formation because of aneugenicity or clastogenicity aligned with this AOP by using the cytokinesis-block micronucleus (CBNM) assay in combination with fluorescence in situ hybridization (FISH) technique using specific centromeric probes for chromosomes (Chung et al., 2002).

3. AOs:

   1. The AOs "chromosome structural damage" and "inherited DNA mutation" align with the HESI GTTC Topo-II Poison AOP **Figure 6.3** and have assays from the standard test battery associated with them. This enforces the fact that the standard test battery recognises the adverse outcome whilst not considering much of the preceding biological events.

   2. The ultimate "malignant neoplasm" AO takes this AOP network further than the HESI GTTC AOP which is more focused on genotoxic outcome. Including "malignant neoplasm" in this instance maps the alert for rodent carcinogenicity and shows the outcome on the organism level.

   3. An interesting other AO was "mitochondrial dysfunction" which was fired by the anthraquinone molecules based on their effects on collapse of the mitochondrial membrane potential and activation of caspases (Yan et al., 2008).

6.4.2 KEs Aligned with MultiFlow and ToxTracker Biomarkers

The key feature of the Derek Nexus AOP network was the intuitive mapping of the MultiFlow and ToxTracker biomarkers to the AOP KEs. This mapping demonstrates "how can this KE be measured or identified" according to one of the AOP key principles (OECD, 2018). It was already acknowledged in the HESI GTTC proposed Topo-II Poison AOP that γH2AX foci can detect DNA double-strand breaks, and that the indirect measure of inhibition of replication can be obtained from the ToxTracker Bscl2-GFP biomarker (Sasaki et al., 2020). This Derek Nexus AOP network mapping provides additional KEs in the AOP network that novel biomarkers can be used to identify and measure the KE. Namely:

1. Mapping of the ToxTracker Rtkn-GFP and MultiFlow γH2AX biomarkers to the "DNA double-strand breaks (DSB)" KE. Topo-II Poisons increase the forward rate of DNA double-strand break formation and inhibit DNA double-strand break repair, leading to the

manifestation of poorly tolerated breaks. Phosphorylated histone H2AX (γH2AX) is produced as a response to nuclear DNA double-strand breaks. The MultiFlow assay uses anti-γH2AX-Alexa Fluor® 647 to directly measure γH2AX expression. The Rtkn-GFP biomarker is responsive to DNA double-strand breaks that recruit ATM for DNA repair (Hendriks et al., 2015) and is strongly upregulated by genotoxic agents that cause DNA damage.

2. Mapping of the MultiFlow p-H3 biomarker to the "mitotic dysregulation" KE. The phosphorylation of histone H3 at serine 10 is fundamental in both chromatin de-condensation during transcriptional activation (Clayton et al., 2000; Nowak and Corces, 2000) and chromosome compaction during cell division (de la Barre et al., 2000). As the cell cycle progresses through G2 checkpoints, strong correlation is observed between histone H3 phosphorylation and chromatin condensation (Van Hooser et al., 1998). The MultiFlow assay uses an anti-phospho-histone H3-PE marker to quantify the frequency of mitotic cells (Bryce et al., 2016).

3. Mapping of the MultiFlow p53 biomarker to the "DNA damage repair" KE. The scientific community recognises that DNA damage can be reverted by the action of elaborate DNA-repair mechanisms, many involving the p53 protein (Lee et al., 1995; Liu and Kulesz-Martin, 2001). Coupled with cell cycle control, p53's role in the DNA double-strand break repair pathways, Non-Homologous End Joining (NHEJ), and Homologous Recombination (HR) is especially significant. The more error-prone process, NHEJ, is active throughout the cell cycle, however more prominent during G1 where the outcome seems to be *via* effects on halting apoptosis (Gao et al., 1998; 2000). HR is predominantly present during late S and G2 with coordinated DNA-DSB repair in synergy with ATM/ATR checkpoint kinases (Romanova et al., 2004; Boehden et al., 2005). The cellular responses to Topo II inhibitor-mediated genotoxicity include nuclear p53 translocation in response to DNA-DSBs. As the cell cycle progresses through G1 to S and G2 phases, p53 acts in a coordinated manner with the cell cycle checkpoints to induce NHEJ- and HR-mediated DNA double-strand break repair and recombination. In concomitant response with γH2AX, p53 expression indicates a response to genome damage. Since the MultiFlow assay measures biomarker expression in detergent-liberated nuclei, p53 nuclear translocation resulting from genotoxic stress can be directly measured using the MultiFlow assay's anti-p53-FITC reagent that labels the N-terminal domain of

p53. The ToxTracker Btg2-GFP biomarker activation corresponds to activation of the p53 response upon DNA damage and can be an indirect measure of p53 activation.

Overall, the Derek Nexus database rearranged into an AOP network shows the types of DNA damage that can be measured versus the current regulatory paradigm of measuring the adverse outcome. It is important to acknowledge the complex crosstalk between some of the biomarkers, for example the preferential upregulation of Rtkn-GFP is coordinated with NF-κB cytokine signalling and is associated with compounds that have an aneugenic mode of action; however, concomitant upregulation of the Bscl2-GFP biomarker suggests that the ATR-Chk1 DNA damage signalling pathway is activated due to stalled replication forks (Hendriks et al., 2015). There is no "stalled replication forks" KE in the Derek Nexus Topo-II Poison AOP Network **Figure 6.4** since there were no key words present in the database to suggest the same KE, however it is suggested that this KE can be added in future work by Lhasa since there is evidence that Bscl2-GFP activation is an indirect measure of stalled DNA replication forks (Sasaki et al., 2020). This KE would be directly before the "DNA double-strand breaks (DSB)" KE in the Derek Nexus Topo-II Poison AOP Network.

6.4.3 A QSAR Model for KE Associated DNA Damage

The expert mapping of the Derek Nexus key words into an AOP network led to developing the secondary hypotheses aimed at using the KE associated tests to develop a proof-of-concept QSAR model to predict compound potency. In this respect, the following questions were asked:

a. Do compound structural properties correlate with potency, relative to activating different Topo-II Poison AOP KEs?
b. Can the potency of a nearest neighbour be predicted based on structural activity relationships (SAR)?

With the aim of answering these questions, the molecular weight. LogP, and Polar Surface Area molecular descriptors were selected for developing a QSAR model predictive of Topo-II Poison compound potency from a BMD metrics training set. The QSAR model was built using the K-Nearest Neighbours algorithm and nearest neighbour distances were evaluated for performance of the model compared to BMD confidence interval based comparative potency measurements that were derived in the previous results chapter.

6.4.3.1 Influence of BMD Metric on Model Outcome

The K-Nearest Neighbours model could only accommodate a single type of BMD metric (BMDL, BMD or BMDU) into the 3-dimensional data space matrix. As previously mentioned, this violates the principle advocated in the scientific literature that the full BMD confidence interval shall be utilised for comparative potency analyses (Wills et al., 2016b; 2017). Nonetheless, the impact of selecting either BMD metric on the K-Nearest Neighbours distances was evaluated. **Figure 6.6** to **Figure 6.11** shows the impact of using either a BMDL, BMD or BMDU value on the K-Nearest Neighbours distances when factored with the combination of molecular descriptors assessed in this QSAR model development. When comparing the K-Nearest Neighbours distances between compounds plotted across the three panels (corresponding to BMDL, BMD, and BMDU values – left to right) it is evident that the choice of BMD metric does not influence the K-Nearest Neighbours distances observed between plotted compounds, irrespective of the molecular descriptor selected. The Rtkn-GFP biomarker was selected as the BMD confidence intervals were consistently wider than other biomarkers studied and presented in the previous results chapter. Additionally, the Bscl2-GFP biomarker was selected since the BMD confidence intervals widths across the compound studies were more variable than those obtained from other biomarkers studied and presented in the previous results chapter. It is evident from the K-Nearest Neighbours matrix plots in **Figure 6.6** to **Figure 6.11** that the impact of the choice of BMD metric does not impact the outcome of K-Nearest Neighbours distances even when there are wide or variable BMD confidence interval widths. These observations are based on superimposable K-Nearest Neighbours matrix plots within each figure where the relative position of each plot does not alter based on the selection of either BMD metric. This is likely because BMD confidence intervals in all cases were relatively small, spanning a maximum of 1-2 Log units where the BMD point estimate is approximately the midpoint. It can be suggested that the impact of selection of BMD metric will be greater in instances with confidence intervals where the BMD point estimate is not the midpoint, with relatively large positive direction confidence intervals. However, in this proof-of-concept approach, the BMD value was arbitrarily selected for further analysis since there was no impact of selection of BMD metric on compound K-Nearest Neighbours plots in the superimposable matrix spaces.

6.4.3.2 Compound Structural Properties for Predicting BMD Potency

In general, the K-Nearest Neighbours analysis results of each MultiFlow or ToxTracker biomarker response presented in **Figure 6.12** to **Figure 6.16** show that the molecular descriptors perform well in correlating compound structure with BMD potency. Each figure shows the results of including the BMD value and the iterative combination of molecular descriptors (LogP vs. Log Polar Surface Area, Log Polar Surface Area vs. Log Molecular Weight, and Log Molecular Weight vs. LogP) in the K-Nearest Neighbours analysis training sets. From each figure the general trend is based on some common observations: where the greatest distances exist between the anthraquinone molecules and the quinolone antibiotics; emodin and genistein show the least K-Nearest Neighbours distance and cluster in the same order of magnitude in the MultiFlow biomarker responses where they were both studied; and, there is a distinct cluster of compounds belonging to the potent podophyllotoxin compounds (teniposide, etoposide) and the potent anthraquinone compounds (mitoxantrone, doxorubicin, and daunorubicin) and all cluster where the K-Nearest Neighbours distance between the members of the cluster is smaller than the K-Nearest Neighbours distance between them and the cluster of emodin and genistein.

A specific observation relates to the relative performance of the different combinations of molecular descriptors based on K-Nearest Neighbours distances compared to observed BMD potencies in the *in vitro* genotoxicity biomarkers. Compounds with similar molecular weights (emodin and genistein) and compounds with similar polar surface area measurements (teniposide, etoposide, and mitoxantrone) form clusters in the same order of magnitude for BMD potency when both measurements (Log Molecular Weight and Log Polar Surface Area) are included in the K-Nearest Neighbours models. This observation is illustrated in **Figure 6.19** where similar molecular weight and polar surface area measurements cluster in 2D and 3D KNN matrix plots. The BMD confidence interval plots of the same compounds are shown for the 24-hour γH2AX biomarker response where the BMD potencies cluster in the same manner. This suggests that measurements of polar surface area and molecular weight may perform as the best molecular descriptors to predict the genotoxic potencies of the compounds studied in this results chapter. It is also interesting to observe that flumequine has a similar molecular weight to emodin and genistein, however has a considerably lower polar surface area measurement than the same two compounds. This translates to greater K-Nearest Neighbours distances for flumequine compared to emodin and genistein, and a

flumequine BMD potency several orders of magnitude different than the same two compounds. This reinforces that molecular weight and polar surface area measurements in tandem may be superior BMD genotoxic potency prediction descriptors.
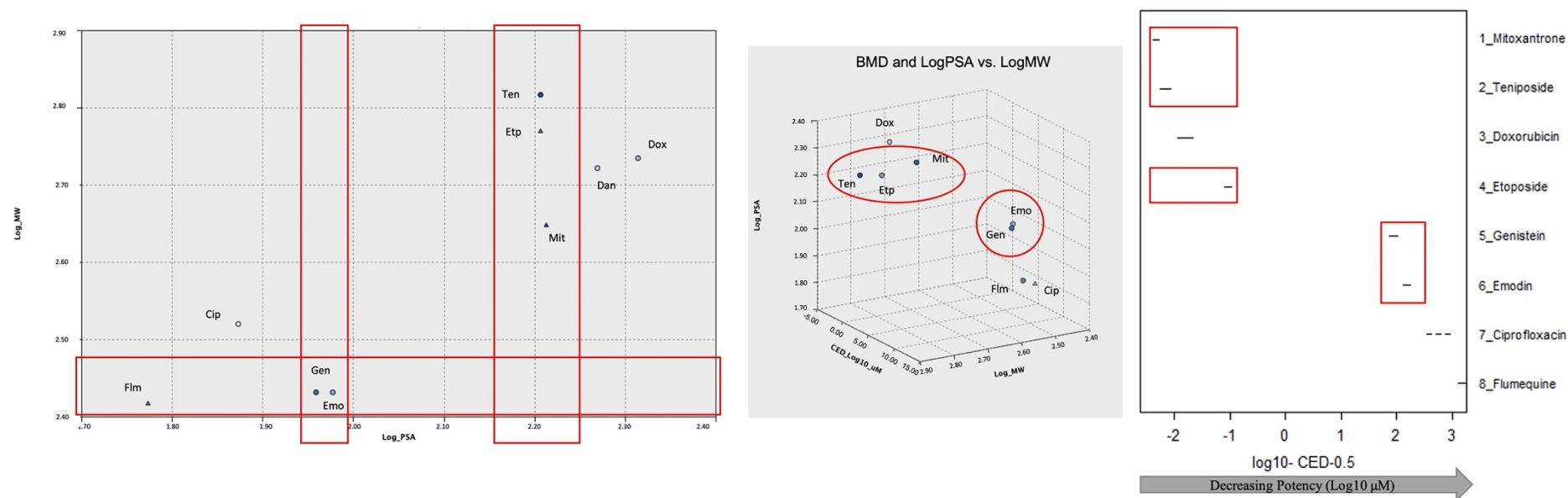
Figure 6.19. Illustration of clusters of emodin and genistein; and teniposide, etoposide and mitoxantrone based on their molecular weight (LogMW) and polar surface area (LogPSA) measurements. The left panel shows the K-Nearest Neighbours distances in a 2D matrix space based on LogPSA (*x*-axis) and LogMW (*y*-axis). Centre panel shows the same molecular descriptor metrics with the addition of BMD in a 3D matrix space plot. The comparison of KNN distances with BMD in the training set match the K-Nearest Neighbours distances when BMD is not included in the analysis. The right panel shows BMD confidence intervals in the 24-hour γH2AX biomarker response. Similar K-Nearest Neighbours distances and clusters are observed in the 2D and 3D KNN models as the BMD confidence interval plot for the same highlighted compounds. Flumequine with a similar molecular weight to emodin and genistein, however a marginally different polar surface area measurement shows distinctly different BMD potency in the right panel, and a larger K-Nearest Neighbours distance than emodin and genistein in the centre panel. Abbreviations: PSA = Polar Surface Area; MW = Molecular Weight.

It can therefore be concluded that the molecular descriptors can be used to predict genotoxic potency of a compound based on its nearest neighbours in a training set. This can be illustrated by taking the plot of doxorubicin in a 2D matrix as an example and comparing the K-Nearest Neighbours distances of its nearest neighbours (**Figure 6.20**). If we imagine that we only know the molecular weight and polar surface area descriptors of doxorubicin, we could predict its relative potency rank position based on known BMD potencies of its K-nearest neighbours (mitoxantrone, etoposide, and teniposide). For example, the built 3D K-Nearest Neighbours matrix plot for γH2AX induction shows that mitoxantrone and teniposide are more potent that etoposide. The 2D K-Nearest Neighbours matrix plot for Log Molecular Weight and Log Polar Surface Area for the same compounds plus doxorubicin shows that the K-Nearest Neighbours distance between doxorubicin and etoposide (0.5) is less than the K-Nearest Neighbours distances between doxorubicin and teniposide (0.6) and doxorubicin and mitoxantrone (0.6) would suggest that doxorubicin's potency is closer to etoposide than teniposide and mitoxantrone, hence one could comparatively rank doxorubicin's potency for DNA double strand breaks relative to etoposide, teniposide and mitoxantrone.
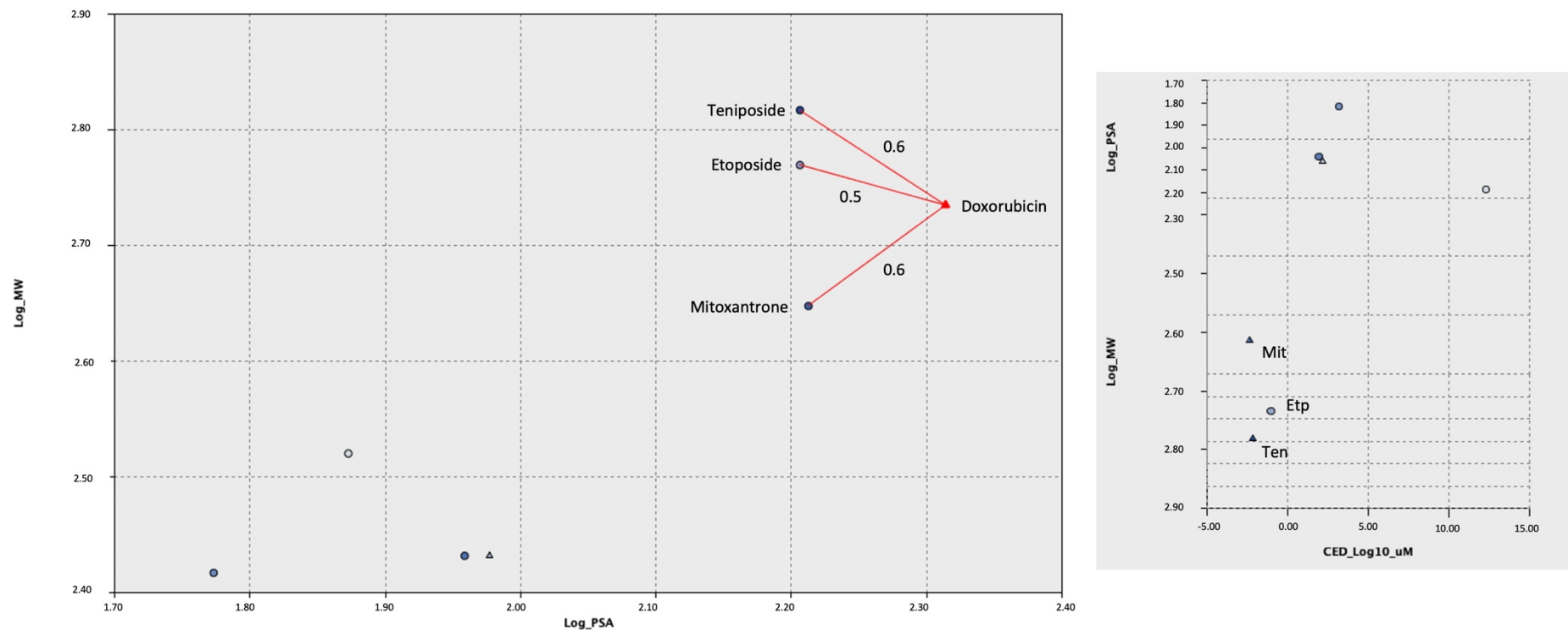
Figure 6.20. Left panel: The K-Nearest Neighbours distances are highlighted for the Topo-II Poison compounds doxorubicin, teniposide, etoposide, and mitoxantrone on a 2D matrix plot based on polar surface area (LogPSA) and molecular weight (LogMW) molecular descriptor variables. Right panel: 3D K-Nearest Neighbours matrix plot (BMD, LogPSA, LogMW) rotated on its axis to show BMD potency on the *x*-axis in a manner that the rank order potency of the highlighted compounds can be visualised (Mit>Ten>Etp). One can place the relative position of doxorubicin's potency based on the 2D K-Nearest Neighbours distances relative to the known potencies of the 3 other compounds and their K-Nearest Neighbours distances. Abbreviations: Mit = mitoxantrone; Ten = teniposide; Etp = Etoposide. Abbreviations: PSA = Polar Surface Area; MW = Molecular Weight.

6.5 Concluding Remarks: Derek Nexus AOP Network and Proof-of-Concept Predictive *in vitro* Genotoxicity QSAR Model Development

The HESI GTTC proposed AOPs are intended to provide the basis for organising key biological events associated with genotoxic outcomes (Sasaki et al., 2020) as well as outlining the tests that can be used to identify and measure the occurrence of key events in biological test systems. The scientists at Lhasa describe how an AOP network can advance compound risk assessment (Ball et al., 2021) by imagining some use cases that could emerge from the application of AOPs to risk assessment, including: 1) what genotoxic properties exist of similar compounds?; 2) what AOP is applicable to a query compound?; and 3) what biological tests are appliable to the AOP? The organisation of key words from the Derek Nexus knowledge database into an AOP network makes the first practical use of the proposed HESI GTTC Topo-II Poison AOP and takes it beyond the current conceptual framework that the AOPs currently stand. The Derek Nexus AOP network shows the chemical stressors which have been included in this first example use case, the associated MIEs, KEs, and AOs based on similar key words shared amongst the compounds and the knowledge used to build the AOP network. The associated standard battery biological tests and novel *in vitro* genotoxicity biomarker assays have been intuitively added to the AOP network KEs and AOs. It is suggested that the Derek Nexus knowledge database can be improved to include replication fork collapse as a key word including an indication that Bscl2-GFP activation is an indirect measure of stalled DNA replication forks (Hendriks et al., 2015; Sasaki et al., 2020). The Derek Nexus Topo-II Poison AOP network that has been developed in this results chapter can be included in Lhasa's Kaptis software platform (https://www.lhasalimited.org/products/kaptis.htm) that is intended to combine AOP knowledge in a meaningful way to be used in risk assessment (Ball et al., 2021).

The development of a proof-of-concept QSAR model using the KE associated tests from the Derek Nexus AOP network has been reported in this results chapter. For example, BMD potency metrics from γH2AX and Rtkn-GFP biomarker responses associated with the DNA double strand breaks KE in the AOP network were used to build genotoxic potency predictive models based on simple effective molecular descriptors. The models performed well based on the LogP, molecular weight, and polar surface area molecular descriptors that have been demonstrated to perform well in QSAR predictive algorithms used elsewhere in genotoxicity systems, particularly for Ames mutagenicity predictions (Mersch-Sundermann

et al., 1994; Votano et al., 2004). It is suggested that the molecular weight and polar surface area descriptors perform best for predictive potency determination – which could be explained by the combination of the descriptor's account of nucleus membrane permeability, surface of reactive atoms, and molecular similarity. The use of the same molecular descriptors in this results chapter, together with BMD potency metrics from the study of Topo-II Poison compounds in multiplexed *in vitro* genotoxicity systems demonstrates the utilisation of molecular descriptors in building predictive algorithms for genotoxicity measures that are broader than just mutagenicity. In doing so, it was possible to fulfil the hypotheses of whether compound structural properties correlate with potency relative to activating the different Topo-II AOP key events, and whether potency of a nearest neighbour could be predicted using SAR techniques.

This approach is novel in the sense that QSAR and BMD metrics are utilised in a single predictive potency algorithm. A concern from one of the assay developer's (Litron, personal communication) is that the QSAR predictive models may one-day replace the need to perform the benchtop experiments using their assay. For this goal to ever be reached, it is suggested that it would be necessary for significant advanced validation and expansion of the QSAR model presented in the proof-of-concept method presented in this result chapter on a relatively small group of specific compounds. Comparing this idea with the industry use case where QSAR methods can be used *in lieu* of performing an Ames test for pharmaceutical impurities under ICH M7 (ICH, 2013) demonstrates that it could take decades for regulation to catch up. Moreover, the Ames test is still widely used today more than 40 years after its development.

Chapter 7. General Discussion.

7.1 General Discussion

The purpose of the project was to investigate the application of the Benchmark Dose (BMD) approach to novel *in vitro* genotoxicity test systems. This has important implications because:

1. BMD modelling had only previously been performed on 'gold standard' genotoxicity endpoints such as *in vitro* micronuclei, where limited mechanistic information can be derived.
2. Novel flow cytometric *in vitro* genotoxicity tests had been initially developed as a follow up to positive micronucleus tests (Bryce et al., 2014).
3. There is a real impetus for a paradigm shift from genotoxic hazard identification to risk-assessment which involves the use of novel *in vitro* test systems that can inform mechanism of action information for detailed genotoxicity risk assessments.

In addition, this study provides an in-depth stepwise application of BMD modelling to specific *in vitro* genotoxicity systems that demonstrates the use of the biomarkers dose-response datasets for potency and MOA determinations. This is consistent with the quantitative dose-response analysis framework of genetic toxicology where the data informs a decision beyond genotoxic/non-genotoxic (Gollapudi et al., 2013b). This all-encompassing report realises some of the main themes from the 'next generation testing strategy for assessment of genomic damage' from Dearfield et al. (2017), specifically:

1. Modern test systems that can be used to determine compound MOA information.
2. Use of quantitative dose-response modelling to genotoxicity datasets.
3. Quantitative Structural Activity Relationships (QSAR) for read across in genotoxicity risk assessments.

7.2 Findings of this Thesis

The following sub-sections include an account of the main findings of this thesis and their implications in genotoxicity dose-response modelling for compound potency determinations in risk-assessments. A summary of the conclusions from each of the results chapters are:

- *In vitro* flow cytometric genotoxicity biomarker (MultiFlow and ToxTracker) dose-response datasets are suitable for BMD analysis to derive compound potency determinations that correlate well with the gold-standard micronucleus endpoint (chapters 3 and 5).

- Use of BMD point estimates in comparative potency analyses mischaracterises or overestimates potency differences. By not considering the BMD confidence intervals which include measures of dose-response uncertainty, false assumptions or misleading results pertaining to compound potency can be derived. The BMD confidence intervals should be considered (chapter 4).

- Compound structure supports BMD potency determinations from *in vitro* genotoxicity biomarker responses for the Topoisomerase II (Topo-II) Poison class of compounds (chapter 5). The application of the BMD dose-response methodology to compound classes of interest supports compound classification and leads to opportunities for read-across.

- The Derek Nexus database can be reorganised into an Adverse Outcome Pathway (AOP) network for the alerts relating to Topo-II Poison compounds (chapter 6).

- A QSAR model built on molecular descriptors of the Topo-II Poison compounds predicts genotoxic potency based on BMD measurements of the AOP Key Events (chapter 6). The application of this methodology provides a predictive quantitative genotoxicity read-across method that is superior to the current categorical QSAR read-outs, for example Ames mutagenicity.

7.2.1 BMD Modelling is Applicable to *In Vitro* Flow Cytometric Genotoxicity Biomarker Dose-Response Data for Compound Potency Determinations (Chapter 3 and Chapter 5).

There is overwhelming support from the expert scientific community that the BMD approach is the most appropriate to describe dose-response relationships of genotoxicity datasets (Gollapudi et al., 2013b; Macgregor et al., 2015). Previously, BMD modelling of *in vitro* datasets extended as far as characterising compound potency from standard endpoints – such as micronucleus – and correlating the potency with an *in vivo* comparison (Soeteman-Hernández et al., 2015; Bemis et al., 2016; Soeteman-Hernández et al., 2016). With the advent of the novel flow cytometry-based biomarker assays for characterising DNA damage such as Litron's MultiFlow and Toxys' ToxTracker assay (Johnson et al., 2014b; Hendriks et

al., 2015; Bryce et al., 2016), it was imperative to determine if the resulting datasets were amenable to BMD analysis for comparative potency determinations beyond the positive micronucleus follow up testing that one of the assays had been developed for (Bryce et al., 2014). Chapter 3 of this thesis reports the application of the BMD approach to MultiFlow DNA damage response biomarkers (γH2AX, p53, and p-H3) for a series of aneugens and clastogens, and determination of comparative compound potency correlations with the *in vitro* micronucleus endpoint for the same compounds. Chapter 5 also shows how BMD potency determinations can be derived from the ToxTracker biomarker responses for DNA damage (Rtkn-GFP, Bscl2-GFP, and Btg2-GFP). Through BMD potency correlations it was demonstrated that the MultiFlow *in vitro* DNA damage response biomarkers are predictive of potency in the *in vitro* micronucleus assay (**Figure 3.39** to **Figure 3.41**). This has implications in high-throughput screening laboratories in drug development where potency predictions of candidate compounds could be made before investing in costly studies required for regulatory submissions. For example, an R&D department investigating 1000+ candidate compounds could screen for *in vitro* mechanistic genotoxicity information and correlate the candidate compound's potency to predict a regulatory genotoxicity assay outcome, including *in vitro* and *in vivo* micronucleus endpoints. This approach could reduce the risk of failure during late preclinical phases of drug development programs.

7.2.2 *In Vitro* Genotoxicity Biomarker Derived Comparative Potency Analyses Should Consider the BMD Confidence Intervals (Chapter 4)

Since the utility of BMD modelling to the MultiFlow biomarker dose-response datasets was demonstrated and published together with collaborating scientists at Litron (Dertinger et al., 2019), researchers began using the BMD method to test different comparative potency hypotheses. One example in the scientific literature was the use of BMD modelling of the MultiFlow dose-response datasets to assess the impact of low concentration S9 on the potency of several clastogens requiring an exogenous source of metabolic activation (Tian et al., 2020). The researchers used BMD point estimates to make their comparative potency conclusions. An associated shortcoming was identified in results chapter 4 of this thesis where the same researchers did not consider dose-response uncertainty or BMD confidence intervals in their comparative potency analyses. Prominent experts in the field have described that potency differences can only be alleged when BMD confidence intervals do not overlap (Wills et al., 2016a; 2016b, 2017). Using the same logic, Chapter 4

reports on a re-analysis of the Tian et al. (2020) MultiFlow dataset but uses BMD confidence intervals in the potency comparisons. It was demonstrated that the use of BMD point estimates in the same dataset mischaracterised or overestimated the impact of S9 on genotoxic potency. The use of BMD confidence intervals remained a contentious topic in discussion with collaborating scientists at Litron, who expressed that some of their clients prefer single numerical values of comparison. To meet this demand and test a hypothesis if it was possible to combine BMD confidence intervals with simple means of comparison, a novel algorithm was developed, and the output values used to perform unsupervised hierarchical clustering. The results of the clustering showed that distinct groups exist that best describe the range of S9-dependent effects into high, low, zero and negative categories (**Figure 4.29**). The method of performing statistical evaluation through hierarchical clustering objectively considering the range of potency differences, removing the influence of bias from the evaluating genetic toxicologist. The findings presented in results chapter 4 has implications in aiding with data interpretation where it is speculated that a genetic toxicologist could be overwhelmed with the results from hundreds of compounds from *in vitro* screening experiments.

7.2.3 Compound Structural Information Supports BMD Potency Ranks for the Topo-II Poison Class of Compounds (Chapter 5).

The basis of results chapter 5 was to use the *in vitro* genotoxicity biomarker dose-response datasets to perform robust potency analyses of compounds within a MOA of interest. The results chapter combined BMD derived potency information and *in vitro* genotoxicity biomarker (MultiFlow and ToxTracker) dose-response datasets for the Topo-II Poison class of compounds. This compound class was selected since it presented a special interest within the HESI GTTC MOA workgroup. The *in vitro* genotoxicity biomarkers responded as expected upon assessment of cells exposed to Topo-II Poisons. Specifically, the Topo-II Poison mechanisms of action were demonstrated in a dose-dependent manner including replication fork collapse, DNA double strand breaks, cell cycle delay and mitotic arrest. By using BMD modelling, it was possible to determine potency ranks for the 9 Topo-II Poison compounds that were studied. A prominent article by Wills et al. (2016a) used compound structural information to account for BMD derived potency differences of benzimidazole compound evaluated *via* the *in vitro* micronucleus endpoint. The same approach was applied to the Topo-II Poison compounds in this study using compound

structural information derived from an expert rule based QSAR system (Derek Nexus). The data obtained in chapter 5 shows that Topo-II Poison compound potency differences can be accounted for based on their structural differences. In general, for all *in vitro* genotoxicity biomarker dose-responses: the podophyllotoxin and anthraquinone derivative compound grouped in the more potent end of the scale; the structurally similar podophyllotoxin emodin to the isoflavone genistein was considered equipotent; and the anthraquinone antibiotics were the least potent. Further differences observed within the podophyllotoxin and anthraquinone compounds were scrutinised based on subtle differences in their compound structures. Results chapter 5 demonstrated that combining compound structural activity information with BMD dose-response potency information derived from the *in vitro* genotoxicity mechanistic biomarkers adds significant information to support the genotoxicity risk assessment related to the exposure of specific compounds. This was the first instance where quantitative potency information had been applied to compound genotoxicity information derived from the Derek Nexus database. In fact, the broader implication of work presented in chapter 5 could be applied to read-across where toxicological information from a characterised molecule is applied to a molecule with similar structure with lacking toxicological data. Substance grouping and read-across is one of the main alternatives for data gap filling that reduces the number of animals used in toxicological research in the European Union (Schultz et al., 2015; Ball et al., 2016), which has gained regulatory acceptance in the field of industrial chemical registration, cosmetics development. However, little guidance or regulatory acceptance of modern read-across methods (except for Ames bacterial mutagenicity QSAR predictions) exist for genotoxicity risk from exposure to pharmaceutical compounds or their impurities. This is one of the central issues to regulatory toxicology testing, that the industry experts try to progress, but are tied to the current regulations, and any gains that are made end up being diminished by ensuring that industry adheres to out of date policies and practices.

7.2.4 The Derek Nexus Database can be Mapped to an AOP Network for the Topo-II Poison Compounds (Chapter 6).

MOA is synonymous to the AOP framework, except that MOAs are specific to the description of adverse outcomes (AOs) for individual chemicals or chemical groups whereas AOPs can be applied to a variety of agents. The definition of an AOP from (OECD, 2018) is "a sequence of events commencing with initial interaction(s) of a stressor with a biomolecule within an organism that causes a perturbation in its biology (i.e., molecular initiating event,

MIE), which can progress through a dependent series of intermediate Key Events (KEs) and culminate in an Adverse Outcome (AO) considered relevant to risk assessment or regulatory decision making". Chapter 6 was concerned with utilising the proposed HESI GTTC AOP for Topo-II Poisons previously described by (Sasaki et al., 2020). Previous discussions with the scientists at Lhasa Ltd led to the development of the hypothesis of whether the Derek Nexus knowledge database could be reorganised into an AOP network. To this end, chapter 6 reports the comprehensive mapping of the Derek Nexus database based on key words into an AOP format aligned with the HESI GTTC proposed Topo-II Poison AOP. The major KEs of "DNA double strand breaks" and "inadequate repair" are included in the Derek Nexus AOP as "DNA double-strand breaks (DSB)" and "DNA damage repair". However, the main shortcoming of the Derek Nexus based AOP is related to the lack of key words associated with the "disrupted replication forks" KE proposed by the HESI GTTC AOP. It was suggested that the Derek Nexus knowledge base could be improved to include the same key words associated with disrupted replication forks or replication fork blockage due to replication stress from the mechanism of Topo-II poisoning (Hendriks et al., 2015; Sasaki et al., 2020). The main advantage of performing the Derek Nexus mapping was to include visualisation of KEs in contrast to the AOs usually associated with compound exposure in the MOA framework. It was then possible to associate the *in vitro* genotoxicity biomarkers with the KEs in the AOP network. To this end, the MultiFlow γH2AX and ToxTracker Rtkn-GFP biomarkers were mapped to the "DNA double-strand breaks (DSB)" KE; the MultiFlow p-H3 biomarker was mapped to the "mitotic dysregulation" KE; and the MultiFlow p53 and ToxTracker Btg2-GFP biomarkers were mapped to the "DNA damage repair" KE. A further KE could be added if the Derek Nexus database was improved to include knowledge related to "replication stress" or "disrupted replication forks". The ToxTracker Bscl2-GFP biomarker could be associated with the same KE since activation of the same biomarker suggests that the ATR-Chk1 DNA damage signalling pathway is activated due to stalled replication forks (Hendriks et al., 2015). The development of this AOP network using the Derek Nexus knowledge base forms the first practical use of the HESI GTTC proposed Topo-II Poison AOP. The AOP network could be integrated into the Lhasa Kaptis system (https://www.lhasalimited.org/products/kaptis.htm) which is proposed to be developed for storing and accessing AOPs for risk assessment purposes.

7.2.5 Predictive genotoxic potency QSAR models can be built based on the Topo-II Poison AOP KEs (Chapter 6).

Chapter 6 includes the development of a proof-of-concept QSAR method for predicting compound genotoxic potency relative to the types of DNA damage measured for each KE in the Topo-II Poison AOP. The following hypothesis questions were developed and presented in the same results chapter:

    c. Do compound structural properties correlate with potency, relative to activating different Topo-II Poison AOP KEs?

    d. Can the potency of a nearest neighbour be predicted based on structural activity relationships (SAR)?

With respect to hypothesis question a, it was determined that the molecular weight, LogP, and polar surface area molecular descriptors are good structural activity relationship measures that correlate with BMD potency for the 9 Topo-II Poisons presented in the same results chapter. The K-Nearest Neighbours algorithm was used as a statistical regression method for building the QSAR model tested in the same results chapter. The K-Nearest Neighbours distances were used to infer conclusions from the QSAR models in which molecular descriptors were plotted relative to BMD potency. Hypothesis question b was fulfilled based on measures of K-Nearest Neighbours by comparing the trained models containing BMD metrics versus K-Nearest Neighbours models without BMD metrics. It was demonstrated that the potency rank order of a query compound can be determined relative to other compounds in the QSAR model by interpreting the K-Nearest Neighbours distance metrics. This approach is novel in the sense that QSAR and BMD metrics are utilised in a single predictive algorithm. Additionally, the QSAR model developed in chapter 6 utilised measures of potency, which aligns with the paradigm shift towards quantitative genotoxicity risk assessment (Gollapudi et al., 2013b; Macgregor et al., 2015; Dearfield et al., 2017), and therefore more superior than other QSAR models that are only predictive of mutagenic dichotomy (Maunz et al., 2013; Benigni and Bossa, 2018; Roberts et al., 2000). One issue pertaining to the QSAR method developed in chapter 6 was the use of the BMD point estimate due to limitations in the number of variables that can be loaded into the K-Nearest Neighbours algorithm. Although it was determined that the choice of BMD versus the combination of confidence limits (BMDL and BMDU) did not influence the K-Nearest Neighbours distance metrics between plotted points in this instance, it is suggested that lower

quality dose-response datasets may not display the same level of reproducibility than the QSAR model built in chapter 6. This would be more impactful where large levels of uncertainty in the fitted dose response would translate to wider overlapping confidence intervals for compounds in a combined covariate (Wills et al., 2016a). Therefore, it is imperative that only good quality dose-response datasets that yield narrow confidence intervals are used to build QSAR models for predicting genotoxic potency.

7.3 *In Vitro* Genotoxicity BMD Potency Determinations: Implications in Compound Development.

7.3.1.1 Implications to *In Vitro* Genotoxicity Testing?

In vitro genotoxicity testing methods remain an integral part of compound safety assessment. Depending on the assay performed, the tests can be valuable to high-throughput screening for R&D prioritisation, and as part of the standard test battery for regulatory submission purposes. Several publications have demonstrated the utility of dose-response analysis of *in vitro* genotoxicity tests – such as the *in vitro* micronucleus endpoint – in compound potency ranking (Soeteman-Hernández et al., 2015; Bemis et al., 2016; Wills et al., 2016a; Allemang et al., 2018). The findings presented in this thesis have added valuable contributions to the literature with regards to the utility of the BMD approach to compound potency ranks from dose-response datasets from multiplexed *in vitro* genotoxicity biomarker responses. The additions to the literature include demonstrating that compound potency from multiplexed DNA damage biomarkers can inform potency in a standard battery endpoint such as the *in vitro* micronucleus (Dertinger et al., 2019); that statistically robust potency comparisons across different treatment conditions can be made using BMD confidence intervals from *in vitro* DNA damage response biomarkers (Wheeldon et al., 2021); and that BMD robust potency comparisons can be made for a compound class of interest with supporting mechanistic and compound structural information (Wheeldon et al., 2020). Some individuals have suggested the development of *in vitro to in vivo* extrapolation methods, where *in vitro* genotoxicity tests could replace *in vivo* alternatives (White et al., 2020); however, this remains a contentious topic due to pharmacodynamic and pharmacokinetic differences observed between *in vitro* and *in vivo* systems. For now, we advocate that *in vitro* genotoxicity tests – particularly *in vitro* genotoxicity biomarker responses – remain valuable to acquiring data for BMD analysis and potency ranking consistent with prominent researchers in the genetic toxicology discipline (Wills et al., 2016a; 2016b; White et al.,

2020). BMD potency ranking is a highly valuable tool for high-throughput laboratories in compound R&D, with the findings presented in this thesis and published in the literature provide a great foundation for laboratories to use in their development efforts.

7.3.1.2 Implications of AOPs in Genetic Toxicology?

The genetic toxicology community has already embraced the application of MOA knowledge to compound safety assessment (Dearfield et al., 2017); however, MOA remains focussed on evaluating the different types of adverse outcomes (AOs) associated with the different MOAs (clastogenicity, aneugenicity, mutagenicity). The advantage of the AOP framework is that the sequence of perturbing key biological events (KEs) that lead to adverse outcomes are highlighted in the AOP, and this can lead to the use and development of appropriate tests to identify and measure the KE (OECD, 2018; Sasaki et al., 2020). The HESI GTTC MOA workgroup have begun to develop AOPs for the main indirect and non-DNA reactive types of DNA damage, including those related to two key types of aneugenic molecular initiating events (aurora kinase inhibition and tubulin binding) and three associated with molecular initiating events (oxidative stress DNA damage, inhibition of DNA synthesis, and Topoisomerase II Inhibition) leading to clastogenic outcomes. The genetic toxicology community remains intent on developing and embracing the use of AOPs to realise knowledge gaps associated with genotoxic KEs and AOs, and to develop new tests in the framework of Integrated Approaches to Testing and Assessment (IATA) (OECD, 2018) and Non-Animal Methods (NAMs). The findings presented in this thesis relate to a specific application of the AOP framework to structure existing knowledge in a manner which could be used by genetic toxicology risk assessors and regulators, specifically outlining the types of novel *in vitro* DNA damage biomarker responses that apply to KEs for the Topo-II Poison AOP.

7.3.1.3 Implications of QSAR in Genetic Toxicology?

Although the use of QSAR in genetic toxicology has advanced with the adoption of the use of various methods in regulatory guidance (ICH, 2013), its application remains constrained to predicting the outcome of Ames mutagenicity (Barber et al., 2017; Benigni and Bossa, 2018). As has been discussed in this thesis, mutagenicity is not a comprehensive measure of genotoxicity since there are different types of genotoxic adverse outcomes, particularly as compound development advances to avoid DNA reactive molecules. Other

QSAR methods have been successfully developed and are utilised to predict toxicities ranging from skin sensitisation (Macmillan and Chilton, 2019), acute toxicity (Mombelli and Pandard, 2021) and cytotoxicity (Sun et al., 2020). Of note, the adoption of QSAR methods for predicting skin sensitisation potential was possible due to the development of the skin sensitization AOP and international consorted efforts to replace animal testing of skin sensitisation (Maxwell et al., 2014; Chilton et al., 2018). As the development and application of indirect and non-DNA reactive genotoxic AOPs gain grounds, it is suggested that QSAR methods predictive of KEs associated with genotoxic adverse outcomes could be developed. Data presented in this thesis has shown how simple QSAR methods are predictive of genotoxic potency of the main KEs associated with Topo-II poisoning, where BMD was central to the development of the QSAR model. Significant refinement and validation of QSAR models using BMD measurements of comparative compound potency would be needed, and it is difficult to imagine the pace of regulatory change adopting the use universal use of the same methods. However, the application of novel predictive genotoxic QSAR models like the one presented in this thesis could add significant value to a genetic toxicology laboratory. One could imagine the use of the QSAR model to predict the relative potency of a query compound to define the concentration range of a query compound for *in vitro* testing without the need to run prior concentration range-finding experiments. This could improve the throughput of the laboratory and reduce some of the time spent in R&D activities.

7.4 Concluding Remarks

The "Next Generation Testing Strategy for Assessment of Genomic Damage" conceptual framework described by Dearfield et al. (2017) includes using information pertaining to MOA, modern test systems, dose-response modelling, and QSAR to advance genotoxicity risk assessment. This can only be achieved with an accompanying regulatory paradigm shift from a hazard identification to exposure-based risk assessment. The data presented in this thesis provides a strong foundation for advancing the use of the preferred BMD approach with modern test systems like the MultiFlow and ToxTracker DNA damage response biomarkers. This thesis also shows how BMD potency metrics can be used in the context of genotoxic AOPs to develop novel QSAR methods for predicting genotoxic potency of a query compound. To this end, when taken together, the approaches presented in this thesis form the basis of a next generation testing strategy.

7.5 Future Work

For further adoption of the BMD approach to genotoxicity risk-assessment approaches, some of the uncertainties that were highlighted in this thesis would need further development. Therefore, the findings in this thesis would have been strengthened if the following could be further investigated:

- A limited number of Topo-II Poison compounds were studied for comparative potency analysis. A broader set of pharmaceutical Topo-II Poison compounds would add to the findings presented in this thesis by producing a richer profile of potency metrics.
- The KNN method utilised to develop the QSAR method could not accommodate the BMDL and BMDU values in addition to the BMD point estimate used to build the predictive model. This violates the principle that the BMD confidence interval should be considered when determining comparative potency differences. Although it was demonstrated that there was no impact due to the presence of good quality distinct confidence intervals, it was suggested that poorer quality datasets may be unsuitable for QSAR model development. A more advanced QSAR model should be developed to include the BMD confidence intervals.
- Most of the in-depth hypothesis testing in this thesis was focussed on Topo-II Poisons. It would be interesting to see if the same hypotheses could hold valid when tested on compounds with different MOAs, such as aurora kinase inhibitors leading to aneuploidy, or tubulin binders leading to aneuploidy and polyploidy.

Appendices

List of Appendices

Appendix 1. Published Literature (Dertinger et al., 2019)

Appendix 2. Published Literature (Wheeldon et al., 2021)

Appendix 3. Published Literature (Wheeldon et al., 2020)

Appendix 4. Genetic Toxicology Association (GTA) 2019 Annual Meeting. Poster Presentation

Appendix 5. Topo-II Poison 4-hour BMD Dose-Response Curves for MultiFlow Biomarker Responses

Appendix 6. Topo-II Poison 4-hour BMD Confidence Intervals for MultiFlow Biomarker Responses

Appendix 7. Individual Dose-Response Curves for the MultiFlow and ToxTracker Biomarker Responses Resulting from Topo-II Poison Exposure

Appendix 8. Richard Williams Award Abstract

Appendix 9. Topo-II Poison Derek Nexus Network Mapping by Lhasa

Appendix 10. Complex Topo-Poison Derek Nexus Mapped AOP Network

Appendix 1. Published Literature (Dertinger et al., 2019)

Research Article

Predictions of Genotoxic Potential, Mode of Action, Molecular Targets, and Potency *via* a Tiered MultiFlow® Assay Data Analysis Strategy

Stephen D. Dertinger,[1]* Andrew R. Kraynak,[2] Ryan P. Wheeldon,[3] Derek T. Bernacki,[1] Steven M. Bryce,[1] Nikki Hall,[1] Jeffrey C. Bemis,[1] Sheila M. Galloway,[2] Patricia A. Escobar,[2] and George E. Johnson [1]D[3]

[1]Litron Laboratories, Rochester, New York
[2]Merck & Co., Inc., West Point, Pennsylvania
[3]Institute of Life Science, Swansea University Medical School, Swansea University, Swansea, Wales, United Kingdom

The *in vitro* MultiFlow® DNA Damage Assay multiplexes γH2AX, p53, phospho-histone H3, and polyploidization biomarkers into a single flow cytometric analysis. The current report describes a tiered sequential data analysis strategy based on data generated from exposure of human TK6 cells to a previously described 85 chemical training set and a new performance from the ML ensemble: 35/40 (88%) concordance with *a priori* genotoxicity expectations and 21/24 (88%) agreement with expected mode of action (MoA). A second tier applied unsupervised hierarchical clustering to the biomarker response data, and these analyses were found to group certain chemicals, especially aneugens, according to their

INTRODUCTION

serine 139 (γH2AX) to detect DNA double strand breaks, (2) phosphorylation of histone H3 at serine 10 (p-H3) to identify mitotic cells, (3) nuclear p53 content as an indicator

514   Dertinger et al.

of nuclei counts to provide information about treatment-related cytotoxicity and cytostasis. Relative to individual standard *in vitro* genotoxicity assays, an advantage of the MultiFlow method and related high information content assays is that

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy 515
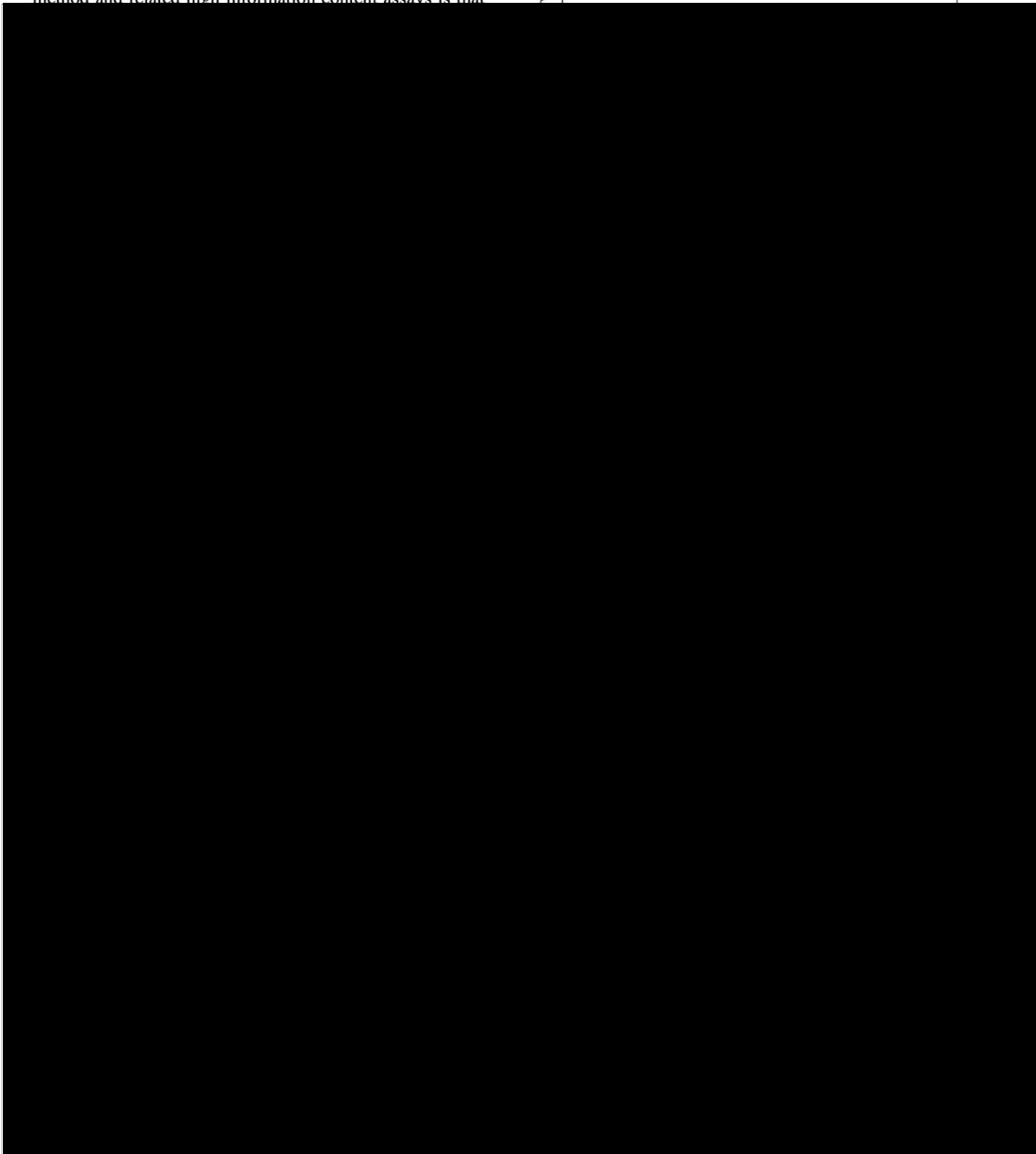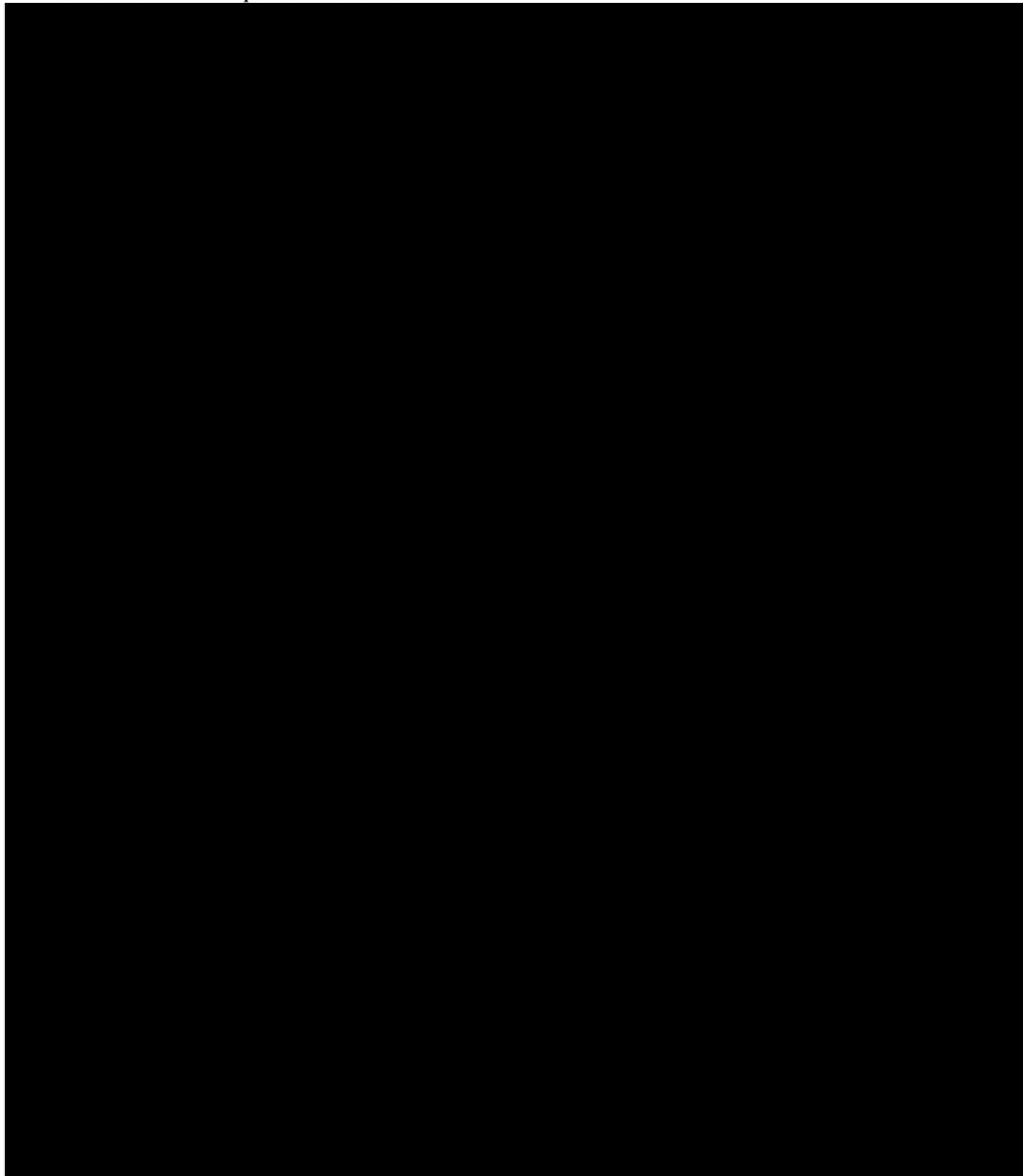
TABLE I. Chemicals and *a priori* Classifications

(2016)
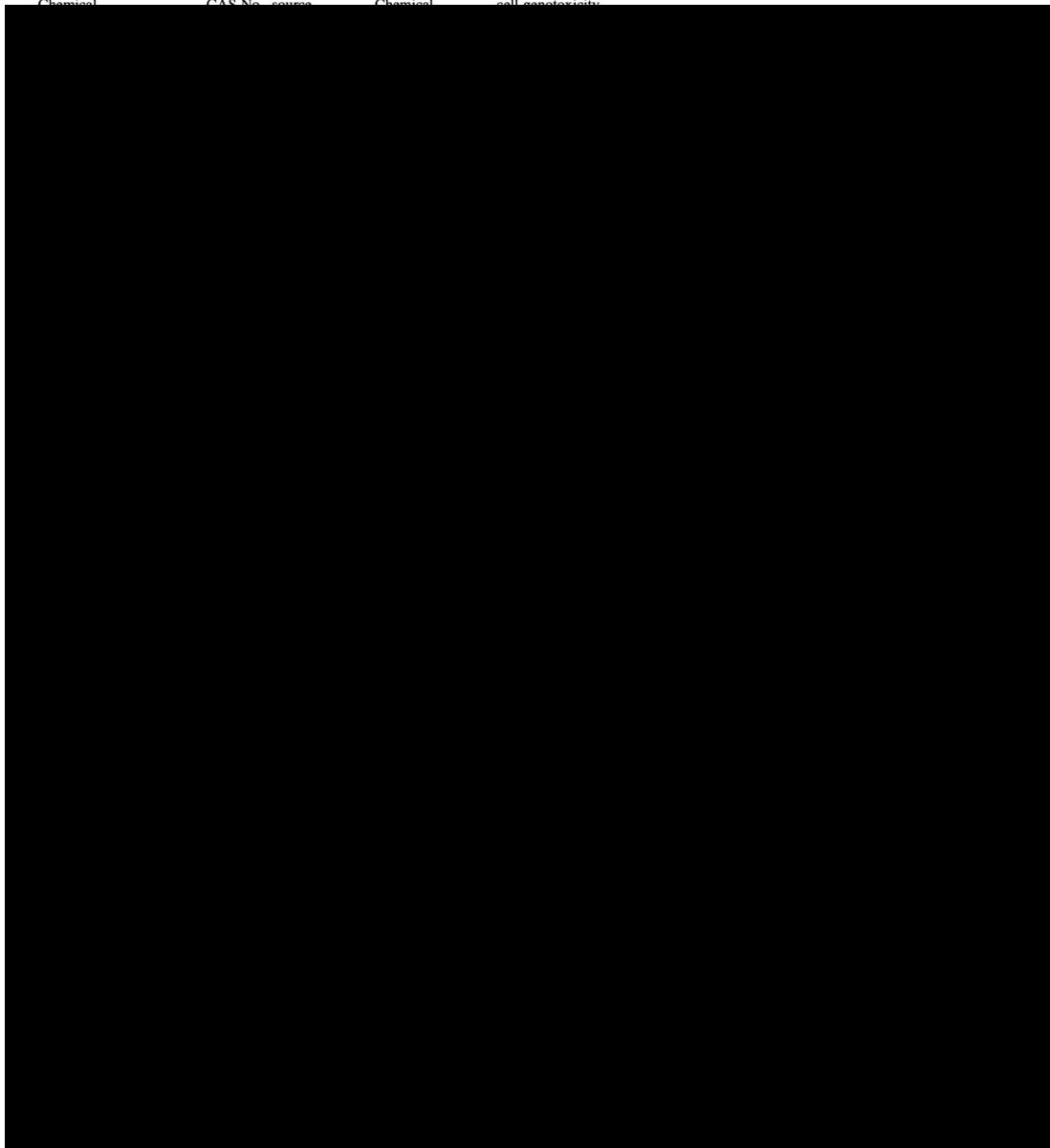
(Continues)

TABLE I. (Continued)

Chemical	CAS No.	source	Chemical	*A priori* mammalian cell genotoxicity

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy   517

TABLE I. (Continued)

A priori mammalian

338

TABLE I. (Continued)

A priori mammalian

MN neg., Rat MN neg., MSD in-house
data

(Continues)

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy 519

TABLE I. (Continued)

Chemical,	CAS No., source	Chemical	*A priori* mammalian cell genotoxicity

TK6 cells were purchased from ATCC® (cat. no. CRL-8015). Cells were grown in a humidified atmosphere at 37°C with 5% CO_2, and were test chemical per well. The highest concentration tested was 1 mM, and the 19 additional concentrations were tested using a square root dilution

scheme—that is, each concentration differed from the one above by a fac-   top concentration was 1 mM or the lowest precipitating concentration,
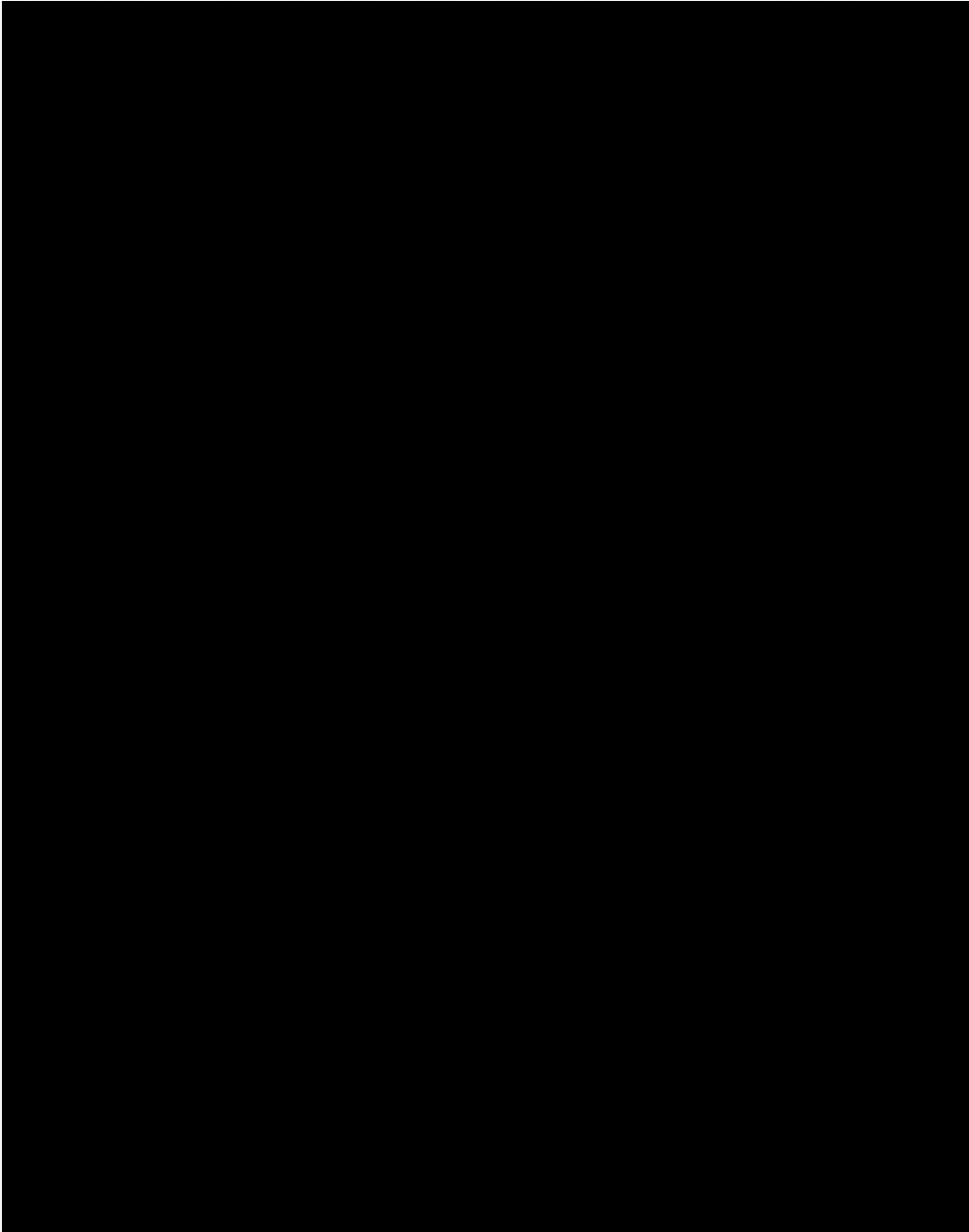
14h and Top as noted above, in the absence of excessive cytotoxicity the   gen MoA, required two successive concentrations to exhibit alleugen

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy   521
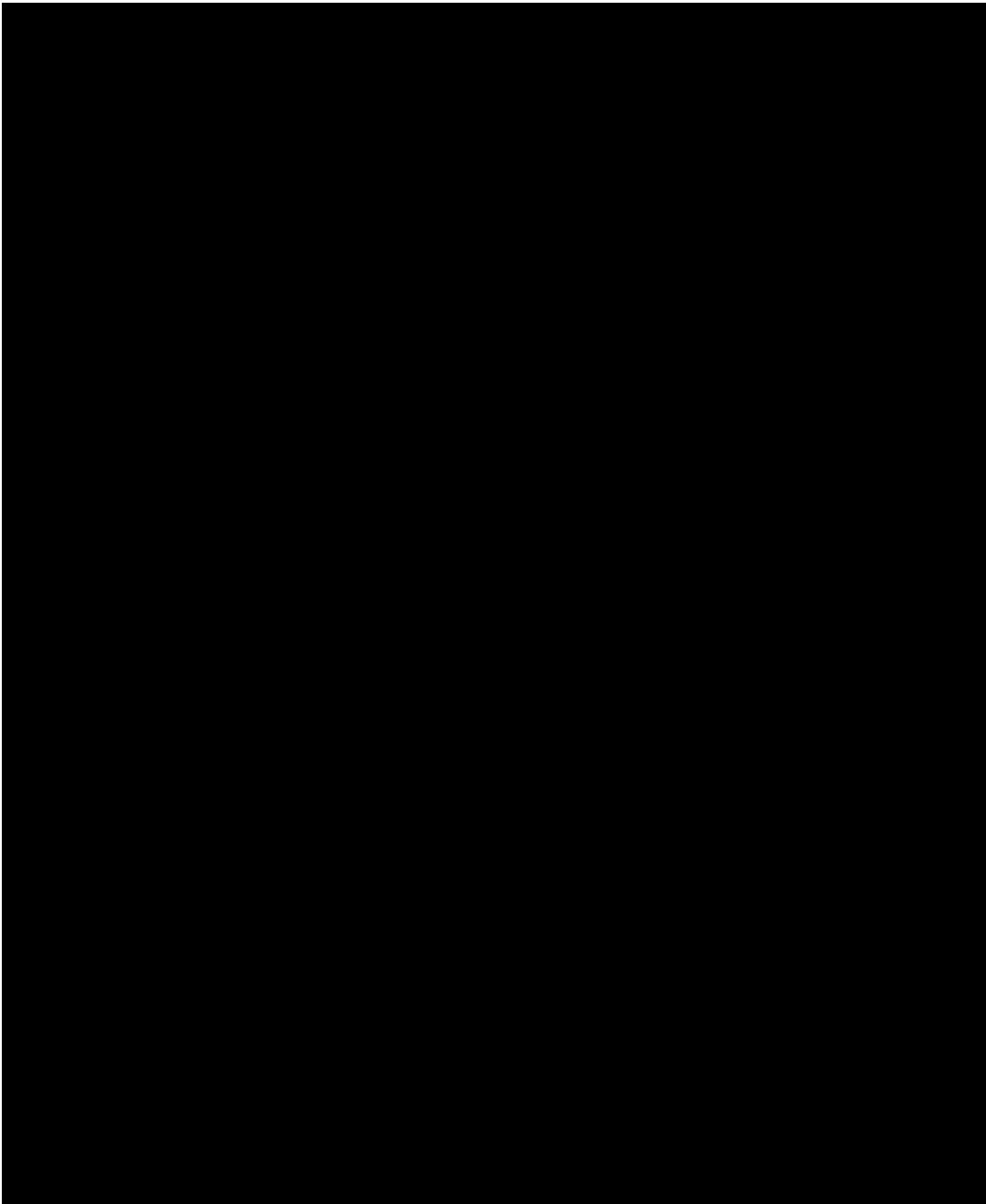
BMR value should be justified. There is much debate over the appropriate   despite that fact that it was tested to cytotoxic concentrations
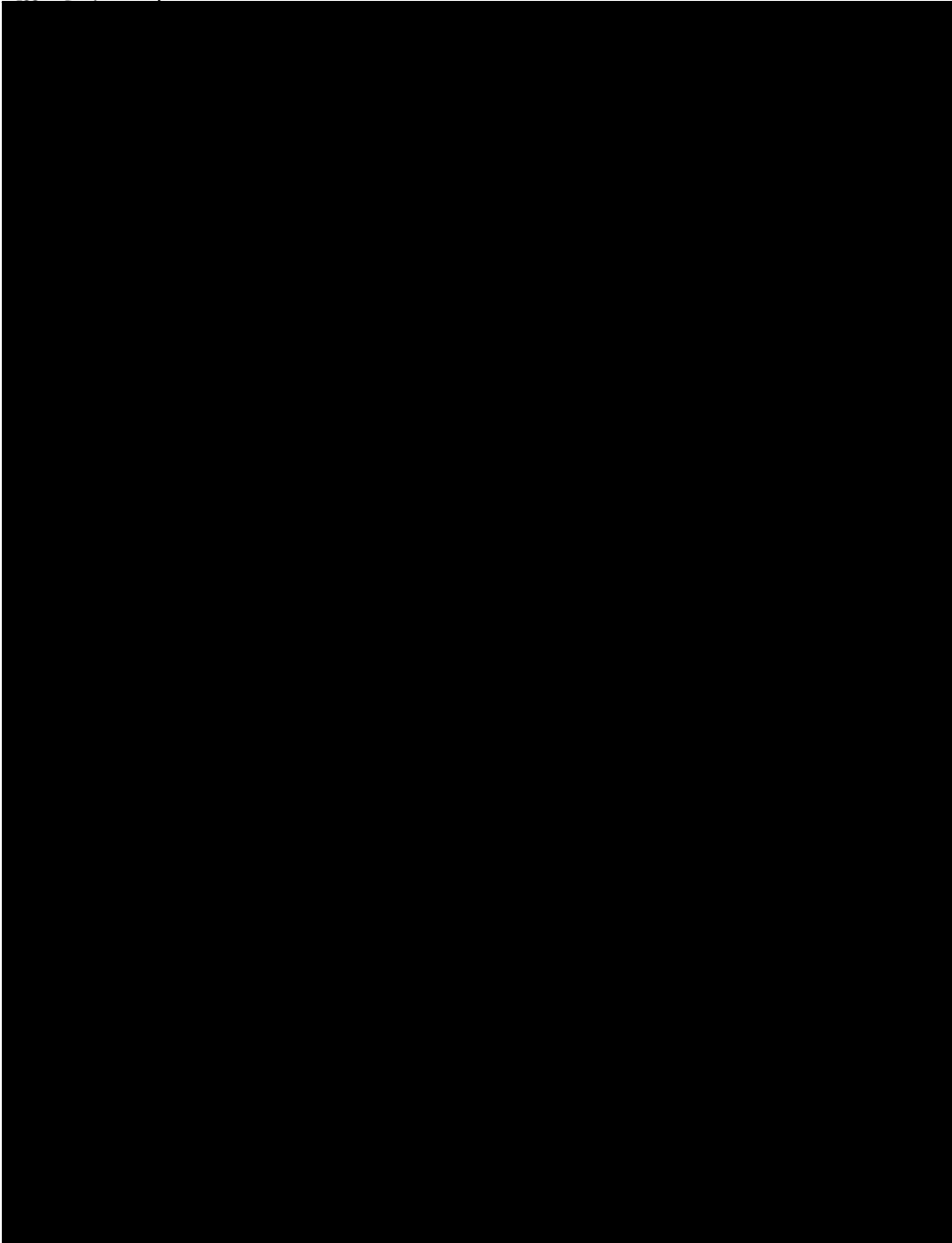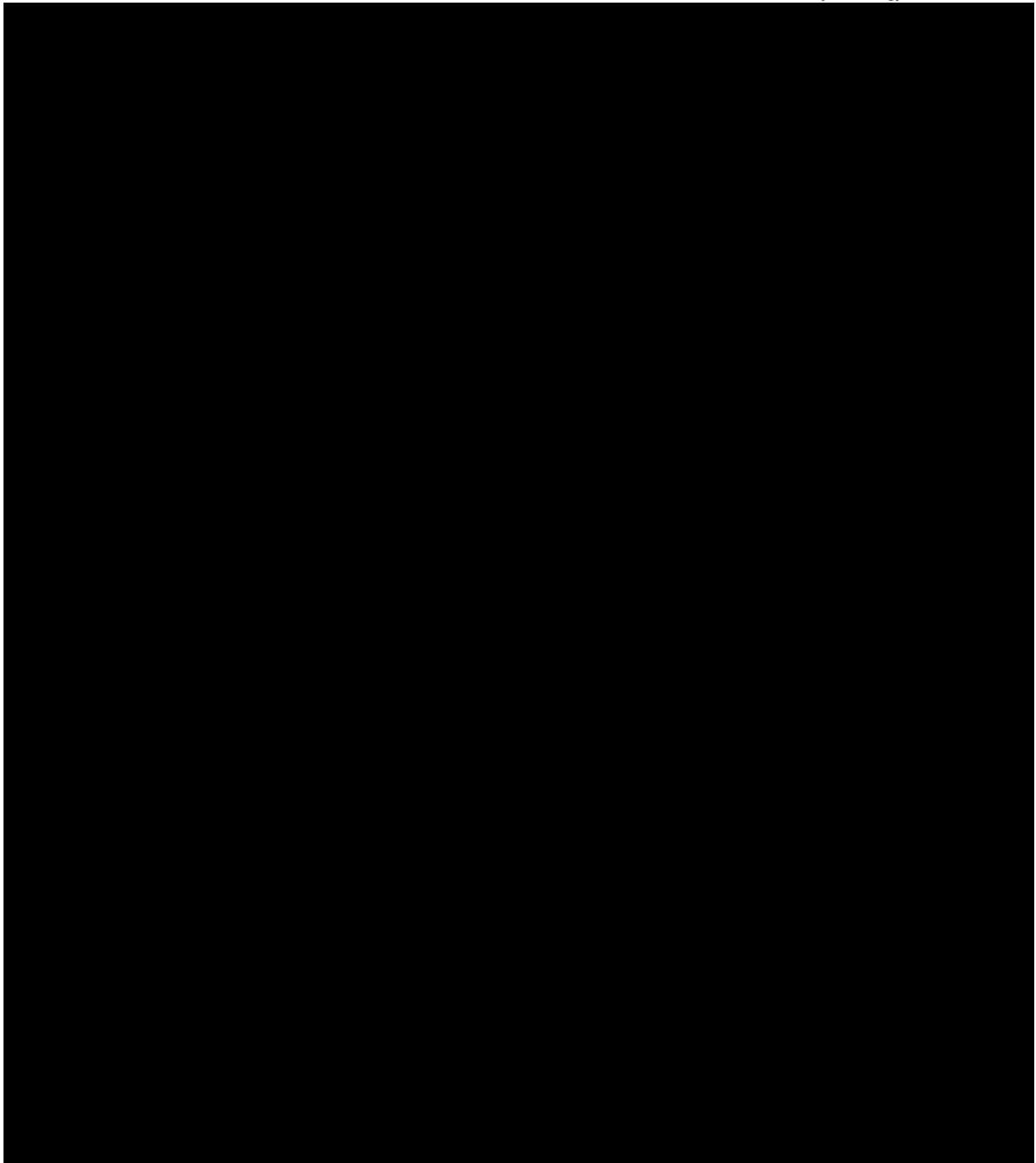
524   Dertinger et al.
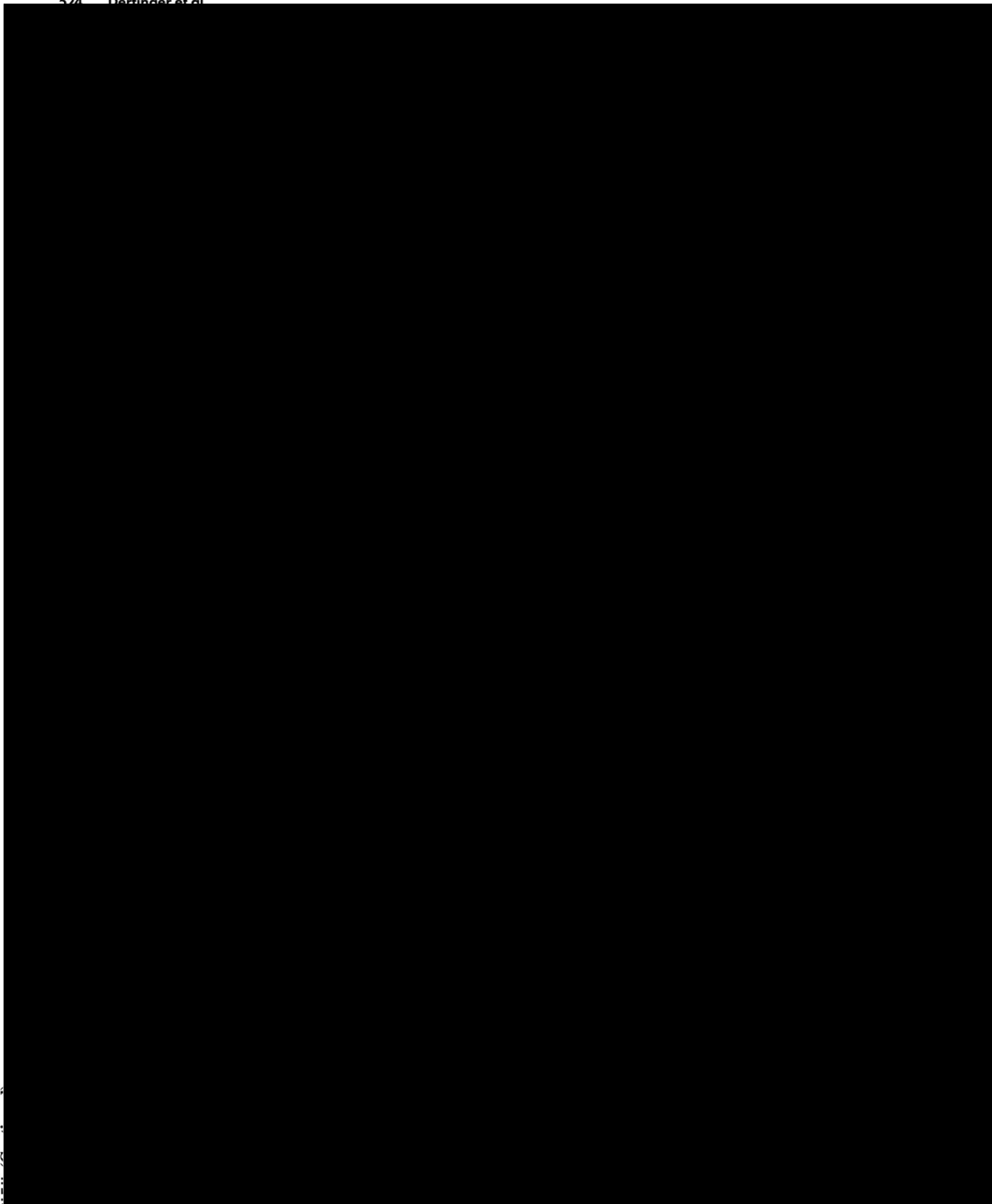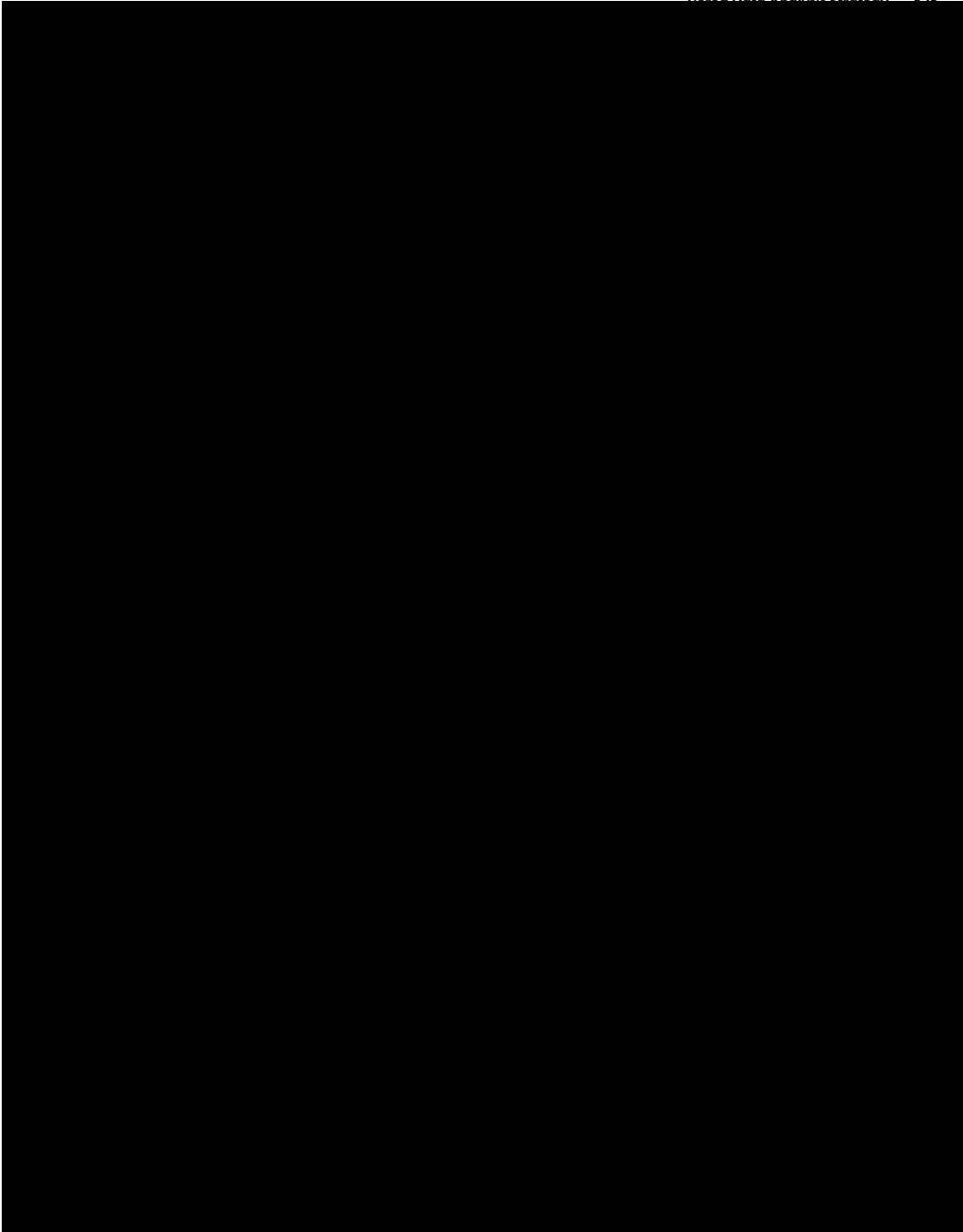
cals, work with compounds that were not used to devise the

to denote clades with chemicals that are known to exhibit these activities. The bottommost graph shows the horizontal distances between join points.
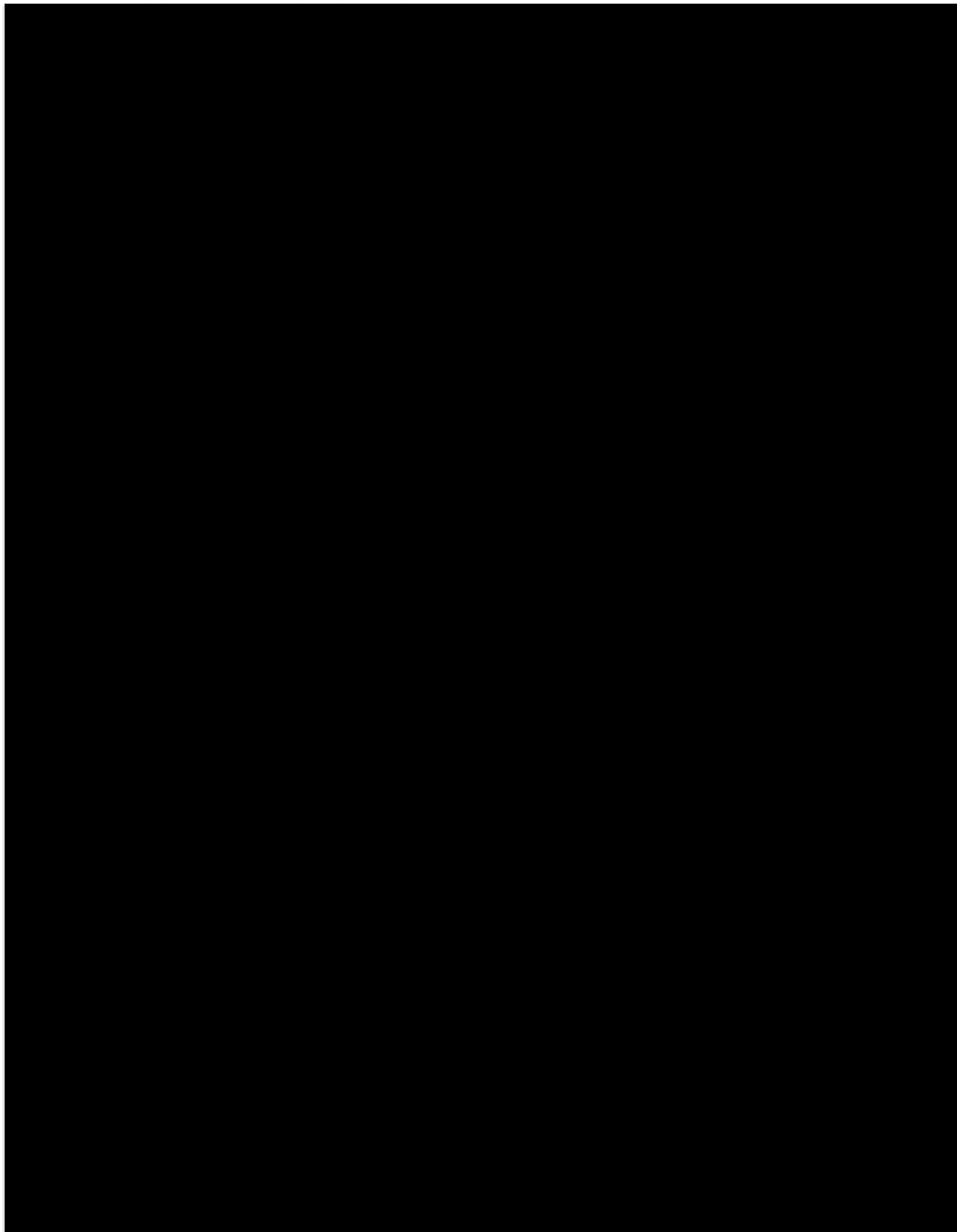
despite the fact that analyses included concentrations that induced up to 63.8% cytotoxicity. Compound 14n was also

A set of 21 *a priori* aneugens and mixed MoA chemicals that were identified as such in Tier 1 ML analyses were

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy 527

TABLE III. Test Set Chemicals N = 40
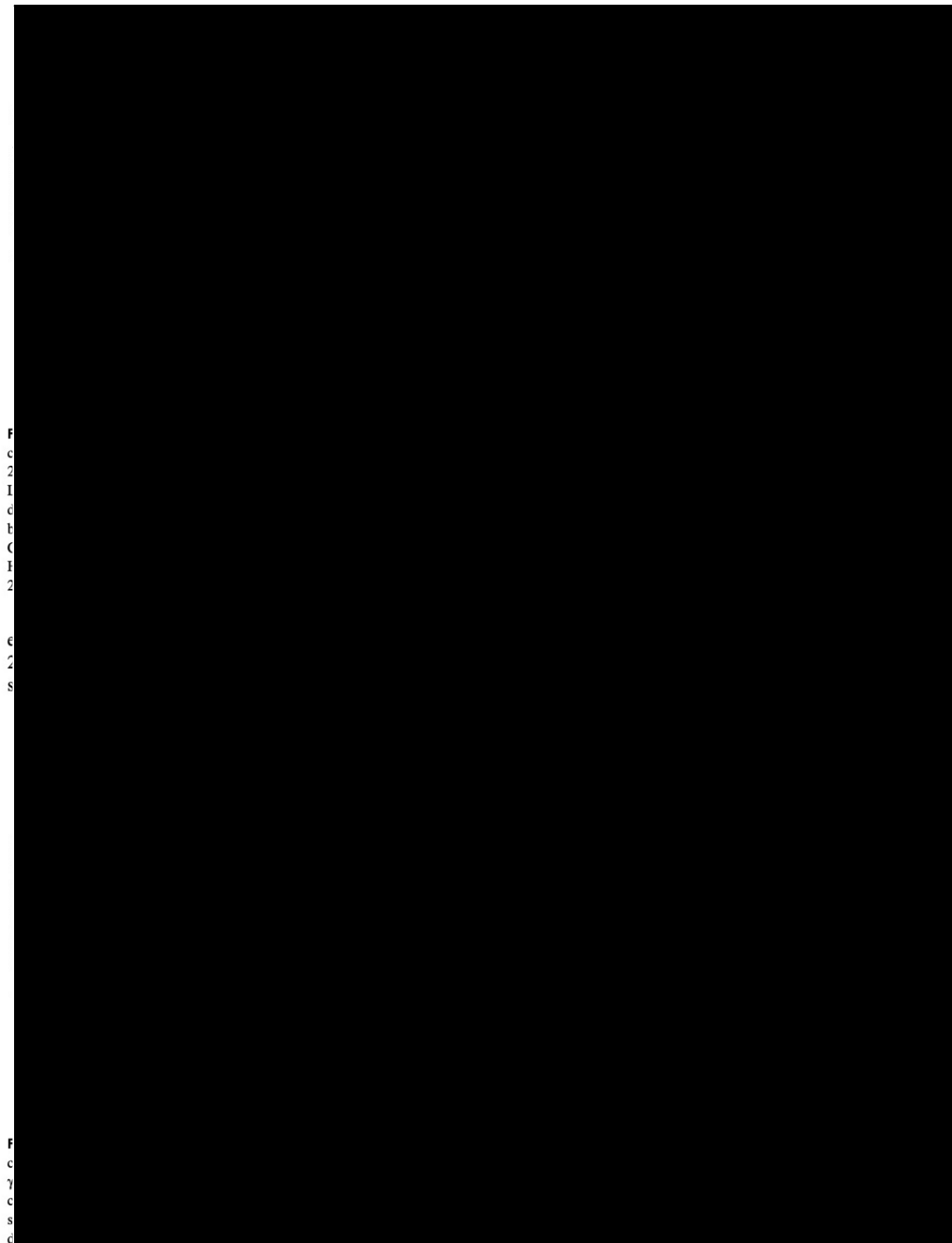
BMD50 CIs for *in vitro* MN vs. BMD50 24 hr p53 responses in TK6 cells,   (BMD, Benchmark Dose; CI, confidence interval; MN, micronucleus).

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy **529**
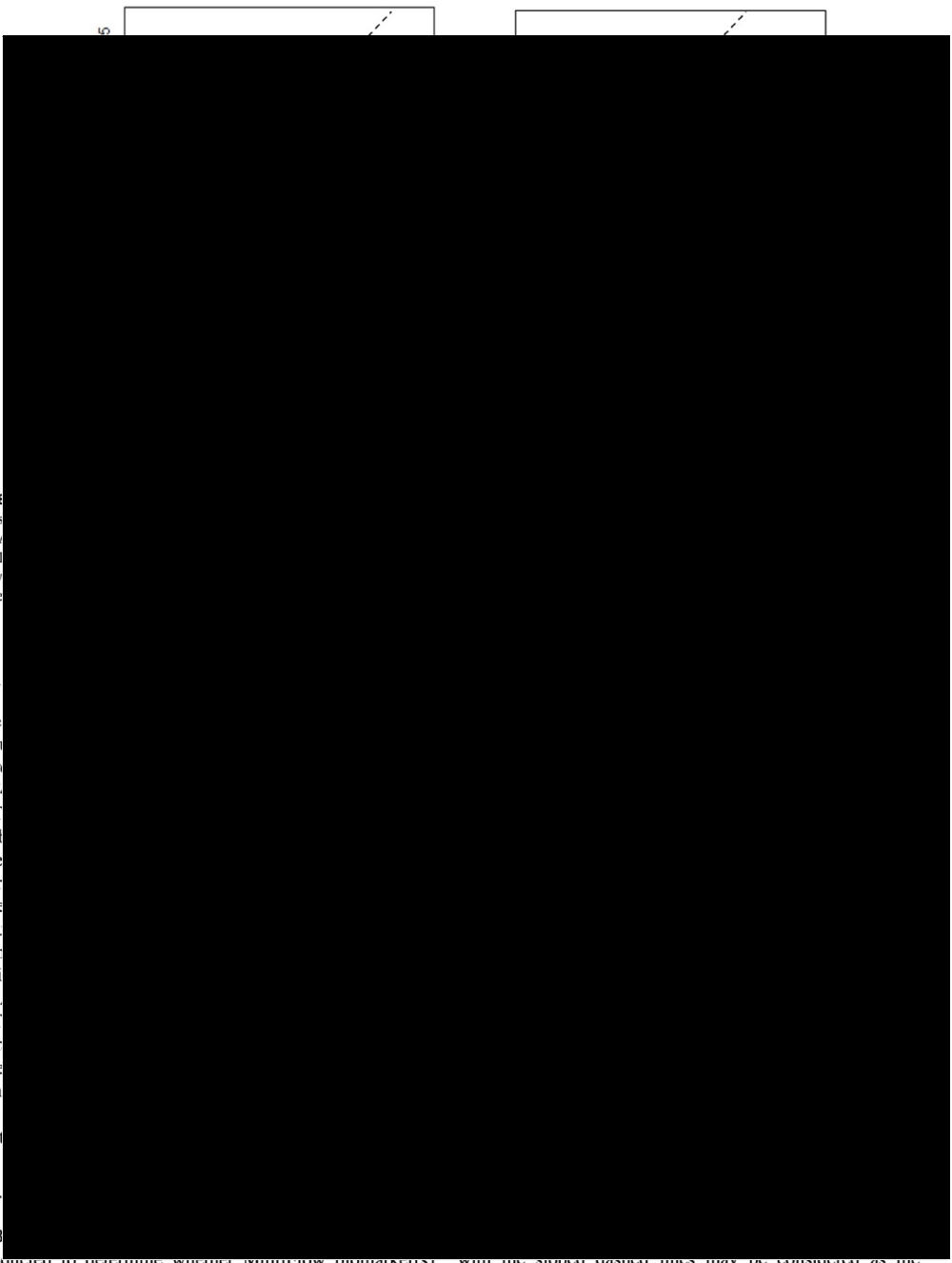
Fig. 8
cross
γH2A
Log10
draw
repre

17q
and
The
inclu
were
4474
T
such
vise
mar
dive
arou
ider
inhi
DN.
T
hier
pote
tion
the
vs. t

Tier

B
conducted to determine whether Multiflow biomarker(s)   with the sloped dashed lines may be considered as the

530   Dertinger et al.

respective upper and lower bounds of the uncertainty range   capable of highlighting biomarker response patterns that

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy 531

each contributed to the statistical analyses strategy.
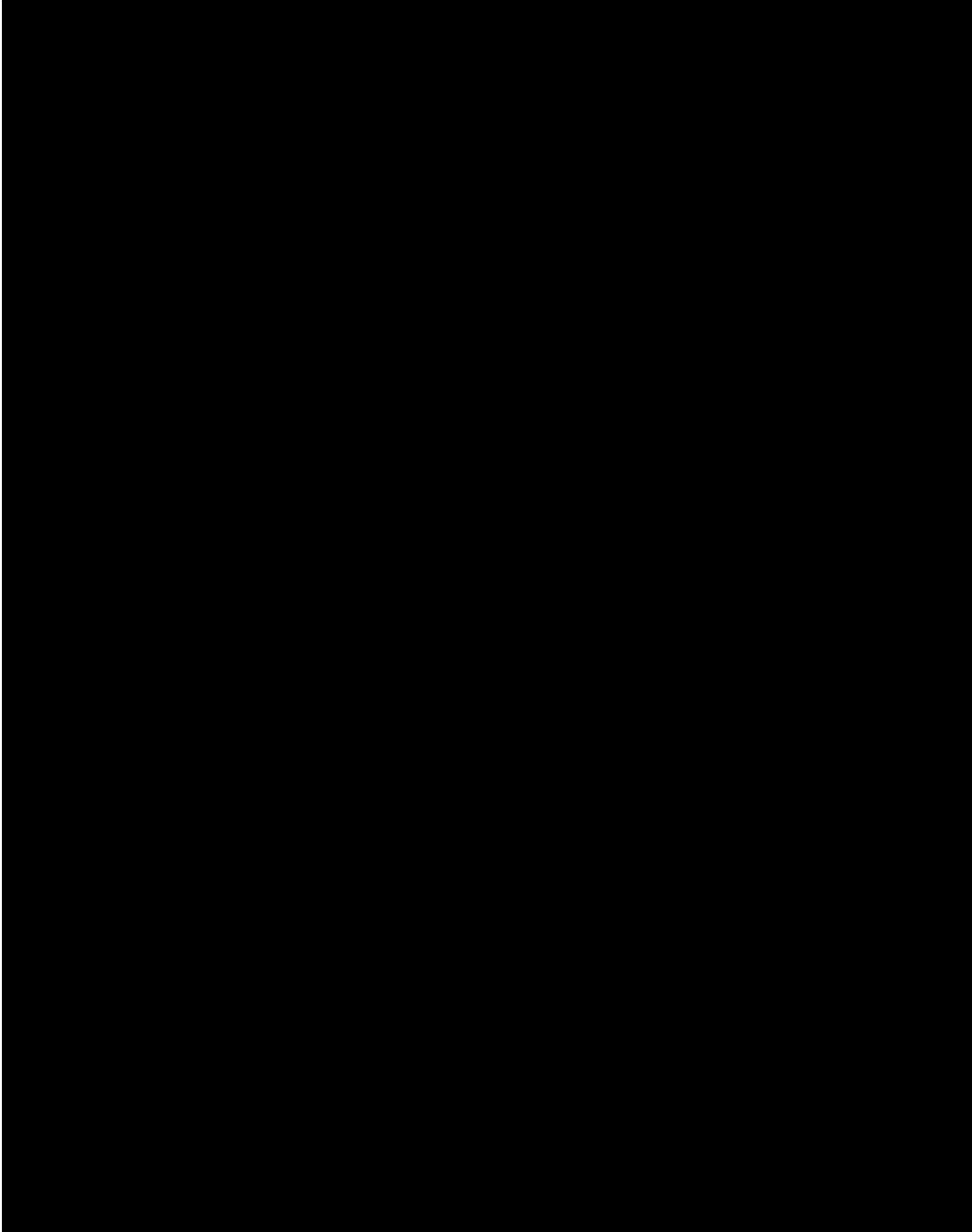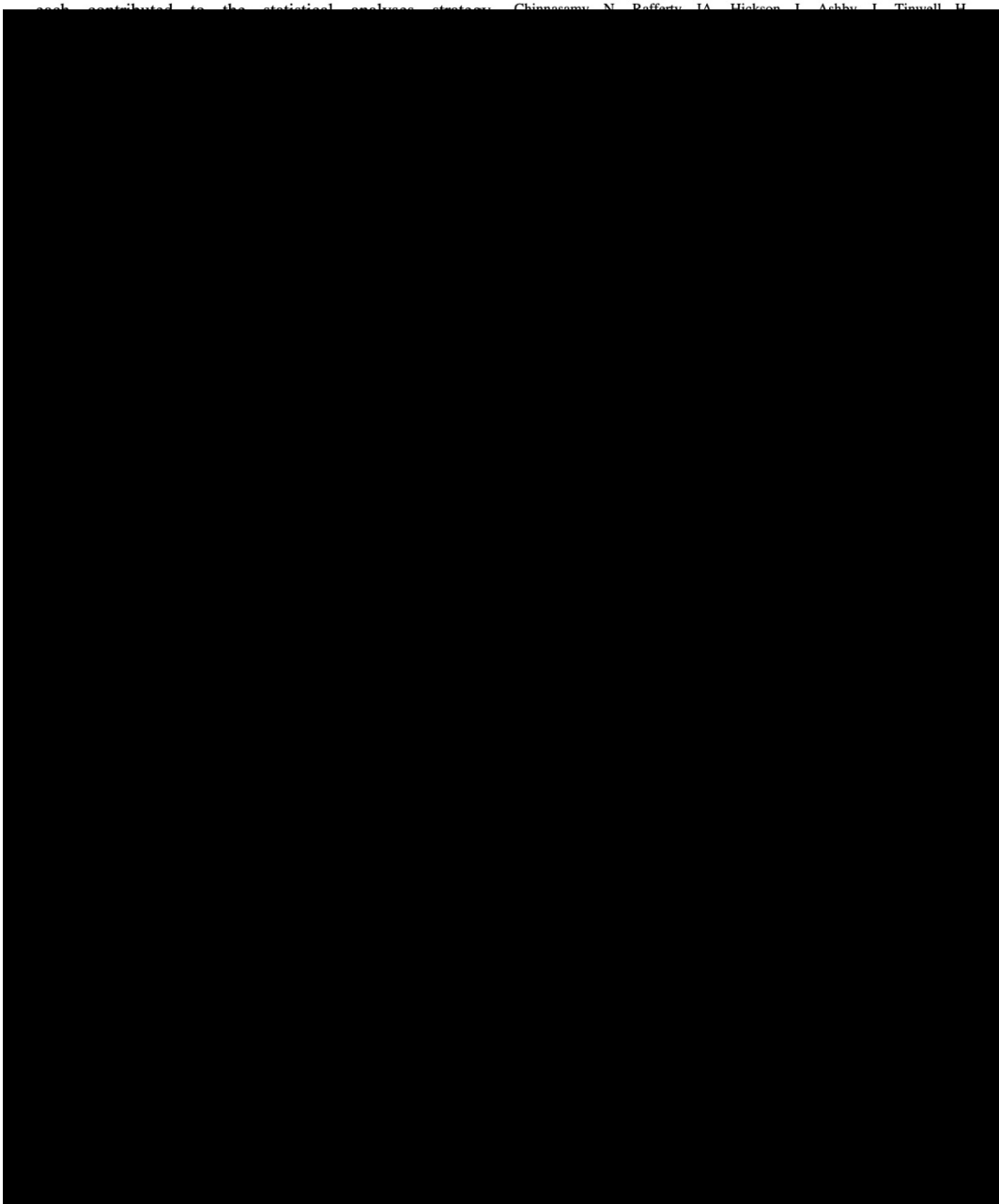
2015. Histone markers identify the mode of action for compounds positive in the TK6 micronucleus assay. Mutat Res 777:7–16.

Chinnasamy N, Rafferty JA, Hickson I, Ashby J, Tinwell H,

genotoxicity, reproductive toxicity, and carcinogenicity of ethyl methanesulfonate. Toxicol Lett 190:254–265.

Gollapudi BB, Johnson GE, Hernández LG, et al. 2013. Quantitative
    approaches for assessing dose-response relationships in genetic

Kimura A, Miyata A, Honma M. 2013. A combination of in vitro comet
    assay and micronucleus test using human lymphoblastoid TK6
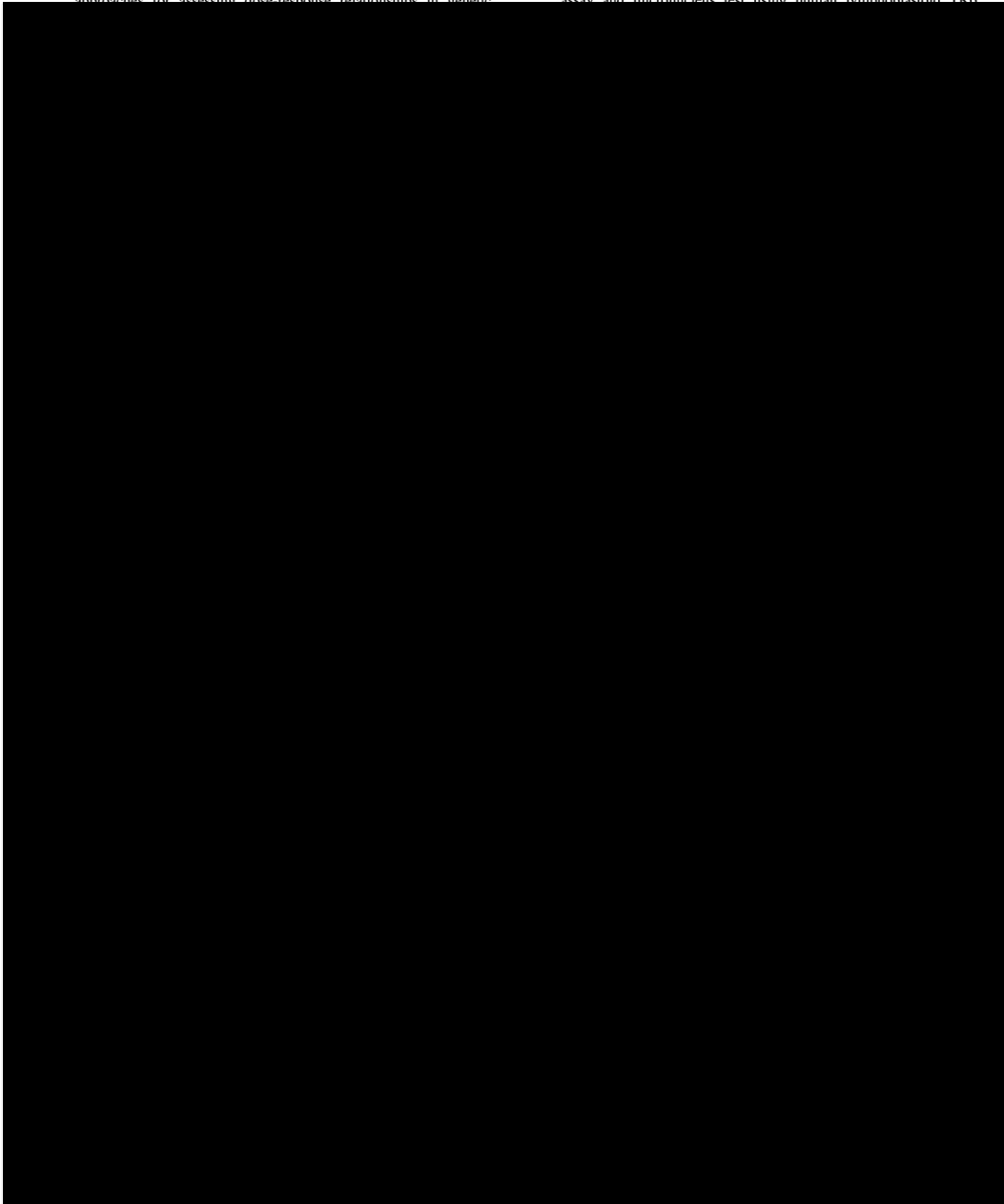
lated histones H2AX and H3 quantification in different cell lines
for genotoxicity screening. Arch Toxicol 90:1983–1995.

2012.   Available   at: http://www.accessdata.fda.gov/drugsatfda_
docs/label/2012/019643s085lbl.pdf. Accessed October 12, 2018.

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Tiered Data Analysis Strategy 533

Moon JL, Kim SY, Shin SW, Park J-W. 2012. Regulation of brefeldin A-induced ER stress and apoptosis by mitochondrial NADP+ -dependent isocitrate dehydrogenase. Biochem Biophys Res Com-

Slob W, Setzer RW. 2014. Shape and steepness of toxicological dose–response relationships of continuous endpoints. Critical Reviews in Toxicology 44:270–297.
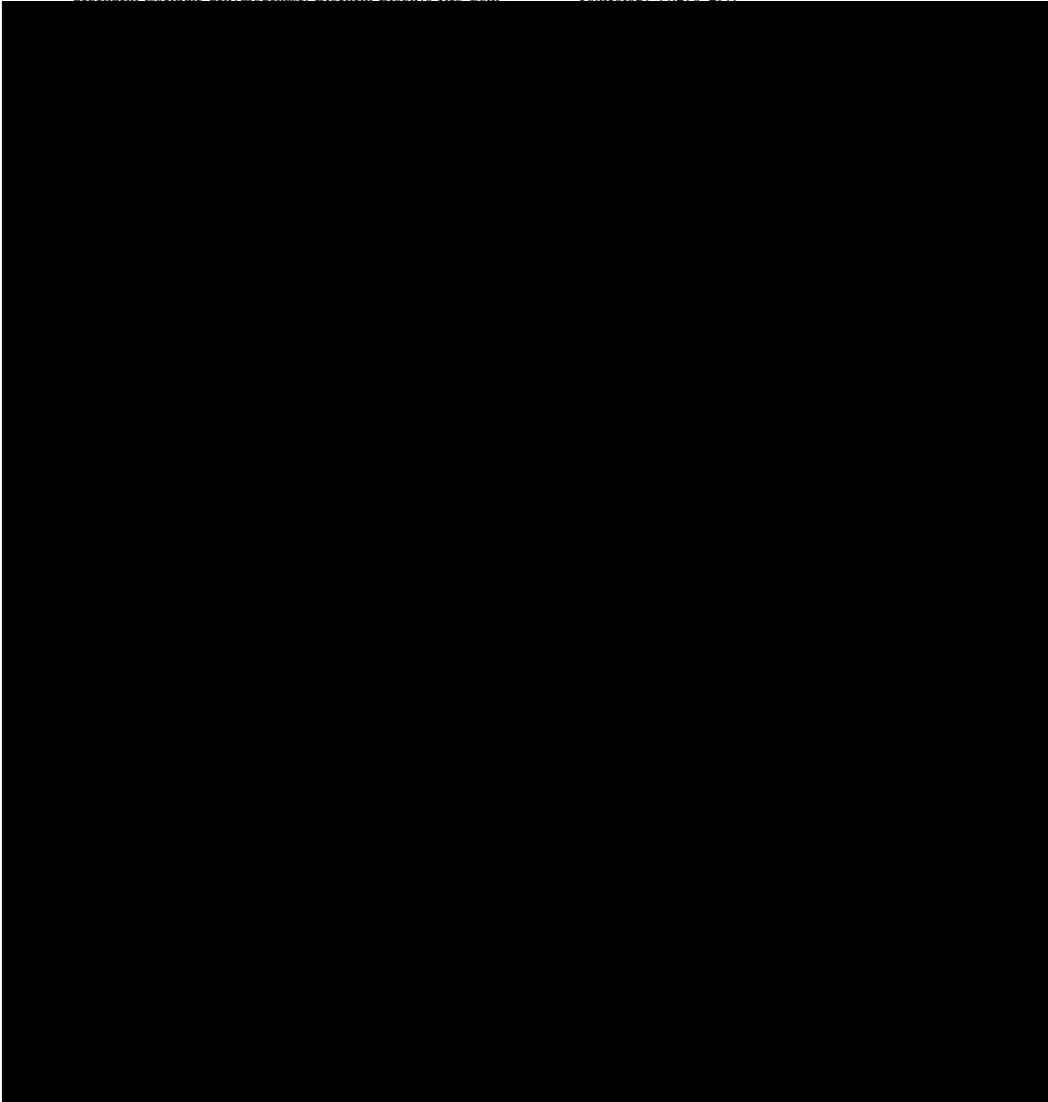
induces aneuploidy as well as polyploidy in cultured human lymphocytes. Mutagenesis 14:51–56.

Slob W. 2016. A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. Crit Rev Toxicol 47:342–351.

www.accessdata.fda.gov/drugsatfda_docs/label/2002/20412S017.pdf. Accessed October 12, 2018.

Appendix 2. Published Literature (Wheeldon et al., 2021)

RESEARCH ARTICLE

Environmental and Molecular Mutagenesis · Environmental Mutagenesis and Genomics Society · **WILEY**

The use of benchmark dose uncertainty measurements for robust comparative potency analyses

Ryan P. Wheeldon[1] [ID] | Stephen D. Dertinger[2] | Steven M. Bryce[2] |
Jeffrey C. Bemis[2] | George E. Johnson[1] [ID]

[1]Institute of Life Science, Swansea University Medical School, Swansea University, Swansea, UK

[2]Litron Laboratories, Rochester, New York

Correspondence
Ryan P. Wheeldon, Swansea University Medical School, Swansea University, Swansea SA2 8PP, UK.
Email: r.wheeldon.702067@swansea.ac.uk

Funding information
National Institute of Environmental Health Sciences, Grant/Award Number: R44ES029014

Accepted by: P. White

Abstract

The Benchmark Dose (BMD) method is the favored approach for quantitative dose–response analysis where uncertainty measurements are delineated between the upper (BMDU) and lower (BMDL) confidence bounds, or confidence intervals (CIs). Little has been published on the accurate interpretation of uncertainty measurements for potency comparative analyses between different test conditions. We highlight this by revisiting a previously published comparative in vitro genotoxicity dataset for human lymphoblastoid TK6 cells that were exposed to each of 10 clastogens in the presence and absence (+/−) of low concentration (0.25%) S9, and scored for p53, γH2AX and Relative Nuclei Count (RNC) responses at two timepoints (Tian et al., 2020). The researchers utilized BMD point estimates in potency comparative analysis between S9 treatment conditions. Here we highlight a shortcoming that the use of BMD point estimates can mischaracterize potency differences between systems. We reanalyzed the dose responses by BMD modeling using PROAST v69.1. We used the resulting BMDL and BMDU metrics to calculate "S9 potency ratio confidence intervals" that compare the relative potency of compounds +/− S9 as more statistically robust metrics for comparative potency measurements compared to BMD point estimate ratios. We performed unsupervised hierarchical clustering that identified four S9-dependent groupings: high and low-level potentiation, no effect, and diminution. This work demonstrates the importance of using BMD uncertainty measurements in potency comparative analyses between test conditions. Irrespective of the source of the data, we propose a stepwise approach when performing BMD modeling in comparative potency analyses between test conditions.

KEYWORDS
BMD, comparative, DNA damage, potency, uncertainty

1 | INTRODUCTION

An appreciation of dose–response analysis of genotoxicity data has accompanied a shift from a hazard identification testing approach, toward quantitative assessment of genotoxicity for risk assessment purposes (Macgregor et al., 2015; Dearfield et al., 2017). Under the auspices of the Quantitative Analysis Workgroup (QAW) of the Health and Environmental Sciences Institute Genetic Toxicology Technical Committee (HESI GTTC), different statistical methods for assessing dose–response relationships in genetic toxicology studies have been evaluated (Gollapudi et al., 2013). An overall conclusion was that the benchmark dose (BMD) approach for analyzing dose–response data derived from genotoxicity

studies exhibits the most favorable characteristics. The BMD for continuous data is defined as the dose that results in a predetermined change, typically ranging between 1 and 10% in the response rate of an adverse effect relative to existing background incidence (Macgregor et al., 2015). Unlike other dose–response analysis methods such as the No Observed Genotoxic Effect Level (NOGEL) approach, the BMD method is advantageous since it is not restricted to the study concentration/dose selection. The BMD method evaluates the entire range of concentrations/doses within the dataset while providing measures of uncertainty such as confidence limits (Crump, 1984; Slob, 2002). The uncertainty in the estimation of the BMD is defined as the range delineated between the upper (BMDU) and the lower (BMDL) confidence bounds.

BMDs and their associated confidence intervals (CIs) can be interpreted in a manner that conveys the potency of the studied test article. For example, CIs have been plotted to illustrate compound potency ranking from in vivo carcinogenicity and genotoxicity studies, as well as providing empirical comparisons across endpoints. Additionally, combined analysis can be performed for multiple dose–response datasets for a shared endpoint differentiated by one or more covariates (Slob and Setzer, 2014). This is known as the BMD combined covariate approach. A prominent study by Wills and colleagues (Wills et al., 2016) applied the BMD combined covariate approach to in vitro genotoxicity data and demonstrated that the precision of the BMD increases when compound, or another condition that pertains to mode of action (MoA), serves as the covariate. The precision of the BMD is defined by the ratio of the BMDL to the BMDU and has potential implications for regulatory decision making when used to define a Point of Departure (PoD) for extrapolation to a human exposure limit (White et al., 2020).

Wills et al. (2016) states that as CIs represent the range in which the true BMD lies, potency differences are only statistically defendable when there is no apparent overlap between intervals. Thus, evaluation of the entire BMD CI (BMDL and BMDU) is imperative to drawing potency conclusions from BMD analysis of dose–response datasets. Researchers have used BMD CIs of chemical classes of interest to plot compound genotoxic and/or carcinogenic potency in rank order (Hernández et al., 2011; Soeteman-Hernández et al., 2015a; Soeteman-Hernández et al., 2015b; Wills et al., 2016, 2017). Some prominent examples where BMD CIs were used in comparative genotoxicity potency analyses include studies by Allemang et al. (2018) and Wheeldon et al. (2020). Allemang et al. (2018) performed BMD analyses to evaluate the relative genotoxic potency of 15 pyrrolizidine alkaloids (PAs) via in vitro micronuclei formation in HepaRG cells. Wheeldon et al. (2020) performed BMD analyses to evaluate the comparative genotoxic potency of 8 Topoisomerase II poisons studied in human lymphoblastoid TK6 cells using the MultiFlow DNA damage response assay. Both publications identified the utility of BMD CIs in compound comparative genotoxicity potency analyses to support read across and MoA determination for a limited number of compounds of interest. In all of the aforementioned studies, compound potency comparisons within and between genotoxicity endpoints were always performed by scrutinizing the shape and steepness of the underlying dose–response curve, and by graphically representing the BMD CIs. Potency comparisons and measures of correlation were consistently

evaluated by CIs spanning orders of magnitude versus deriving numerical values for comparison purposes.

Herein we consider a previously published in vitro genotoxicity dataset where the authors performed BMD analyses and drew conclusions without considering the uncertainty associated with the BMD measurements. Tian et al. (2020) investigated the use of phenobarbital/β-napthoflavone-induced rat liver S9 at maximal non-cytotoxic concentration (0.25% vol/vol final) in a flow cytometry based multiplexed DNA damage response assay. The laboratory was able to maintain the S9 enzyme/co-factor mix with cells and test compound for the entire exposure period of 24 hr in the TK6-based assay. The investigators focused on 15 chemicals: 8 of which are clastogens that are known to require metabolic activation to maximize formation of DNA-reactive metabolites; 5 are cytotoxicants; and 2 are direct acting clastogens that do not require metabolic activation. In addition to determining compound MoA through the use of biomarker responses, the authors applied the BMD approach with the aim of calculating a numerical "S9 potentiation ratio" value, which served as a comparison metric obtained by division of the BMD value in the absence of S9 by the BMD value in the presence of S9. Upon further consideration of this approach, we believe that there are significant shortcomings in the use of solely a BMD point estimate in their comparative potency analysis. Using the same logic highlighted by Wills et al. (2016); it is irreconcilable to rely upon a BMD point estimate to robustly compare potencies, since the BMD CI represents the range in which the true BMD lies. To our knowledge, there are no published reports that critique the use of a BMD point estimate, and how this differs from use of BMD CIs in comparative analysis between experimental conditions. Herein, the Tian et al. (2020) data is re-analyzed to further inform use of the same in vitro genotoxicity/low concentration S9 dataset. Primarily, we stress that the use of a BMD point estimate value does not provide an accurate representation of the likely potency range of the test compound. Said another way, the BMD point estimates and associated "S9 potentiation ratio" comparison metrics did not convey information about the uncertainty of the measurements that are consistent with the BMD uncertainty measurement approach that is advocated in the scientific literature.

This current report focuses on reanalysis and augmentation of the previously published Tian et al. (2020) dataset. Specifically, we convey the BMD uncertainty measurements that relate to the relative genotoxic potency of the 10 clastogenic compounds. To this end, we calculate "S9 potency ratio CIs" using the BMD uncertainty measurements (BMDL and BMDU) between S9 exposure conditions (the presence and absence of S9) and utilize said ratios to derive robust potency conclusions as a follow up to the Tian et al. (2020) work. Readers are encouraged to refer to the original Tian and colleagues' article for further context (Tian et al., 2020).

2 | MATERIALS AND METHODS

2.1 | In vitro genotoxicity dataset

The data were derived from a previously published article in which 15 compounds were studied using the in vitro MultiFlow® DNA

WHEELDON ET AL.

Environmental and
Molecular Mutagenesis

Environmental
Mutagenesis and
Genomics Society

WILEY | 205

Damage Assay in the presence and absence of low dose (0.25% vol/vol) S9 (Tian et al. 2020). The in vitro MultiFlow DNA Damage Assay multiplexes several biomarkers that are responsive to diverse forms of DNA damage into a single flow cytometric analysis. The multiplexed biomarkers include: (a) phosphorylation of H2AX at serine 139 (γH2AX) for the detection of DNA double strand breaks, (b) phosphorylation of histone H3 at serine 10 (p-H3) to identify mitotic cells, (c) nuclear p53 content as an indicator of p53 activation, (d) frequency of 8n+ cells to monitor polyploidization, and (e) relative nuclei counts (RNC) to provide information about treatment related cytotoxicity (Bryce et al., 2016). A detailed description of the assay falls out of scope of this article; thus, interested readers are encouraged to refer to several publications that describe the MultiFlow assay (Bryce et al., 2016; Bryce et al., 2017; Bryce et al., 2018; Dertinger et al., 2019). As previously mentioned, the Tian et al. (2020) data included 8 direct acting clastogens that require metabolic activation, 5 cytotoxicants and 2 direct acting clastogens that do not require metabolic activation. In this reanalysis, we focused on the 10 clastogens, since this is where Tian et al. (2020) suggested the greatest differences in potency between S9 exist. The raw data were provided by Litron Laboratories and included 4 and 24 hr p53, γH2AX, and 24 hr RNC responses for the 10 clastogens: 2-acetylaminofluorene, 2-aminoanthracene, 7,12-dimethylbenzanthracene, benzo[a]pyrene, cyclophosphamide, dibenzo[a,l]pyrene (also known as dibenzo[def,p]chrysene), diethylnitrosamine, mitomycin C, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), and resorcinol.

2.2 | BMD analyses

Tian et al. (2020), performed BMD analysis on individual compounds with the S9 exposure (with or without [+/−] 0.25% vol/vol S9) condition serving as the covariate. Hence, individual dose–response curves on a compound basis already existed for this dataset. Here, the individual dose–response curves were scrutinized so that compounds and/or endpoints with little to no evidence of a dose–response could be disqualified from reanalysis. Hence, we excluded 4 hr p53 dose responses for diethylnitrosamine; 24 hr p53 dose–responses for diethylnitrosamine; and 24 hr γH2AX dose responses for 7,12-dimethylbenzanthracene.

PROAST version 69.1 operating in R 4.0.2 was used to analyze the continuous dose–response data (http://www.proast.nl) for the 4 hr and 24 hr γH2AX, 4 hr and 24 hr p53, and 24 hr RNC endpoints. We applied the exponential model as a sequence of nested models with increasing number of parameters (EFSA 2009; Slob and Setzer 2014; EFSA 2017). In this reanalysis, compound was selected as the covariate—differing from Tian et al. (2020) where S9 condition was the covariate—since this includes more dose–responses in a single analysis to increase the precision of the BMDs. PROAST provided the option to select model 3 or 5 from the family of models. In doing so, the number of parameters were increased to test whether the associated log likelihood is significantly increased, thus informing if the additional parameters were required for describing the dose–response (Slob, 2002; EFSA, 2009; Slob and Setzer, 2014; EFSA, 2017). Either model 3 or 5 was finally automatically selected as the

most appropriate model to describe the shape of the dose–response curve. Readers are advised to refer to Slob and Setzer (2014) for detailed information about the dose response models (including models 3 and 5) and their algorithms.

Consistent with the approach utilized by Tian et al. (2020), an arbitrary critical effect size (CES) of 0.3 was selected for this reanalysis. A CES of 0.3 represents a 30% change in response compared to the concomitant control. CES −0.3 for the RNC endpoint represents a 30% decrease in response for this endpoint analyses. There is a lack of consensus on the appropriate choice of CES reported in the literature for in vivo endpoints (White et al., 2020), thus a lengthy discussion on CES falls outside the scope of this report. In any event, we justify the use of CES 0.3 for these in vitro endpoints since the resulting BMDs do not lie in the extremities of the dose response curves where the associated uncertainties could be relatively high.

The value of the BMD is a point estimate with an associated level of precision (the ratio between the BMDL and BMDU), thus 90% BMD CIs were obtained for each compound and endpoint combination in the presence and absence of S9 and were graphically plotted.

2.3 | Unsupervised clustering

Lower and upper values of the "S9 potency ratio CIs" (algorithm described in the results and discussion section) were evaluated using JMP software's unsupervised clustering platform (JMP, v12.0.1). Lower and upper "S9 potency ratio CI" values associated with the following 5 biomarkers were used as variables: 4 hr p53, 24 hr p53, 4 hr γH2AX, 24 hr γH2AX, and 24 hr RNC. The analysis options were set as follows: clustering method = hierarchical; method for calculating distances between clusters = "Ward"; data as usual = "Standardize Robustly"; data visualization = "Dendrogram," with "two-way clustering."

3 | RESULTS AND DISCUSSION

3.1 | BMD confidence intervals

Values of the BMD, BMDL, and BMDU were estimated and collated for all compounds (+/− S9) and endpoints that were included in the BMD analysis. The BMD CIs were plotted for each compound with +/−S9 CIs side-by-side (Figures 1–10). The BMD point estimates are included in the comparative potency plots to aid in graphical representation of the BMD point estimate relative to the corresponding BMDL and BMDU. This is important since some individuals misinterpret the BMD to be at the midpoint (or geometric mean) between the BMDL and BMDU. This is a misconception as evident by the differing lengths of the CIs either side of each BMD point estimate displayed in Figures 1–10.

CIs that span a maximum of approximately 1–2 Log units (i.e., 1–2 orders of magnitude) are considered good quality and consistent
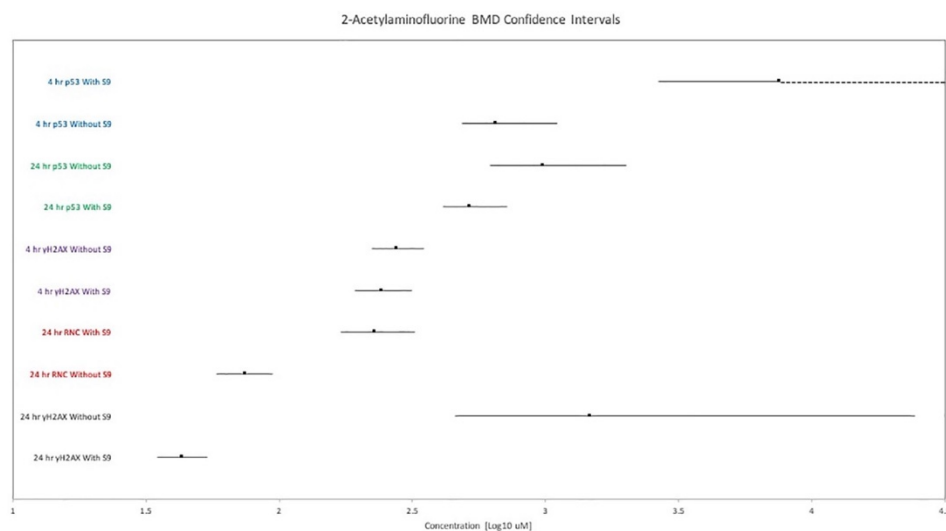
FIGURE 1 BMD confidence intervals for 2-acetylaminofluorene. 4 hr p53 endpoint with S9 infinite BMDU indicated with dashed positive direction lines. BMD point estimates are displayed as data points
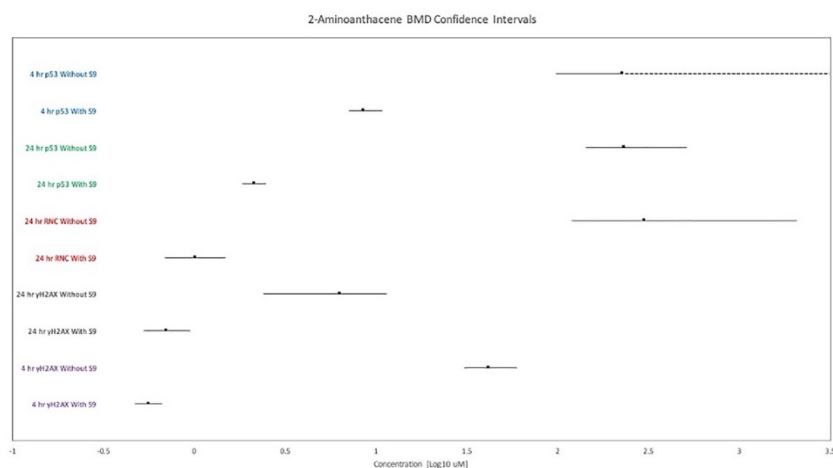
FIGURE 2 BMD confidence intervals for 2-aminoanthracene. 4 hr p53 endpoint without S9 infinite BMDU indicated with dashed positive direction lines. BMD point estimates are displayed as data points

with other BMD CI spans that are reported in the literature for other in vitro genotoxicity systems (Soeteman-Hernández et al., 2015a; Bemis et al., 2016; Wills et al., 2016; Allemang et al., 2018). While most compound's BMD analysis yielded tightly bound CIs, some displayed wide CIs or dose responses with unbound upper confidence limits. Specifically, the 4 hr p53 endpoint for 2-acetylaminofluorene −S9 yielded an infinite BMDU

(BMD 3.88 Log10 μM, BMDL 3.42 Log10 μM, BMDU Infinite Log10 μM) (Figure 1). Second, both the 4 hr and 24 hr p53 (Figure 9) endpoints for PhIP −S9 yielded unbound BMDUs (BMD 4.39 Log10 μM, BMDL 3.03 Log10 μM, BMDU Infinite Log10 μM; BMD 5.20 Log10 μM, BMDL 3.22 Log10 μM, BMDU Infinite Log10 μM, respectively). The 4 hr γH2AX endpoint for 7,12-dimethylbenzanthracene returned a zero BMDL value

358
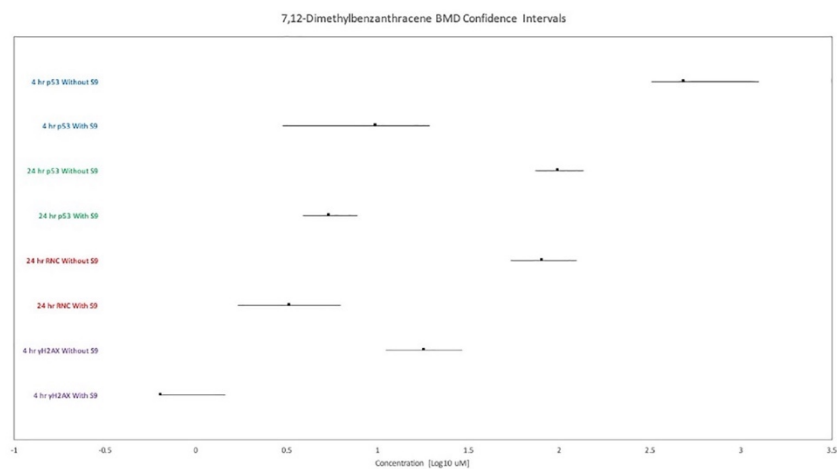
FIGURE 3 BMD confidence intervals for 7,12-dimethylbenzanthracene. 4 hr γH2AX endpoint with S9 yielded a zero value lower confidence bound. BMD point estimates are displayed as data points
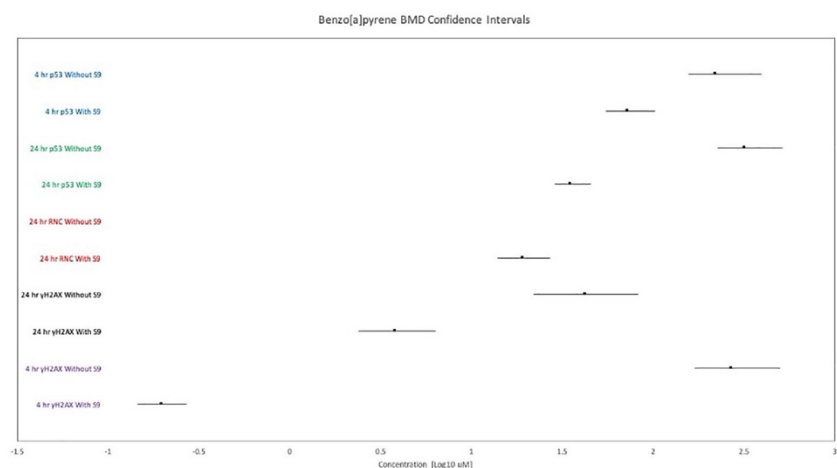
FIGURE 4 BMD confidence intervals for benzo[a]pyrene. 24 hr RNC endpoint without S9 yielded a disproportionately high BMD with infinite BMDU: indicated as a dashed line spanning the width of the plot. BMD point estimates are displayed as data points

indicating that the lower bound of the BMD is not significant from zero. We considered these limited instances where dose–response analyses yielded infinite BMDUs to have exhibited no evidence of dose–response relationships, and hence are unsuitable for drawing statistically robust conclusions from the analyses. In the single instance of a zero BMDL, the resulting BMDL to BMDU ratio is unobtainable (problems associated with division by 0), and hence justifies omission of this compound/endpoint combination from further analysis.

In addition to having unbound BMDUs, some compounds also displayed BMD estimates that are grossly disproportionate to the BMDs from other endpoints or S9 condition for the same compound. The BMD CIs for these instances were not plotted, since the relative disproportionality would make the other CIs appear comparatively small. These instances are indicated in the comparative potency plots with dashed lines spanning the entire graph range. Specifically, the 24 hr RNC endpoint −S9 for benzo[a]pyrene [BMD 12.33 Log10 μM] (Figure 4), the 24 hr RNC endpoint −S9 for dibenzo[a,l]pyrene [BMD
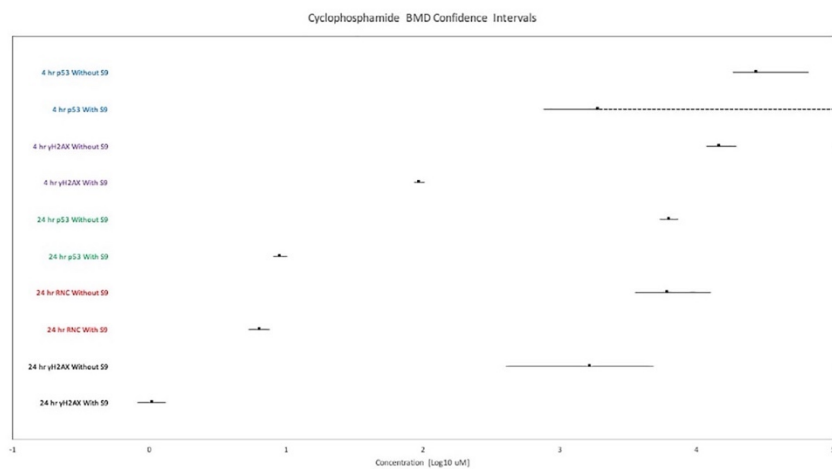
FIGURE 5 BMD confidence intervals for cyclophosphamide. Two-sided confidence intervals were obtained for all endpoint's BMD analyses. BMD point estimates are displayed as data points
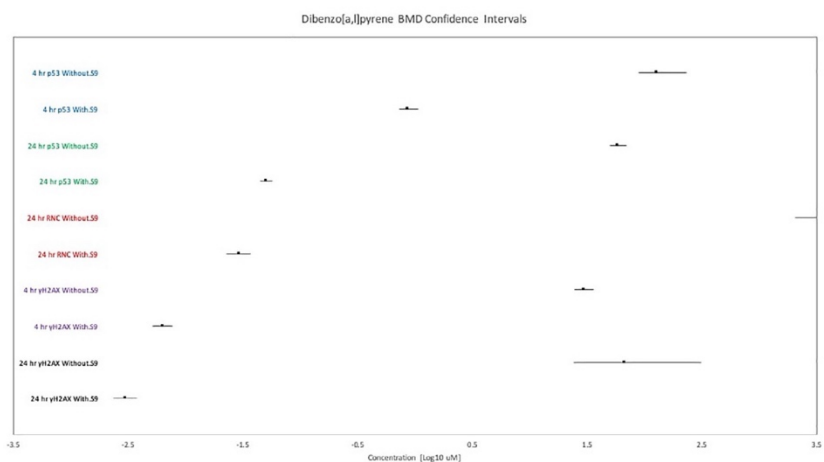
FIGURE 6 BMD confidence intervals for dibenzo[a,l]pyrene. 24 hr RNC endpoint without S9 yielded a disproportionately high BMD with infinite BMDU: BMD restricted and only displaying the BMDL. BMD point estimates are displayed as data points

11.99 Log10 μM] (Figure 6), and the 24 hr γH2AX endpoint −S9 for diethylnitrosamine [BMD 20.74 Log10 μM] (Figure 7). These endpoints also yielded infinite BMDUs, and consequently displayed no evidence of a dose response upon which one can then accurately define potency.

Visual scrutiny of each compound's BMD potency plot allows efficient visual discrimination of the potency difference of each endpoint that results from S9 exposure. Endpoints where the S9 exposure condition yielded BMD CIs that largely overlap results in potency differences that are statistically insignificant. Conversely, non-overlapping CIs are statistically significantly different and show that the S9 condition exerts an impact on the compounds potency in the same in vitro system.

For the most part, as one would expect, the tested clastogens that require metabolic activation to exert genotoxic effects show an increase in potency when exposed to the low concentration S9 system. This is evidenced by +S9 CIs that reside in more potent regions of the comparative potency plot by several orders of magnitude compared to −S9 CIs. An outlier is 2-acetylaminofluorene where all

360

WHEELDON ET AL.

Environmental and
Molecular Mutagenesis

Environmental
Mutagenesis and
Genomics Society

WILEY 209
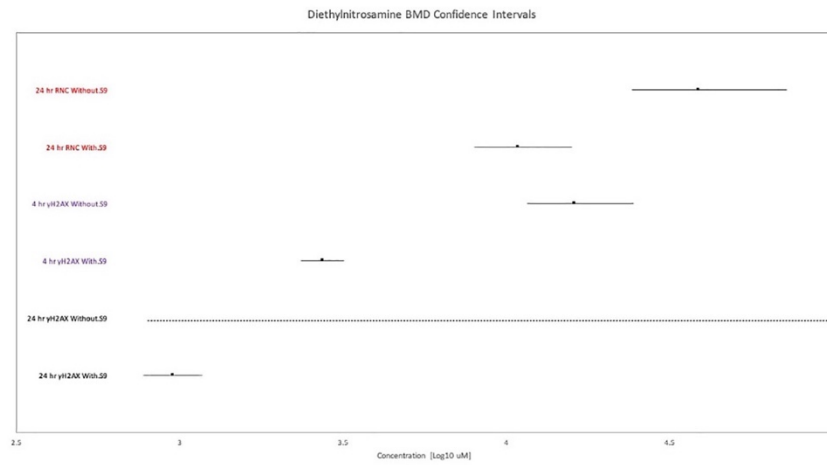
FIGURE 7   BMD confidence intervals for diethylnitrosamine. 24 hr γH2AX endpoint without S9 yielded a disproportionately high BMD with infinite BMDU: indicated as a dashed line spanning the width of the plot. BMD point estimates are displayed as data points
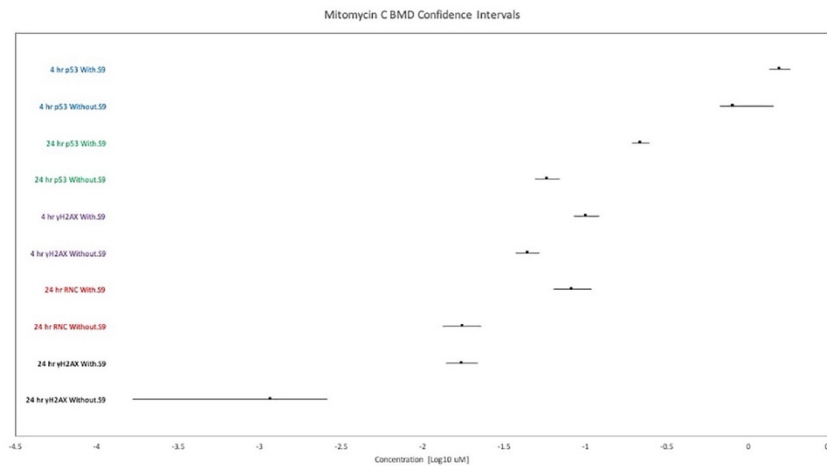
FIGURE 8   BMD confidence intervals for mitomycin C. Two-sided confidence intervals were obtained for all endpoint's BMD analyses. BMD point estimates are displayed as data points

endpoints except for 24 hr γH2AX and 24 hr RNC display CIs that reside in the same order of magnitude (Figure 1). This suggests that S9's potentiating effect was restricted to these endpoint/timepoint combinations. The comparative potency plots also show that the direct acting clastogens, mitomycin C and resorcinol, were not potentiated by S9—rather, to a certain extent, the presence of S9 slightly decreased their genotoxic potency.

3.2 | S9 potency ratio CIs

Tian et al. (2020) derived S9 potentiation ratios by comparative analysis of the BMD point estimates across the S9 exposure conditions for each compound and endpoint. There is a concern that the use of BMD point estimates can misrepresent the potency effect of S9 in this in vitro system since the BMD measure of uncertainty is disregarded.

PhIP BMD Confidence Intervals
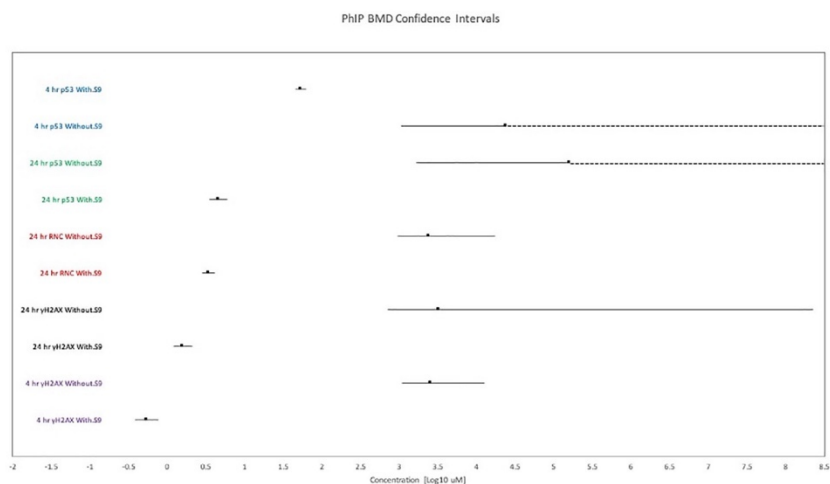
FIGURE 9 BMD confidence intervals for PhIP. 24 hr p53 endpoint without S9 and 4 hr p53 without S9 infinite BMDUs indicated with dashed positive direction lines. BMD point estimates are displayed as data points

Resorcinol BMD Confidence Intervals
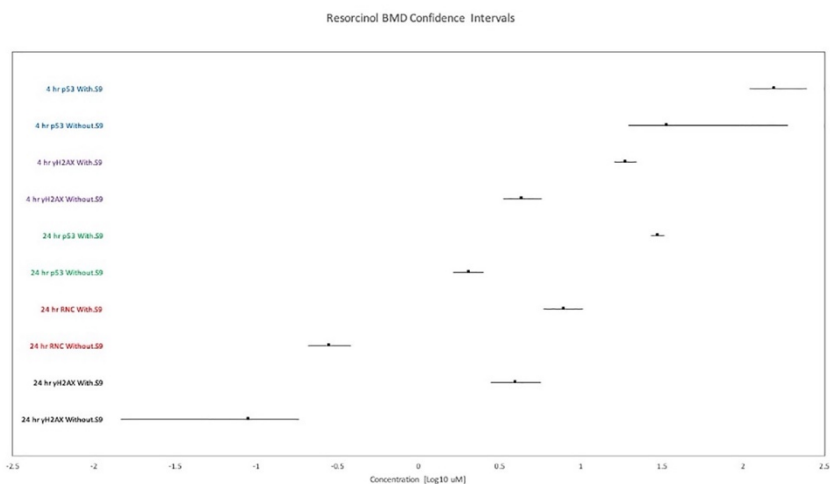
FIGURE 10 BMD confidence intervals for resorcinol. Two-sided confidence intervals were obtained for all endpoint's BMD analyses. BMD point estimates are displayed as data points

Tian et al. (2020) correctly pointed out that the BMD estimate can be beyond the top concentration used in the benchtop experiment. In these instances, the authors calculated the "S9 potentiation ratio" by restricting the BMD to the top concentration (denoted with a greater than [>] symbol in the tabulated results section of their article). However, these tight BMD restrictions led to mischaracterization of the true BMD ratio. We contend that one should only limit BMD

values for graphical display where the value is disproportionately larger than other endpoints and S9 conditions for the same compound. In the BMD analysis reported in this re-analysis, disproportionately high BMD values also coincided with unbound BMDUs.

The BMD CIs obtained in this re-analysis were used to calculate an "S9 potency ratio CI" that is derived for each compound and endpoint BMD analysis that returned a dose response by comparing the

TABLE 1 Compound/endpoint combination BMD ratios and "S9 potency ratio CIs" (original and log scale) compared with the Tian et al. (2020) "S9 potentiation ratios"

Compound	Endpoint	S9 potentiation ratio from Tian et al. (2020) (μM)	BMD-BMD ratio[a]	S9 potency ratio CI range (μM)[b]	S9 potency ratio CI range (Log10 μM)[c]
2-Acetylaminofluorine	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	NC	1.89	0	0
	4 hr γH2AX	>2.6[d]	1.14	0	0
	24 hr γH2AX	>8.9	34.23	8.53–704.02	0.93 to 2.85
	24 hr RNC	0.53	0.33	0.18–0.55	−0.74 to −0.26
2-Aminoanthracene	4 hr p53	>7	No DR	No DR	No DR
	24 hr p53	>22	108.72	57.26–278.80	1.76 to 2.45
	4 hr γH2AX	167	74.55	46.03–126.81	1.66 to 2.10
	24 hr γH2AX	16.3	8.95	2.51–21.59	0.40 to 1.33
	24 hr RNC	>26.2	295.11	79.87–3,004.35	1.90 to 3.48
7,12-Dimethylbenzanthracene	4 hr p53	>32.9	49.90	16.63–418.60	1.22 to 2.62
	24 hr p53	23.7	18.06	9.52–34.87	0.98 to 1.54
	4 hr γH2AX	88.7	28.07	No DR	No DR
	24 hr γH2AX	NC	No DR	No DR	No DR
	24 hr RNC	21.8	24.43	8.59–73.37	0.93 to 1.87
Benzo[a]pyrene	4 hr p53	NC	3.06	1.54–7.23	0.19 to 0.86
	24 hr p53	>80	9.14	4.98–17.88	0.70 to 1.25
	4 hr γH2AX	>463	1,357.44	629.63–3,448.28	2.80 to 3.54
	24 hr γH2AX	1.8×10^6	11.18	3.47–34.73	0.54 to 1.54
	24 hr RNC	>31.4	No DR	No DR	No DR
Cyclophosphamide	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	870	702.02	535.64–923.65	2.93 to 2.97
	4 hr γH2AX	>101	158.03	115.69–227.54	2.06 to 2.36
	24 hr γH2AX	3,249	1,589.81	305.34–5,922.33	2.48 to 3.77
	24 hr RNC	797	963.44	466.40–2,405.30	2.67 to 3.38
Dibenzo[a,l]pyrene	4 hr p53	>56	148.74	88.94–316.87	1.95 to 2.50
	24 hr p53	>962	1,162.83	50.24–1,578.48	1.70 to 3.20
	4 hr γH2AX	22,727	4,666.876	24.49–6,967.37	1.39 to 3.84
	24 hr γH2AX	>38,462	22,167.08	6,325.46–130,672.27	3.80 to 5.12
	24 hr RNC	>1,220	No DR	No DR	No DR
Diethylnitrosamine	4 hr p53	NC	No DR	No DR	No DR
	24 hr p53	NC	No DR	No DR	No DR
	4 hr γH2AX	>3.5	5.90	3.64–10.47	0.56 to 1.02
	24 hr γH2AX	>10	No DR	No DR	No DR
	24 hr RNC	NC	3.58	1.53–9.10	0.18 to 0.96
Mitomycin C	4 hr p53	0.51	0.52	0.37–1.06	−0.43 to 0.03
	24 hr p53	0.73	0.27	0.20–0.36	−0.70 to −0.44
	4 hr γH2AX	0.82	0.44	0.31–0.62	−0.51 to −0.21
	24 hr γH2AX	0.25	0.07	0.01–0.19	−2.00 to −0.72
	24 hr RNC	0.31	0.21	0.12–0.36	−0.92 to −0.44
PhIP	4 hr p53	>3.1	No DR	No DR	No DR
	24 hr p53	>34.6	No DR	No DR	No DR
	4 hr γH2AX	>1,633	4,719.67	1,436.10–33,246.07	3.16 to 4.52
	24 hr γH2AX	>781	2067.12	336.19–189,075,630.25	2.53 to 8.28
	24 hr RNC	>18.2	703.43	230.83–6,142.75	2.36 to 3.79

(Continues)

TABLE 1 (Continued)

Compound	Endpoint	S9 potentiation ratio from Tian et al. (2020) (µM)	BMD-BMD ratio[a]	S9 potency ratio CI range (µM)[b]	S9 potency ratio CI range (Log10 µM)[c]
Resorcinol	4 hr p53	NC	0.22	0	0
	24 hr p53	0.11	0.07	0.0498–0.0937	−1.30 to −1.03
	4 hr γH2AX	0.26	0.23	0.1502–0.3594	−0.82 to −0.44
	24 hr γH2AX	0.05	0.03	0.0026–0.0659	−2.59 to −1.18
	24 hr RNC	0.047	0.04	0.0202–0.0645	−1.69 to −1.19

[a]The BMD-BMD ratio is presented to compare with the "S9 potentiation ratio" from Tian et al. (2020). In essence, the Tian et al. (2020) "S9 potentiation ratio" is a BMD-BMD ratio. The majority of BMD-BMD ratios from the re-analysis are similar to the Tian et al. (2020) "S9 potentiation ratios" showing that the BMD analyses do not differ significantly (same order of magnitude). One outlier is the analysis of the benzo[a]pyrene 24 hr γH2AX endpoint where the BMD-BMD ratio is significantly smaller than the "S9 potentiation ratio." This is likely due to increased precision resulting from combined covariate analysis of the clastogen compounds in the reanalysis of this endpoint. The significant difference between "S9 potentiation ratios" and the BMD-BMD ratios is the tight restrictions placed on the BMDs to calculate the "S9 potentiation ratio" denoted by the greater than symbol (>). This restriction resulted in mischaracterized comparison of potency.

[b]The S9 potency ratio CIs ranges are provided in the original scale to provide a like-for-like comparison with the Tian et al. (2020) "S9 potentiation ratio" presented in the original scale.

[c]The S9 potency ratio CI ranges are provided in the log scale to provide a like-for-like comparison with the comparative potency plots (Figures 1–10) presented in this article in the log scale.

[d]Greater than sign (>) from Tian et al. (2020) where tight restrictions were place on the BMD based on the upper concentration tested.

BMDU +S9 to BMDL −S9, to the BMDL +S9 to BMDU −S9. The "S9 potency ratio CI" is synonymous with the magnitude of the potency difference (range) exhibited for each compound/endpoint combination after exposure to S9. Although there were only 2 experimental conditions (+/− S9) included here, the same algorithm could apply to 2 or more experimental conditions where one potency is compared to another potency of interest. The "S9 potency ratio CIs" were calculated in the original scale so that the values can be compared to the "S9 potentiation ratios" values obtained from BMD-BMD ratios by Tian et al. (2020). Values greater than 1 represent increased compound potency for a particular endpoint after S9 exposure, whilst values less than 1 confer the inverse. 0 values represent CIs that overlap and hence potency differences are statistically indefensible. The BMD-BMD ratios from our analysis are also shown for comparative purposes. We have also displayed the S9 potency ratios in the Log scale for the sole purpose of aiding in demonstration of differences in orders of magnitude. The values are displayed in Table 1 in comparison with the S9 potentiation ratios from Tian et al. (2020).

Comparing our "S9 potency ratio CIs" with the S9 potentiation ratio obtained by Tian et al. (2020) shows that in almost all cases, the "S9 potentiation ratio" that was derived from BMD point estimates is either mischaracterized through tight restrictions on the BMD (where the researchers denoted these ratios with a > sign) or is in the upper range of our "S9 potency ratio CI" range, indicating a tendency for over estimation of the S9 bioactivation effect.

Calculating an "S9 potency ratio CI" for these experiments accurately conveys the uncertainty measurements which should be accounted for when assessing potency comparisons across conditions. We can summarize the "S9 potency ratio CIs" calculated here by stating that S9 increases the potency of the compounds ranging from approximately 1–2 orders of magnitude (Log10 µM) for lower potency differences, to approximately 3–5 orders of magnitude (Log10 µM) for higher potency differences. The effect of S9 is statistically insignificant for some compound/endpoint combinations where BMD CIs overlap. In other instances, S9 exposure decreased potency (expressed as negative values) of compounds by approximately 0.4 to 2.0 orders of magnitude.

3.3 | S9 potency ratio CI range

As previously mentioned, the "S9 potency ratio CI" values span several orders of magnitude. In order to objectively define the range of "S9 potency ratio CIs" obtained, we analyzed the lower and upper values (Log scale) of each "S9 potency ratio CI" via unsupervised hierarchical clustering. The clustering was based on squared Euclidean distance "Ward's method" (Ward and Hook 1963) between points. The resulting groups are presented in Figure 11 in the form of a two-dimensional dendrogram with accompanying heat map. The upper and lower values of the "S9 potency ratio CIs" associated with the in vitro biomarkers (4 hr p53, 24 hr p53, 4 hr γH2AX, 24 hr γH2AX, and 24 hr RNC) are displayed on the X axis. Compounds are plotted on the Y axis. The heat map represents the order of magnitude difference in "S9 potency ratio CIs." Compound/endpoint combinations where S9 exposure increased genotoxic potency are displayed with varying intensities of red on the heat map, whereas compounds and endpoints where S9 exposure decreased potency are displayed with varying intensities of blue on the heat map. Compound/endpoint combinations where the dose–response following S9 exposure was not statistically significantly different were plotted as zero values and displayed as gray-blue on the heatmap. Additionally, the clustering platform could not accommodate missing values for compounds/endpoint combinations that showed no dose–response. In these instances, a zero value was entered to accommodate the clustering method.

WHEELDON ET AL.

Environmental and
Molecular Mutagenesis

Environmental
Mutagenesis and
Genomics Society

WILEY | 213
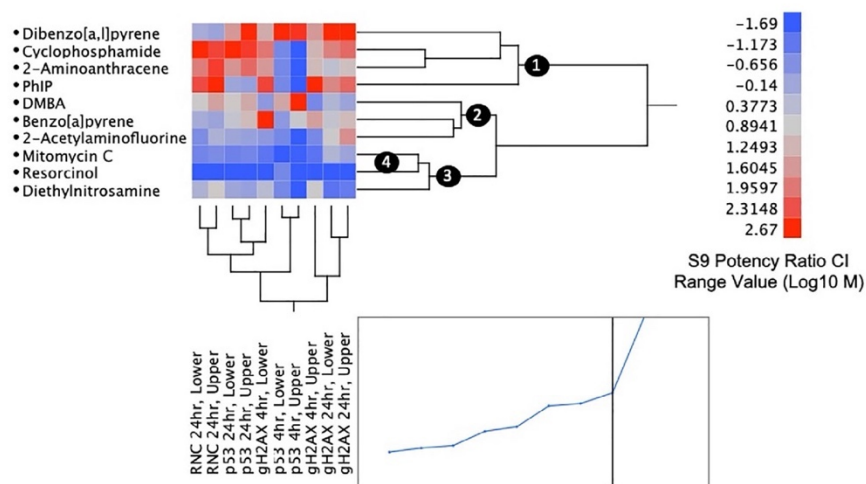
FIGURE 11 Unsupervised hierarchical clustering results are shown as a two-dimensional dendrogram with heatmap for the 10 clastogens "S9 potency ratio CIs." The values of the lower and upper confidence limits of the "S9 potency ratio CIs" were included in the analysis. The lower and upper bound of the "S9 potency ratio S9 CIs" are plotted on the *X*-axis per endpoint combination. Compounds are plotted on the *Y*-axis. Increasing intensities of red indicate a strong tendency for S9 exposure to increase the genotoxic potency of a compound for a specific endpoint. An increasing intensity of blue indicates the converse, where the presence of S9 decreases the genotoxic potency of a compound for a specific endpoint. Gray-blue represents compound/endpoint combinations where the dose–response following S9 exposure was not statistically significantly different (overlapping CIs), or where a zero value was included to accommodate the clustering method in instances where no S9 potency ratio CIs were obtained (infinite BMDUs). There are 4 distinct clades that group compounds into (1) high, (2) low, (3) zero, and (4) negative (as a subset of clade 3), effects on potency as a result of S9 exposure. Abbreviation: DMBA = 7,12-dimethylbenzanthracene

There are 4 distinct clades identified in the dendrogram: (a) Chemicals whose genotoxic potency was *dramatically* increased in the presence of S9 (dibenzo[a,l]pyrene, cyclophosphamide, 2-amino anthracene, and PhIP); (b) chemicals whose genotoxic potency was increased in the presence of S9 (dimethylbenzanthracene, benzo[a] pyrene, and 2-acetylaminofluorene); (c) chemicals whose genotoxic potency was not increased in the presence of S9 (mitomycin c, resorcinol, and diethylnitrosamine); and (d) a subset of clade 3, clade 4, whose genotoxic potency was reduced in the presence of S9 (mitomycin c, and resorcinol; indicated by a strong dark blue in the heat map).

The clustering results demonstrate that distinct groups exist that best describe the range of S9-dependent effects into high, low, zero and negative categories. The division of the S9 potency ratio CIs into groups follows application of quantitative metrics allowing objective interpretation in a hazard-based scale context. It is likely that when a more diverse set of clastogens are analyzed, it will be possible to detect further subdivisions of potency effect (e.g., low, medium, high, very high). The same clustering method could be applied to other dose–response comparative analyses. Several examples are envisioned, for instance the study of compound potency effects between different exposure durations (subchronic vs. chronic dosing regimens), between different target tissues (liver vs. bone marrow), or between different in vitro cell lines (liver HepaRG vs. HepG2, or V79 vs. CHO).

4 | CONCLUSIONS

The results of the data analysis presented here illustrate the necessity to utilize the full BMD CIs to draw conclusions from dose–response datasets. We have demonstrated that in comparative potency analysis, use of BMD point estimates can yield mischaracterized or overestimated potency ratios during instances where the width of the CIs varies considerably between conditions.

Interpretation of BMD CIs relies upon visual scrutiny by the evaluating scientist, and we speculate that one could be overwhelmed with the results from hundreds of compounds from in vitro screening experiments. Therefore, the evaluating scientist may wish to use statistical methods such as hierarchical clustering to aid in data interpretation. When objectively considering the range of "S9 potency ratio CIs" presented in our analysis, one is provided with a scale of high to low, zero, and negative S9 potency effects. While we illustrated the potency effect of S9 in an in vitro genotoxicity test system, the same approach can be applied to any comparative genotoxicity potency analysis. We propose that investigators apply the following stepwise approach when performing comparative potency analysis of 2 or more experimental conditions using the BMD methodology:

1. Utilize the results of combined covariate BMD analyses to plot BMD CIs that represent potency across the endpoints evaluated

under different dependent variables. For graphical purposes only, limit disproportionately high BMDs that display infinite BMDU values.

2. Derive the lower and upper values of "potency ratio CIs" between conditions by comparing BMDU condition 1 to BMDL condition 2, and BMDL condition 1 to BMDU condition 2 (ad infinitum), respectively.

3. Objectively identify groups in the data that best describe the magnitude of the difference in potency observed between conditions. While a scientist can successfully identify patterns in CI plots of a few experimental conditions (a small number of compounds, animal sex, limited number of cell lines, etc.) via visual techniques, we contend that hierarchical clustering may add significant value to the interpretation of large datasets, particularly those from in vitro screening experiments of large numbers of compounds and biomarker combinations.

CONFLICT OF INTEREST

S. D. D., S. M. B., and J. C. B. are employed by Litron Laboratories where the benchtop work was conducted. Litron owns a patent covering the flow-cytometry based assay described in this manuscript and sells a commercial kit based on these procedures: MultiFlow® DNA Damage kit-p53, γH2AX, Phospho-Histone H3.

AUTHOR CONTRIBUTIONS

R. P. W. is the lead author of the manuscript and identified the issues with the S9 potentiation ratio published in Tian et al. (2020). R. P. W. performed the BMD analysis reported here as well as leadership on the results analysis and interpretation. S. D. D's laboratory provided the raw data. R. P. W. suggested the use of unsupervised hierarchical clustering to evaluate the S9 potency ratio CIs, and S. D. D. achieved this. R. P. W. provided the first draft of the manuscript, and all authors contributed to the revisions that followed.

ORCID

Ryan P. Wheeldon https://orcid.org/0000-0002-4893-246X
George E. Johnson https://orcid.org/0000-0001-5643-9942

REFERENCES

Allemang, A., Mahony, C., Lester, C. and Pfuhler, S. (2018) Relative potency of fifteen pyrrolizidine alkaloids to induce DNA damage as measured by micronucleus induction in HepaRG human liver cells. *Food and Chemical Toxicology*, 121, 72–81.

Bemis, J.C., Wills, J.W., Bryce, S.M., Torous, D.K., Dertinger, S.D. and Slob, W. (2016) Comparison of in vitro and in vivo clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. *Mutagenesis*, 31, 277–285.

Bryce, S.M., Bernacki, D.T., Bemis, J.C., Spellman, R.A., Engel, M.E., Schuler, M., Lorge, E., Heikkinen, P.T., Hemmann, U., Dertinger, S.D., et al. (2017) Interlaboratory evaluation of a multiplexed high information content in vitro genotoxicity assay. *Environmental and Molecular Mutagenesis*, 58, 146–161.

Bryce, S.M., Bernacki, D.T., Smith-Roe, S.L., Witt, K.L., Bemis, J.C. and Dertinger, S.D. (2018) Investigating the generalizability of the Multi-Flow® DNA damage assay and several companion machine learning models with a set of 103 diverse test chemicals. *Toxicological Sciences*, 162, 146–166.

Bryce, S.M., Bernacki, T., Bemis, J.C. and Dertinger, S.D. (2016) Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environmental and Molecular Mutagenesis*, 57, 171–189.

Crump, K.S. (1984) A new method for determining allowable daily Intakes1. *Toxicological Sciences*, 4, 854–871.

Dearfield, K.L., Gollapudi, B.B., Bemis, J.C., Daniel Benz, R., Douglas, G.R., Elespuru, R.K., Johnson, G.E., Kirkland, D.K., LeBaron, M.J., Luijten, M., et al. (2017) Next generation testing strategy for assessment of genomic damage: a conceptual framework and considerations. *Environmental and Molecular Mutagenesis*, 58, 264–283.

Dertinger, S.D., Kraynak, A.R., Wheeldon, R.P., Bernacki, D.T., Bryce, S.M., Hall, N., Bemis, J.C., Galloway, S.M., Escobar, P.A. and Johnson, G.E. (2019) Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered Multiflow® assay data analysis strategy. *Environmental and Molecular Mutagenesis*, 60, 513–533.

European Food Safety Agency (EFSA). (2009) European food safety authority. Guidance of the scientific committee on use of the benchmark dose approach in risk assessment. *EFSA Journal*, 1150, 1–72.

European Food Safety Agency (EFSA). (2017) Update: use of the benchmark dose approach in risk assessment. *EFSA Journal*, e04658, 15.

Gollapudi, B.B., Johnson, G.E., Hernandez, L.G., Pottenger, L.H., Dearfield, K.L., Jeffrey, A.M., Julien, E., Kim, J.H., Lovell, D.P., Thybaud, V., et al. (2013) Quantitative approaches for assessing dose–response relationships in genetic toxicology studies. *Environmental and Molecular Mutagenesis*, 54, 8–18.

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. *Environmental and Molecular Mutagenesis*, 52, 518–528.

Macgregor, J.T., Frötschl, R., White, P.A., Crump, K.S., Eastmond, D.A., Fukushima, S., Guérard, M., Hayashi, M., Soeteman-Hernández, L.G. and Thybaud, V. (2015) IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). *Mutation Research – Genetic Toxicology and Environmental Mutagenesis*, 783, 55–65.

Slob, W. (2002) Dose-response modeling of continuous endpoints. *Toxicological Sciences*, 66, 298–312.

Slob, W. and Setzer, R. (2014) Shape and steepness of toxicological dose-response relationships of continuous endpoints. *Critical Reviews in Toxicology*, 44, 270–297.

Soeteman-Hernández, L.G., Fellows, M.D., Johnson, G.E. and Slob, W. (2015a) Correlation of in vivo versus in vitro benchmark doses (BMDs) derived from micronucleus test data: a proof of concept study. *Toxicological Sciences*, 148, 355–367.

Soeteman-Hernández, L.G., Johnson, G.E. and Slob, W. (2015b) Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. *Mutagenesis*, 31, 347–358.

366

Tian, S., Cyr, A., Zeise, K., Bryce, S.M., Hall, N., Bemis, J.C. and Dertinger, S.D. (2020) 3Rs-friendly approach to exogenous metabolic activation that supports high-throughput genetic toxicology testing. *Environmental and Molecular Mutagenesis*, 61, 408–432.

Ward, J.H. and Hook, M.E. (1963) Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement*, 23, 69–81.

Wheeldon, R.P., Bernacki, D.T., Dertinger, S.D., Bryce, S.M., Bemis, J.C. and Johnson, G.E. (2020) Benchmark dose analysis of DNA damage biomarker responses provides compound potency and adverse outcome pathway information for the topoisomerase II inhibitor class of compounds. *Environmental and Molecular Mutagenesis*, 61, 396–407.

White, P.A., Long, A.S. and Johnson, G.E. (2020) Quantitative interpretation of genetic toxicity dose-response data for risk assessment and regulatory decision-making: current status and emerging priorities. *Environmental and Molecular Mutagenesis*, 61, 66–83.

Wills, J.W., Johnson, G.E., Battaion, H.L., Slob, W. and White, P.A. (2017) Comparing BMD-derived genotoxic potency estimations across variants of the transgenic rodent gene mutation assay. *Environmental and Molecular Mutagenesis*, 58, 632–643.

Wills, J.W., Johnson, G.R., Doak, S.H., Soeteman-Hernández, L.G., Slob, W. and White, P.A. (2016) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. *Mutagenesis*, 31, 255–263.

Research Article

Benchmark Dose Analysis of DNA Damage Biomarker Responses Provides Compound Potency and Adverse Outcome Pathway Information for the Topoisomerase II Inhibitor Class of Compounds

Ryan P. Wheeldon [1]*, Derek T. Bernacki,[2†] Stephen D. Dertinger,[2] Steven M. Bryce,[2]
Jeffrey C. Bemis,[2] and George E. Johnson [1]

[1]*Institute of Life Science, Swansea University Medical School, Swansea University, Wales, United Kingdom*
[2]*Litron Laboratories, Rochester, New York*

INTRODUCTION

Wit...

Level (Td) analysis, strongest support was given to the Benchmark Dose (BMD) approach for estimating a PoD.
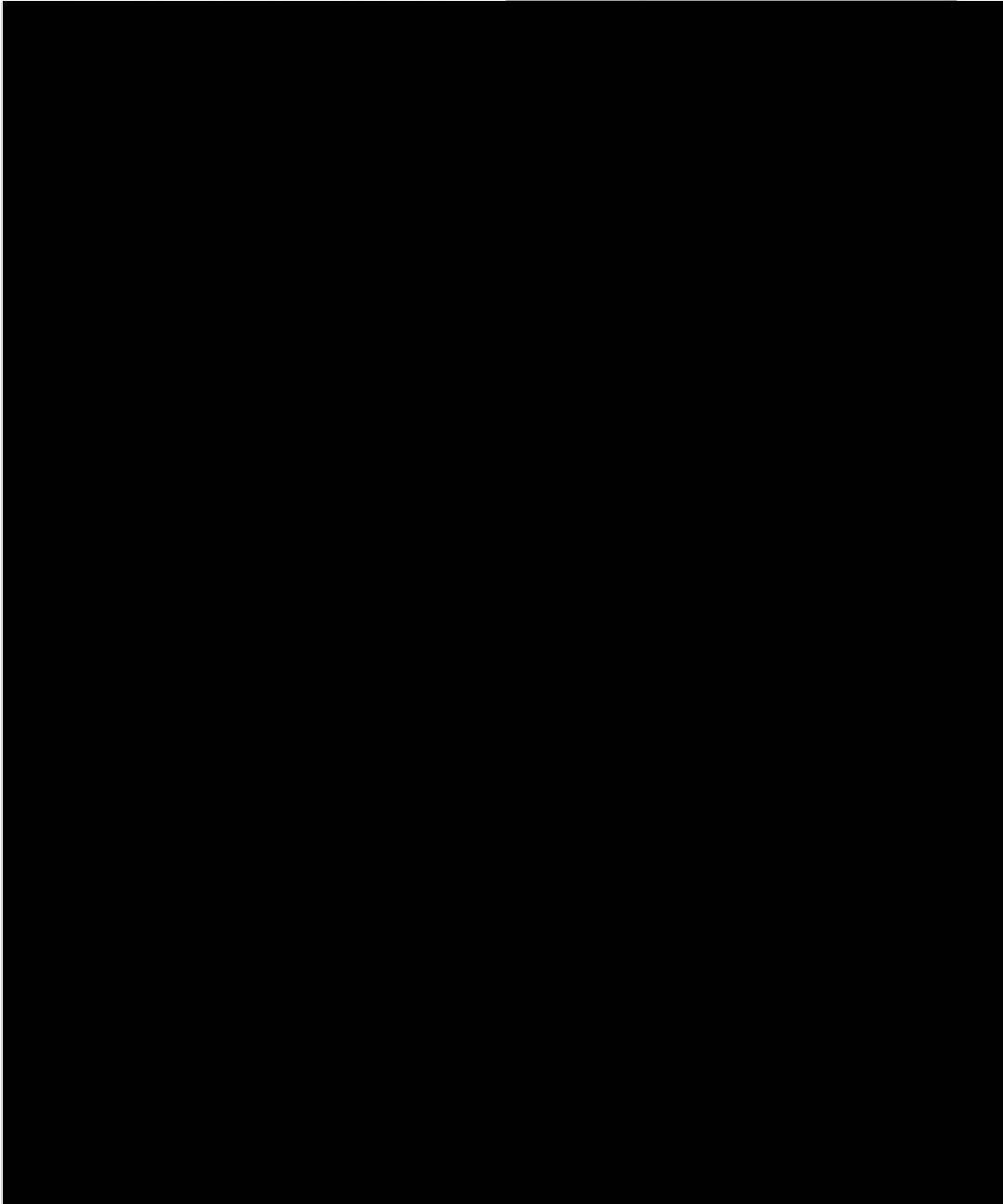
Crump (1984) introduced the statistical strength of the BMD

2014, 2016, 2017, 2018; Bernacki et al., 2016). However

aneugens, and non-genotoxic compounds (Bryce et al.,
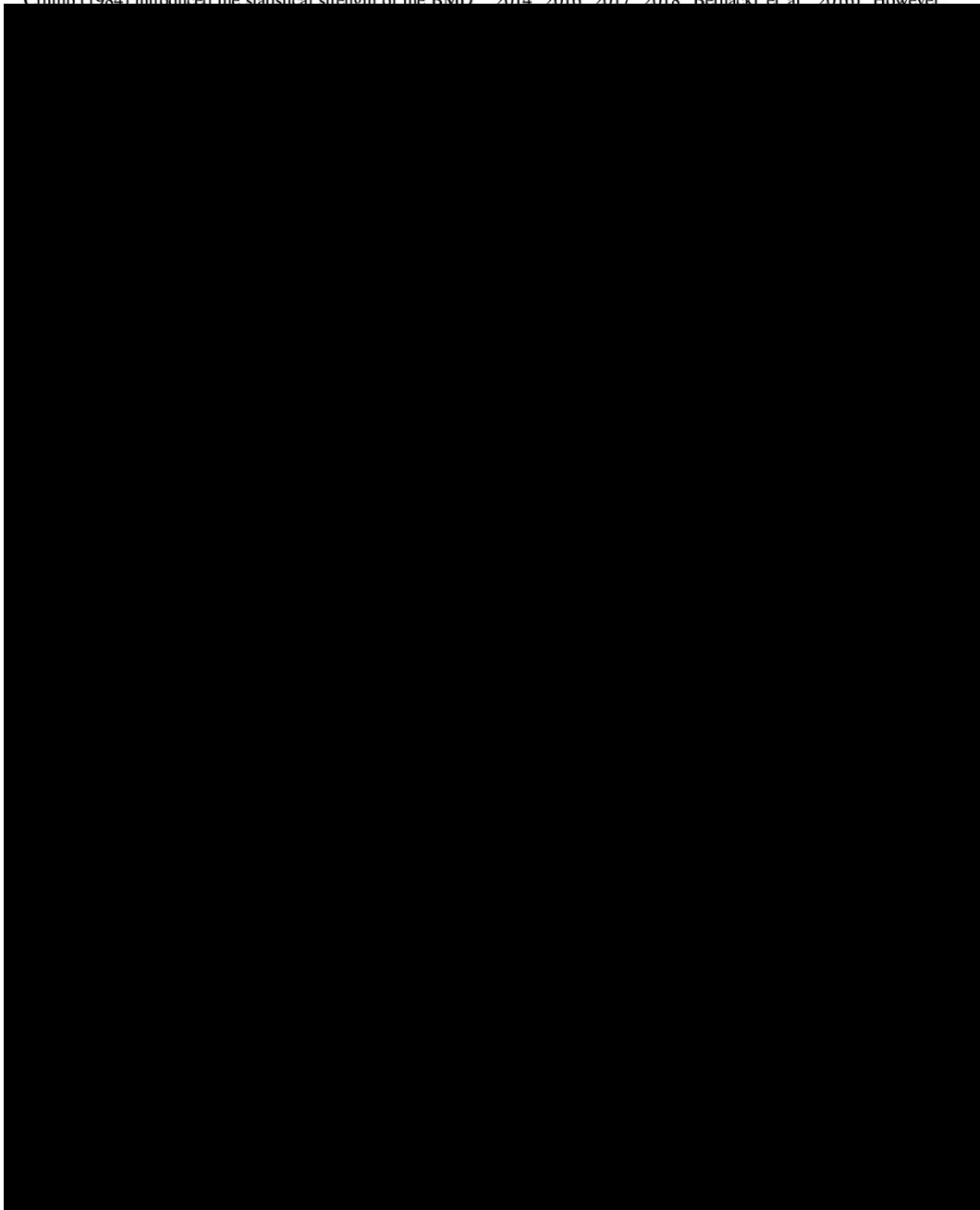
compound complex forms, and subsequent differences in

the compounds chemical structure elicit their action in the

TABLE I. List of Topo II poison compounds

organ, or organism/population level. The AOP concept   associated with the nucleosomes: fluorescein isothiocyanate (FITC)

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Benchmark Dose Analysis of DNA Damage Biomarker Responses   399
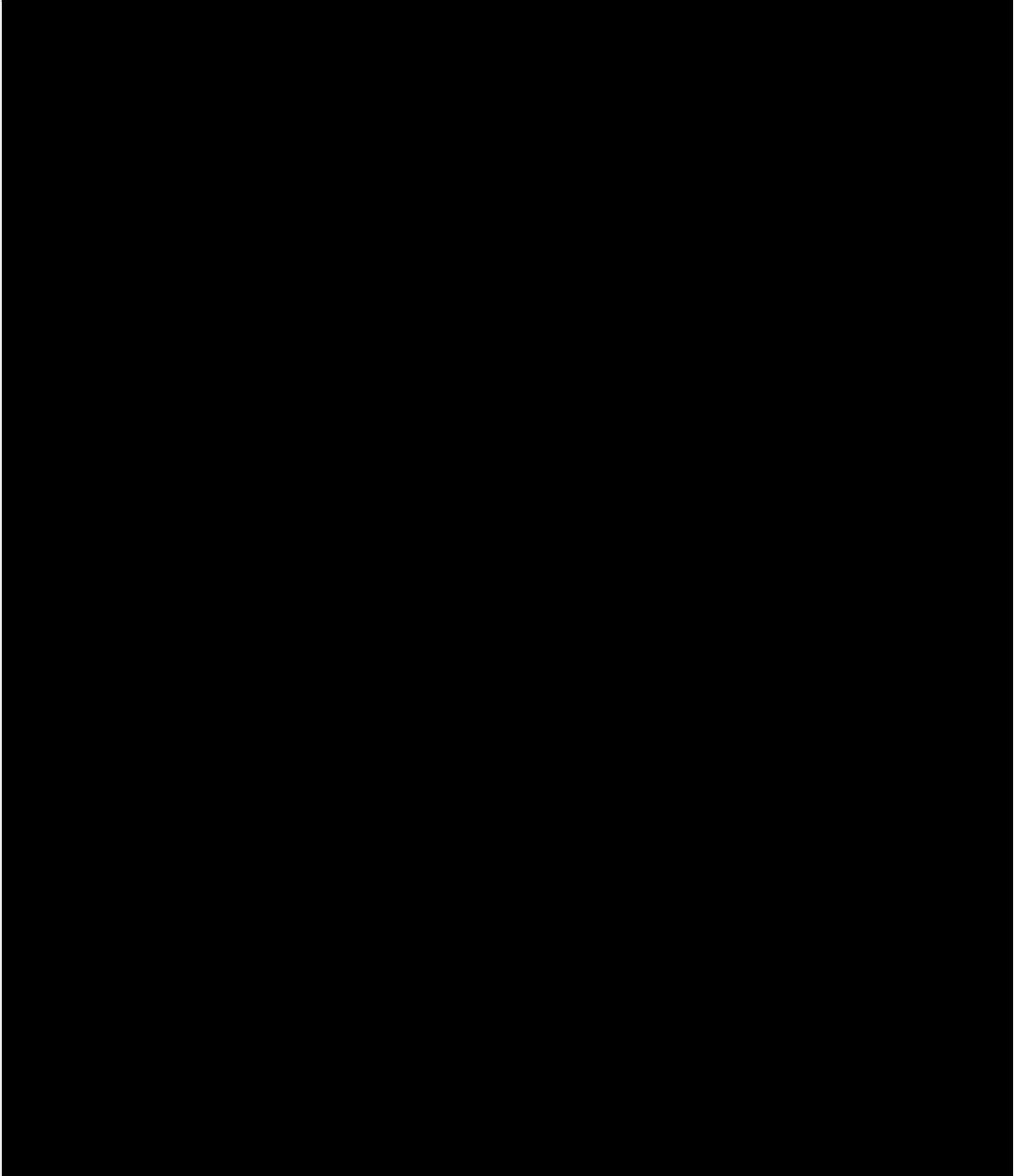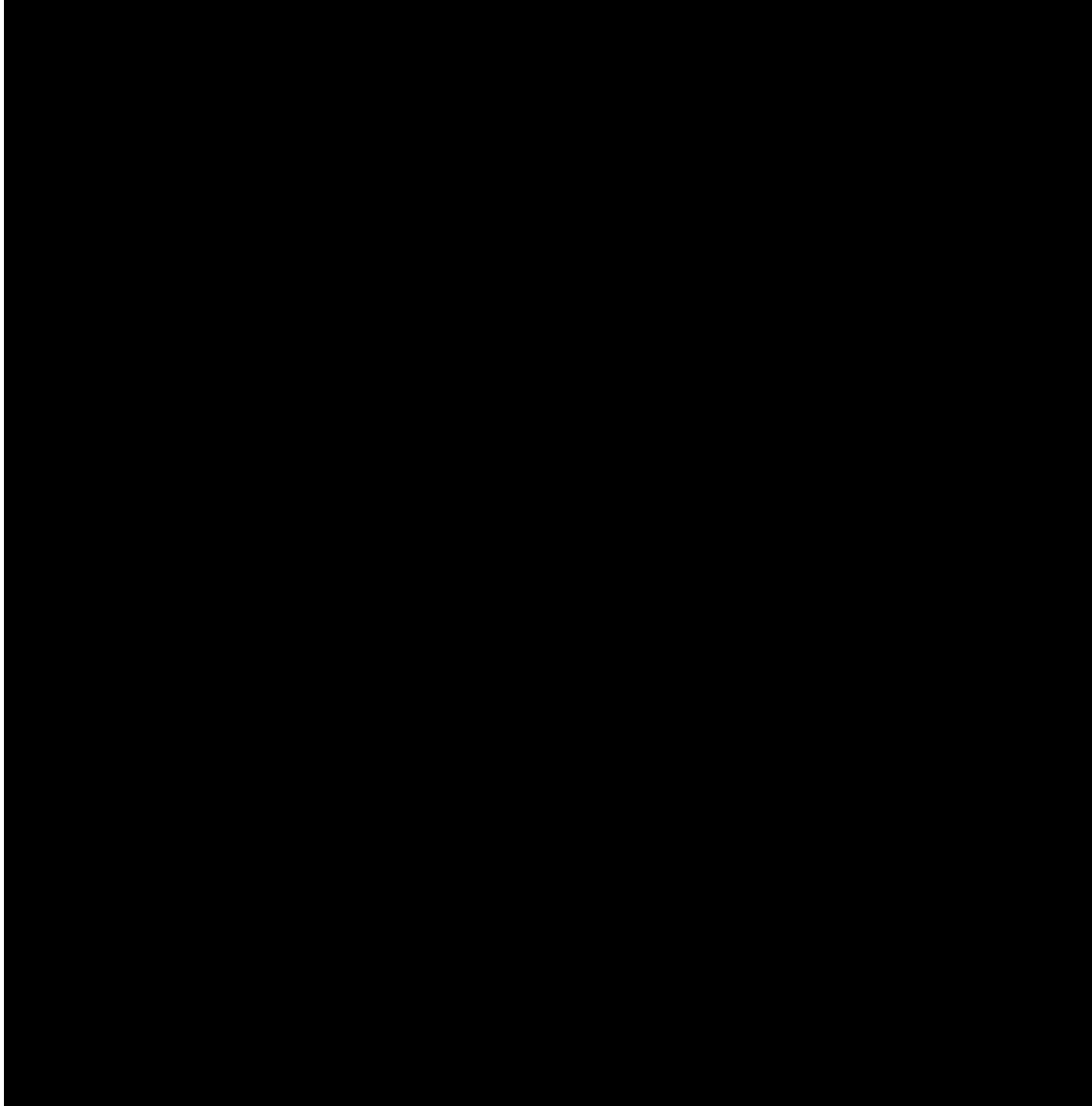
(BMD$_U$, BMD$_L$) were used to calculate the confidence interval around the   (Marchant et al., 2008). Derek Nexus has regulatory
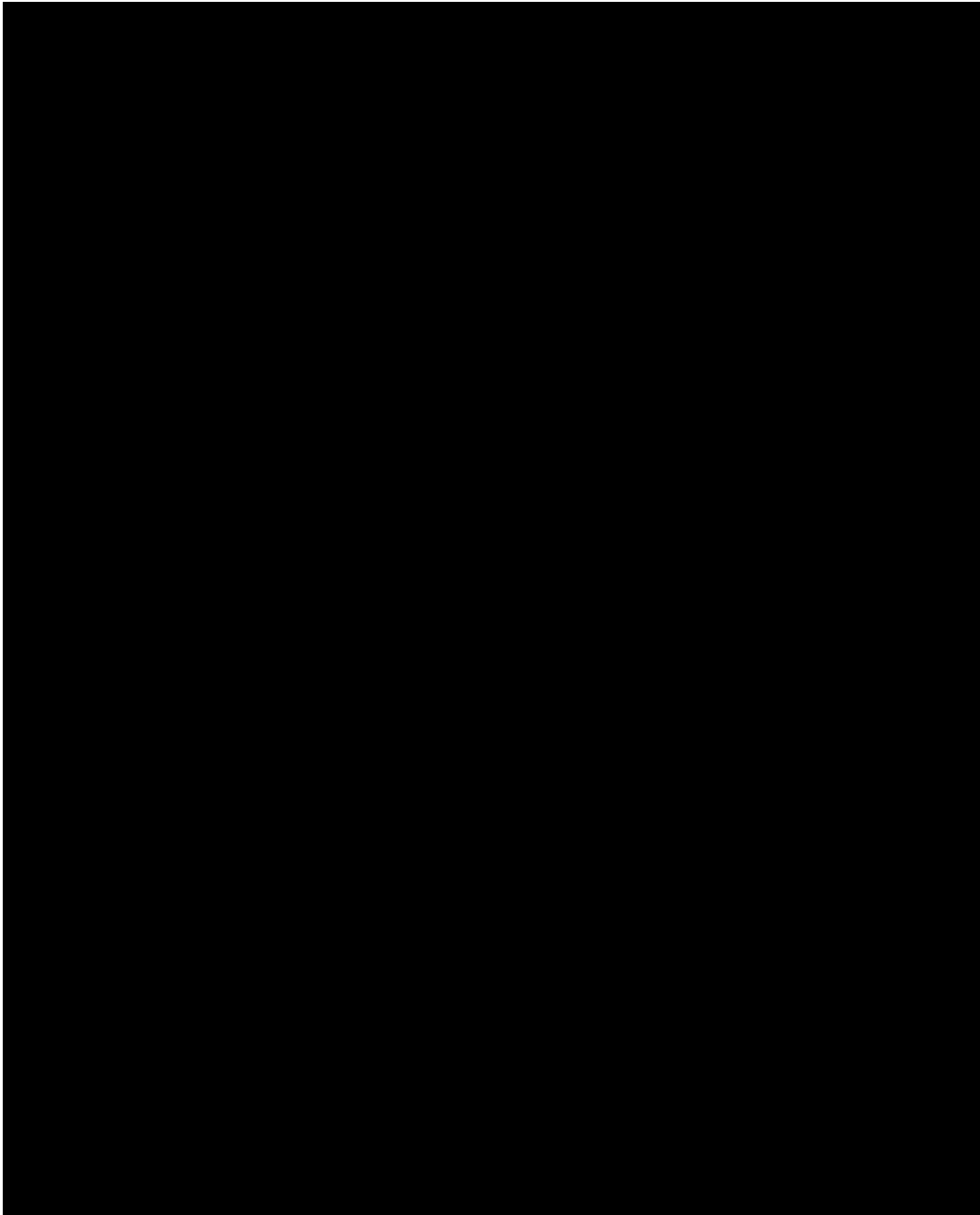
[...] response, ciprofloxacin failed to yield a dose response in the 4 hr exposure; hence, no BMD could be obtained. Infinite upper bound confidence interval for ciprofloxacin's response with 24 hr exposure is with absolute lower and upper confidence intervals. (C) Infinite upper bound confidence interval for the 24 hr p53 response to flumequine exposure.
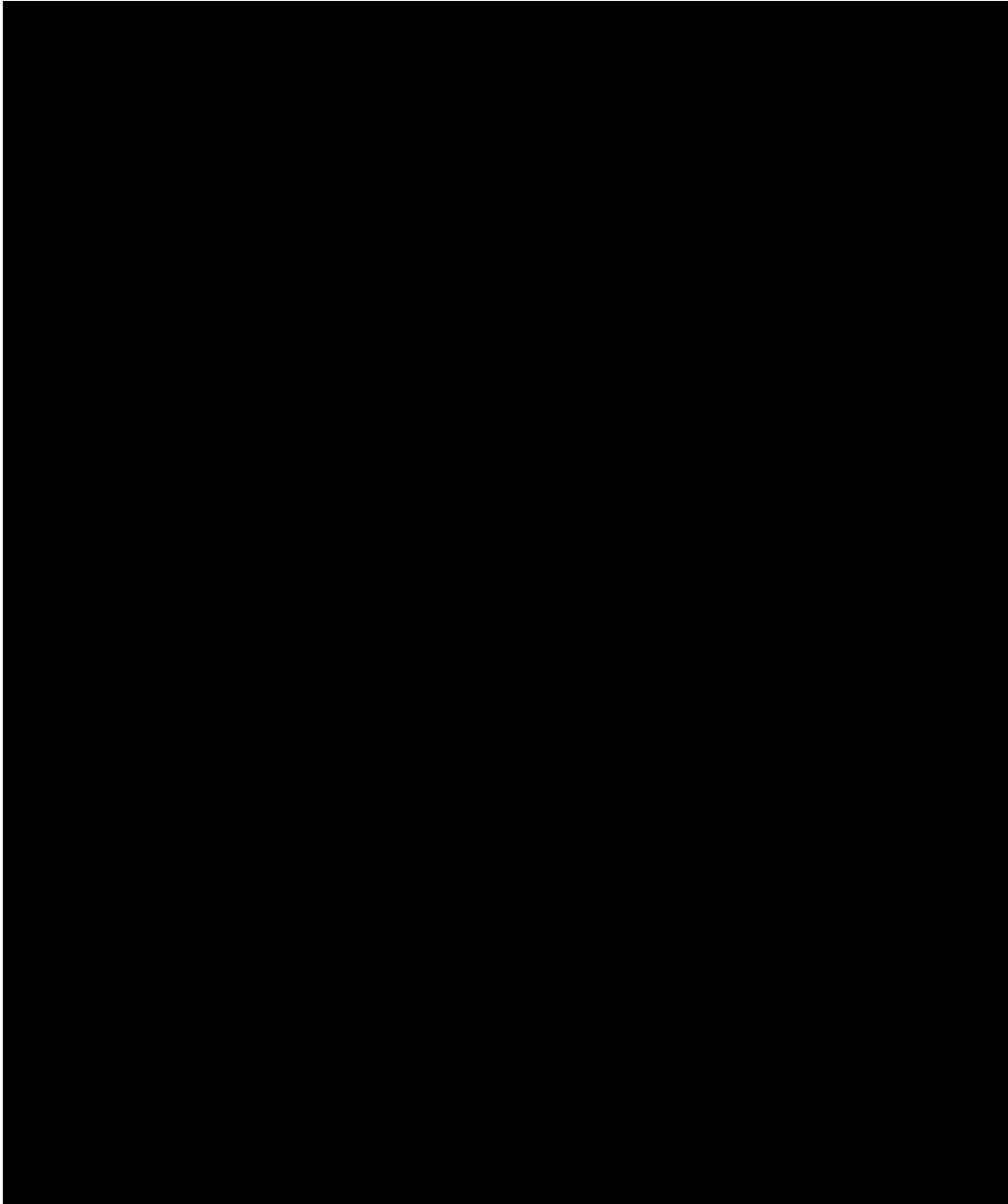
TABLE II. Topo II poison compound Derek Nexus derived structural information

(Continues)

TABLE II. (Continued)

of the compounds studied in the fluorescence-based assay (Shapiro and Austin 2014). Potency decreasing from left to right.
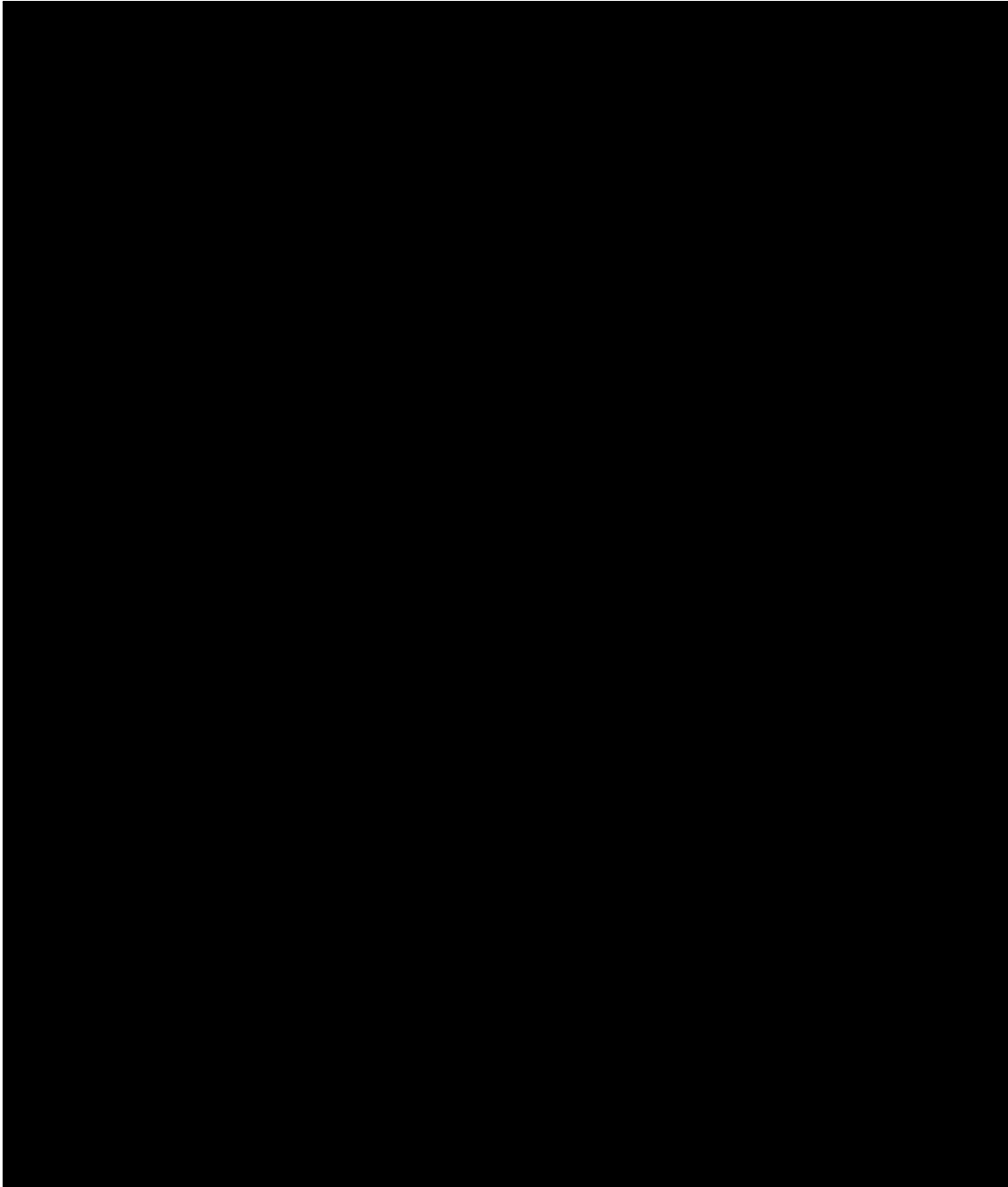
Environmental and Molecular Mutagenesis. DOI 10.1002/em

Benchmark Dose Analysis of DNA Damage Biomarker Responses   403

Topo II poisons that appear to act by covalently adducting   erase IV enzyme's serine and acidic residues. On the other
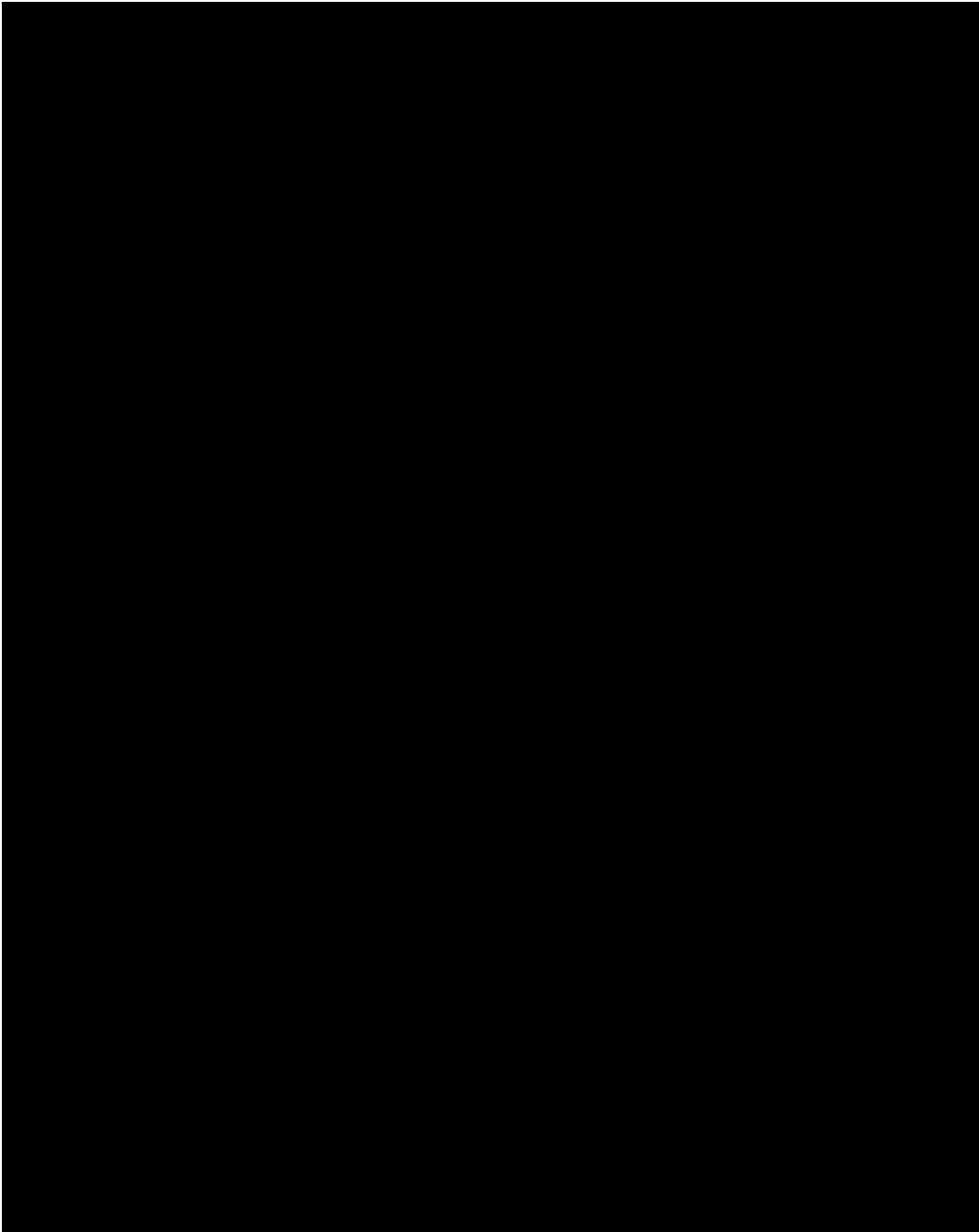
ships (RLRC), through to an AO. The AOP-KB is designed   strand break repair, leading to the manifestation of poorly
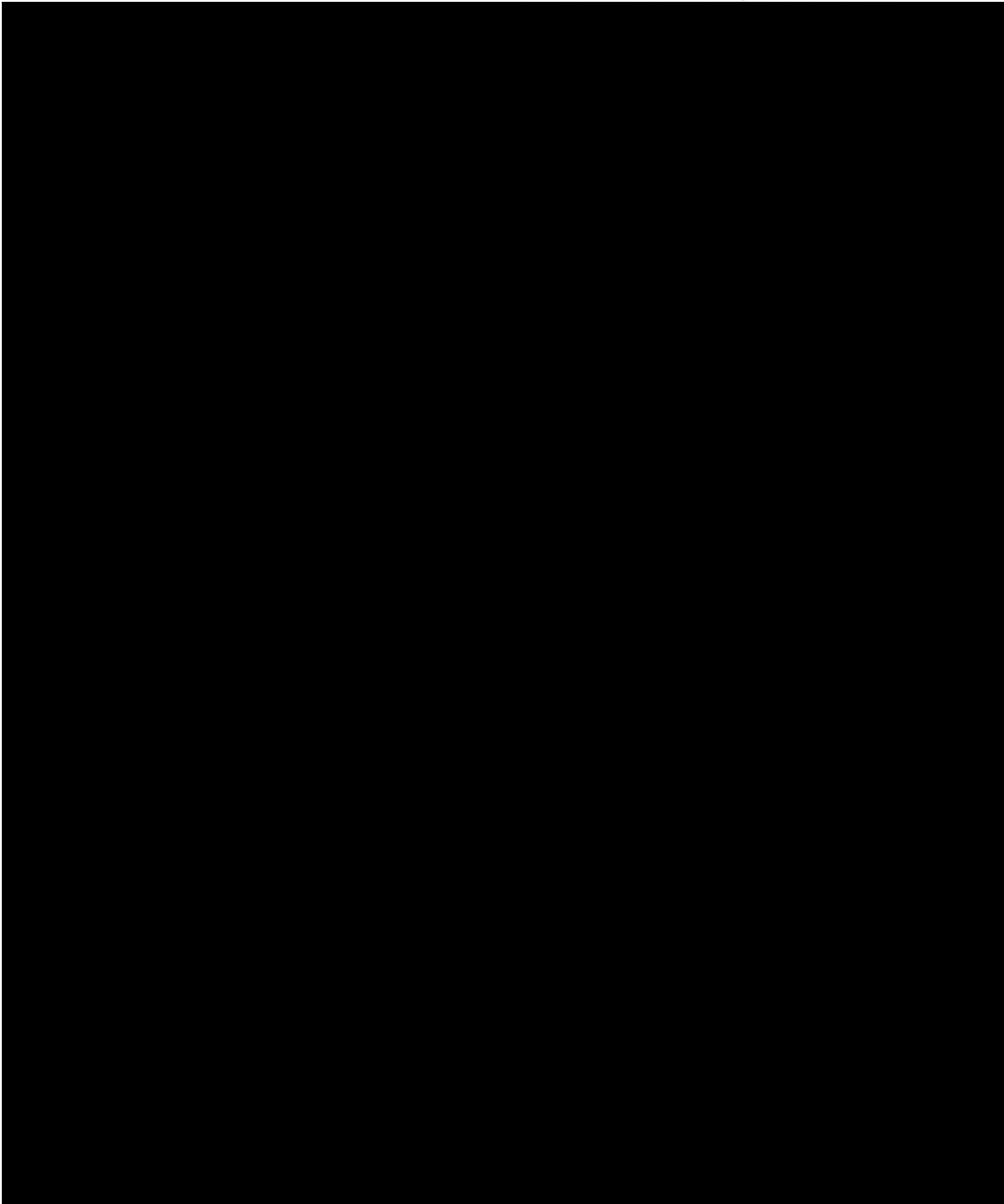
MultiFlow assay uses an anti-phospho-histone H3-PE marker to quantify the frequency of mitotic cells (Bryce
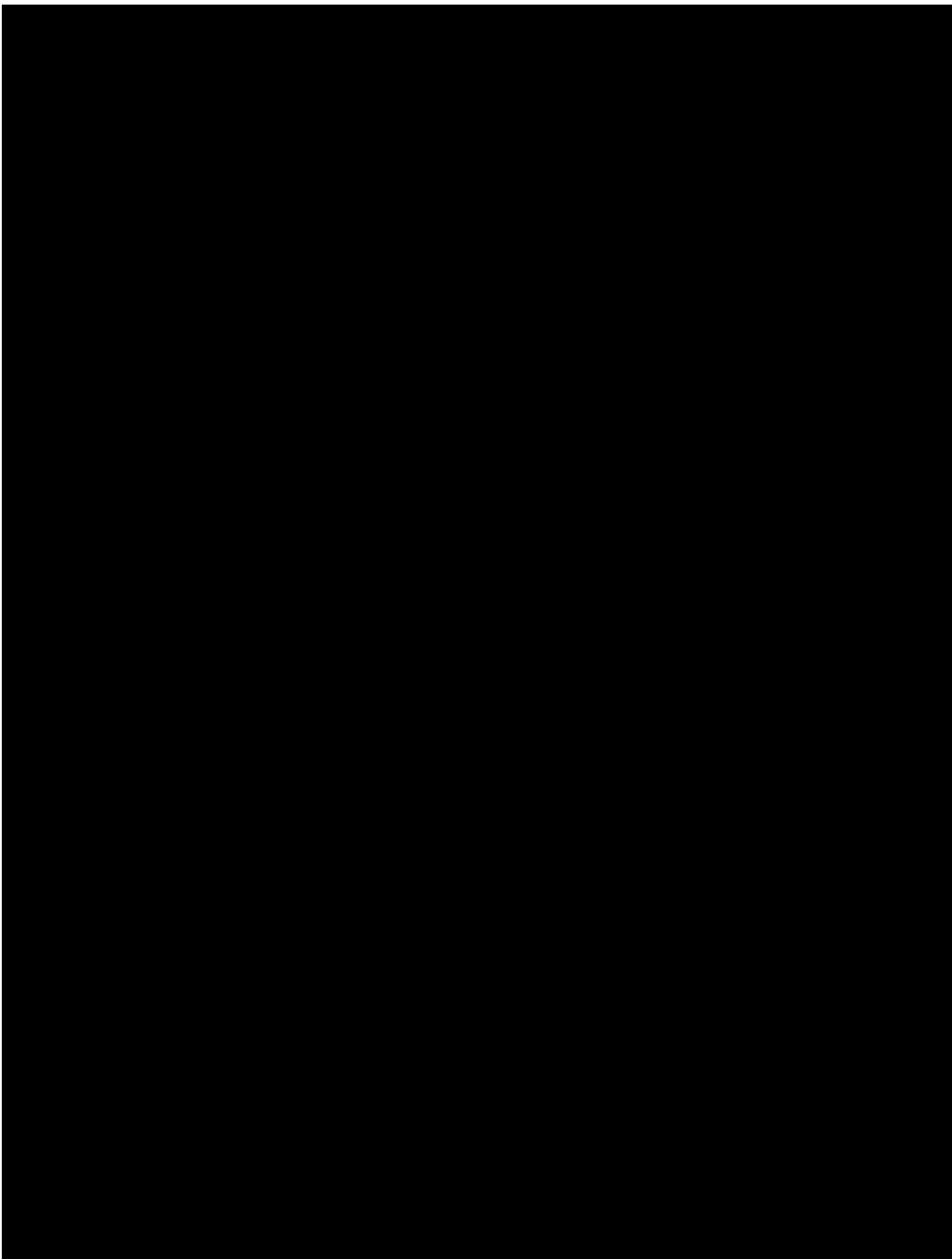
Schuler MJ, Dertinger SD. 2014. Interpreting in vitro micronucleus positive results: Simple biomarker matrix discriminates clastogens, fluoroquinolines with human topoisomerase II a and b. In Silico Pharmacol 5(1):4.

Environmental and Molecular Mutagenesis. DOI 10.1002/em

Benchmark Dose Analysis of DNA Damage Biomarker Responses   407
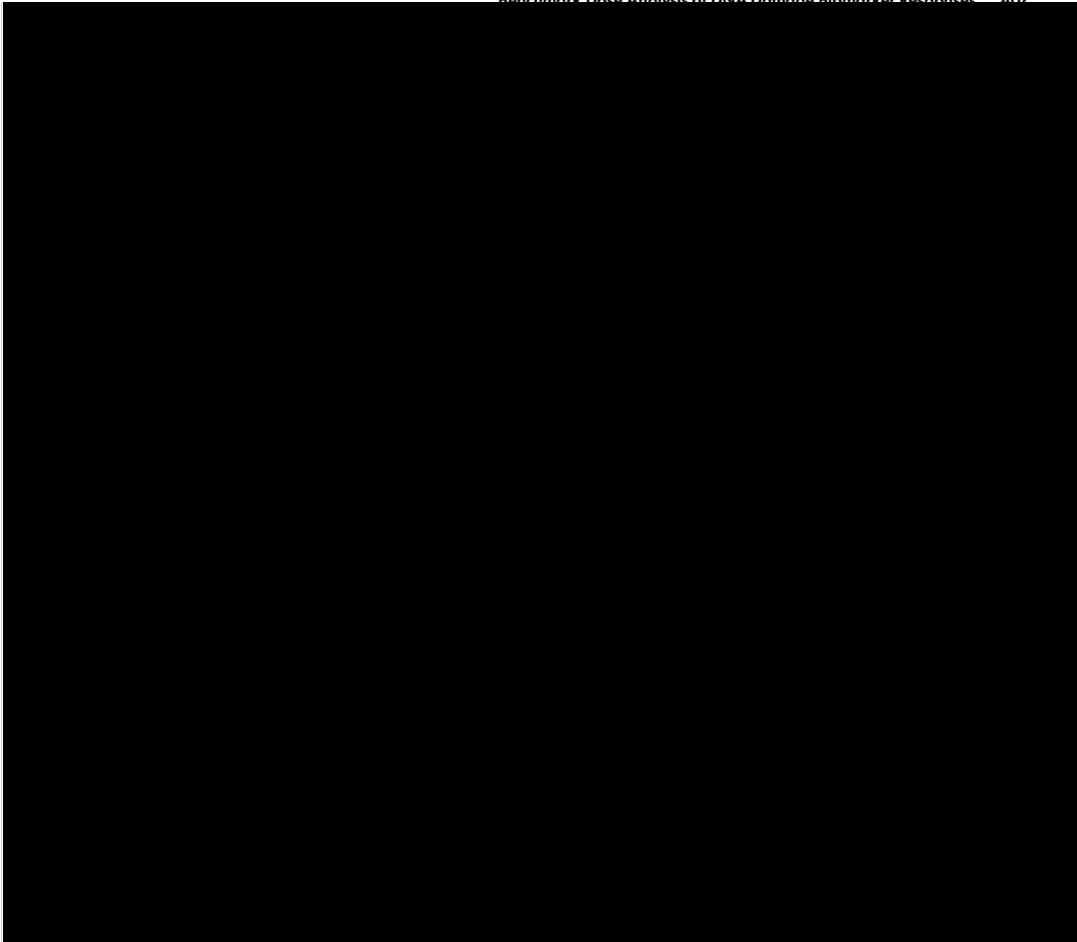
Scientific Committee EFSA. 2017. Use of the benchmark dose approach in risk assessment. EFSA J 15(1):4658.

Shapiro AB, Austin CA. 2014. A high-throughput fluorescence anisotropy-based assay for human topoisomerase II β-catalyzed

dose–response relationships in genetic toxicology. Arch Toxicol 91 (12):3799–3807.

Benchmark Dose analysis of DNA damage response endpoints coupled with QSAR analysis provides robust potency ranks and Adverse Outcome Pathway information for Topoisomerase II Inhibitors.

Ryan Wheeldon[1], Stephen Dertinger[2], George Johnson[1]
[1]Swansea University Medical School, Swansea, UK; [2]Litron Laboratories, Rochester NY, USA
Communication:

Appendix 5. Topo-II Poison 4-hour BMD Dose-Response Curves for MultiFlow Biomarker Responses

Figure A5.1. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 24-hour γH2AX biomarker response. Induction (increase) of the γH2AX response is indicative of DNA double strand breaks from Topo-II Poison exposure.
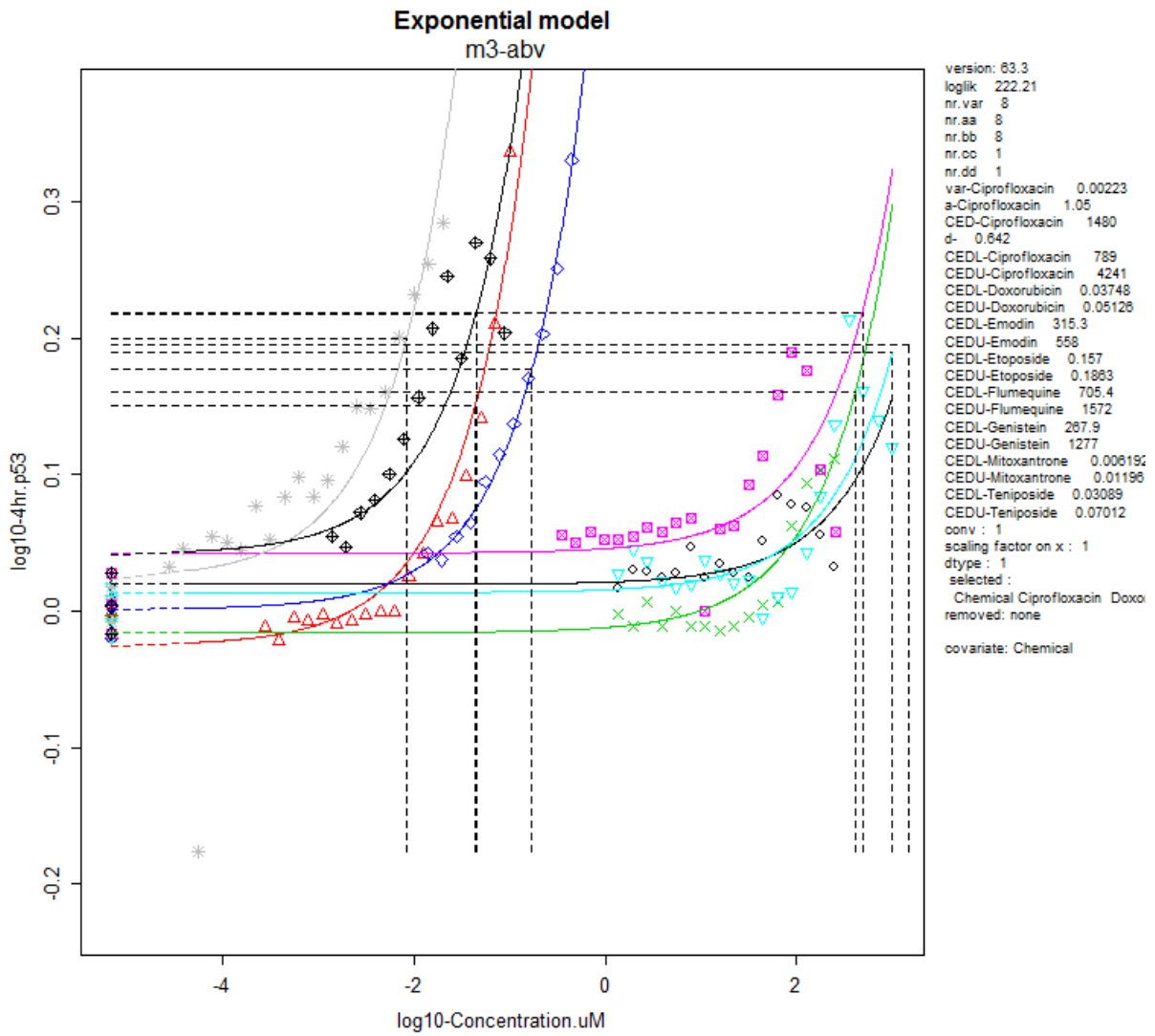
Figure A5.2. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 4-hour p53 biomarker response. Induction (increase) of the p53 response is indicative of cell cycle delay from Topo-II Poison exposure.
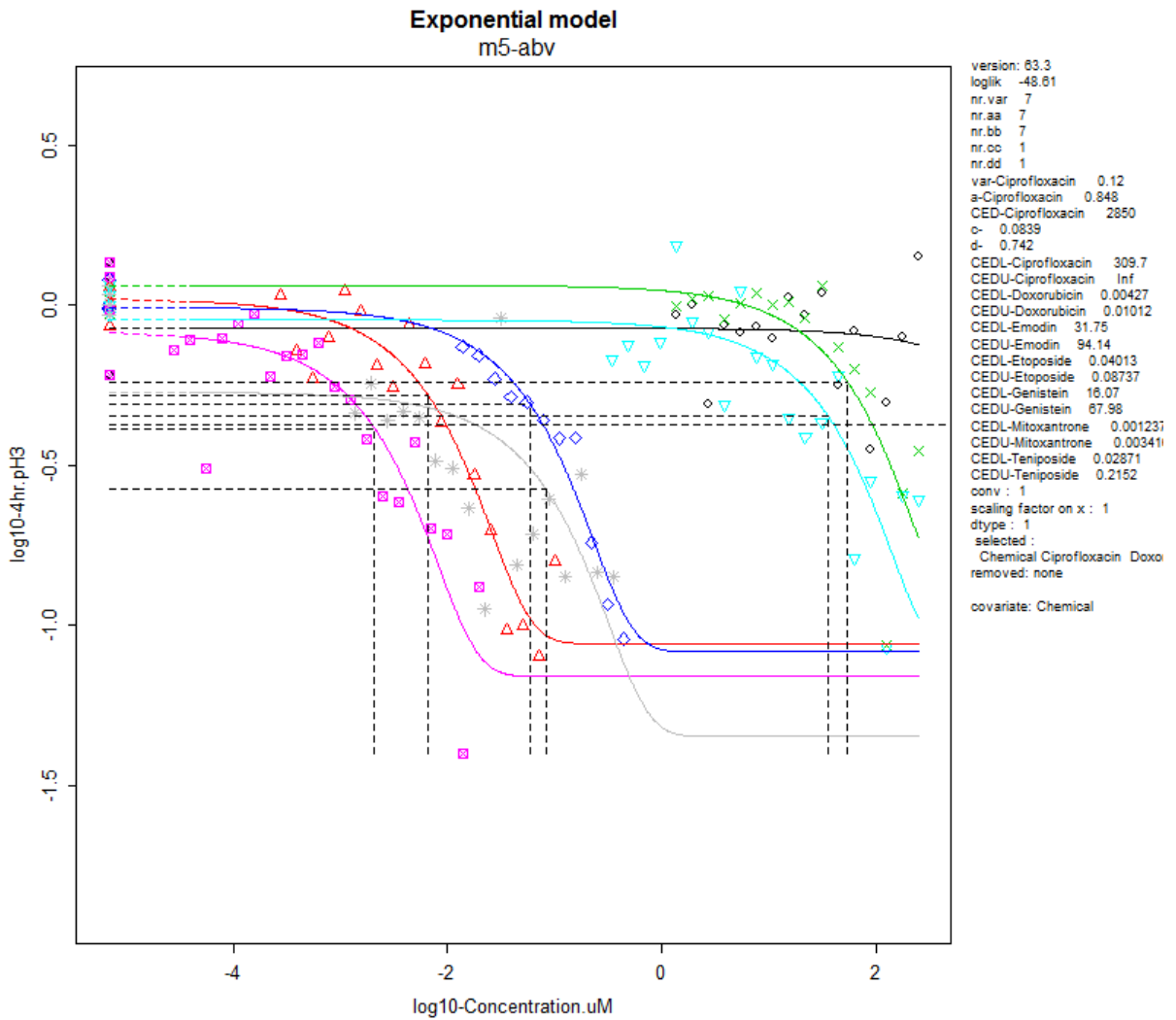
Figure A5.3. BMD dose-response model for the Topo-II Poisons exposure to TK6 cells 24-hour p-H3 biomarker response. Downregulation (decrease) of the p-H3 response is indicative of replication fork blockage and mitotic arrest from Topo-II Poison exposure.

Appendix 6. Topo-II Poison 4-hour BMD Confidence Intervals for MultiFlow Biomarker Responses
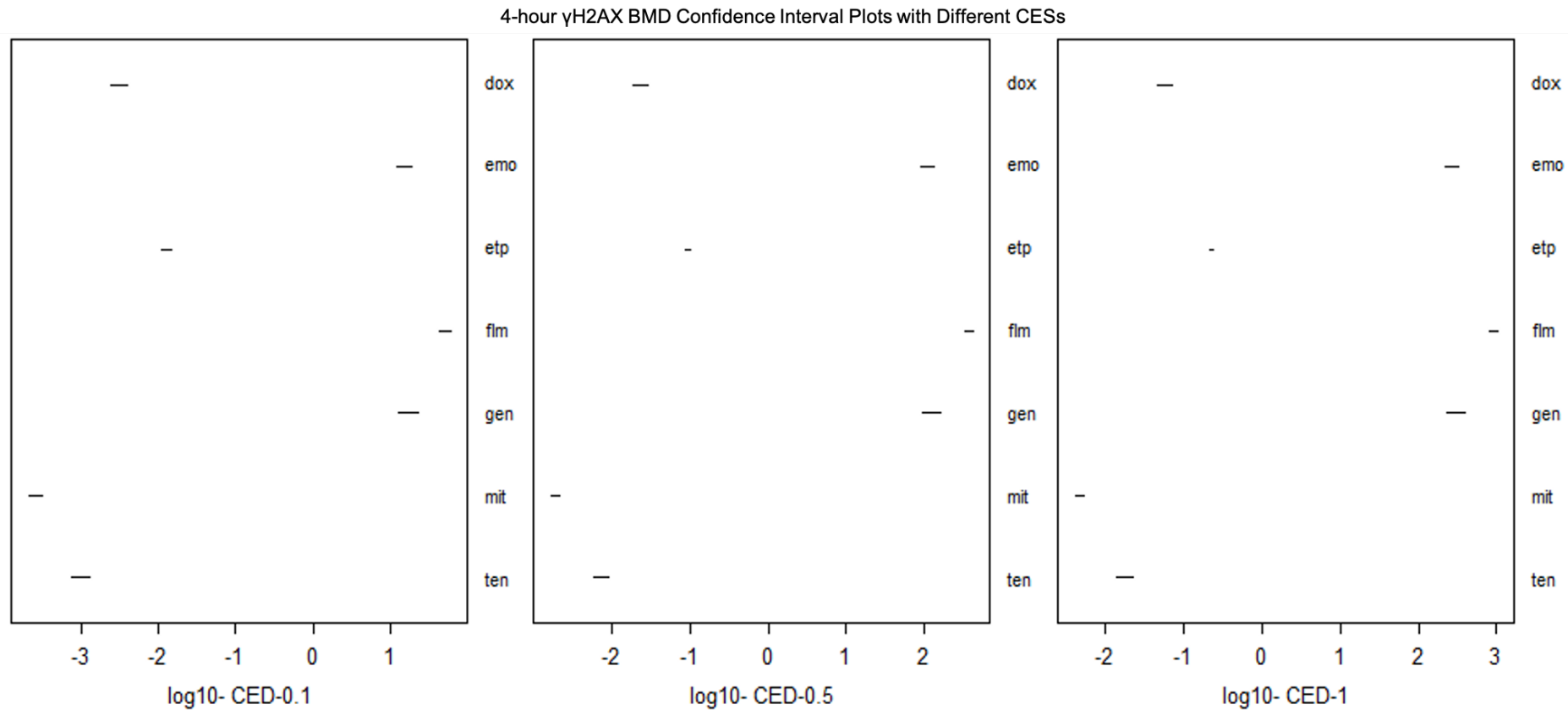
Figure A6.1. BMD Confidence Interval Plots resulting from different CES values. Panel left to right CES 0.1 (10%), 0.5 (50%) and 1.0 (100%) for the 4-hour γH2AX biomarker responses to Topo-II Poison compounds. Compound abbreviations on the right-hand side of each panel following the abbreviations in **Table 3.1** of the main text. The potency rank order does not differ between the three different CESs.
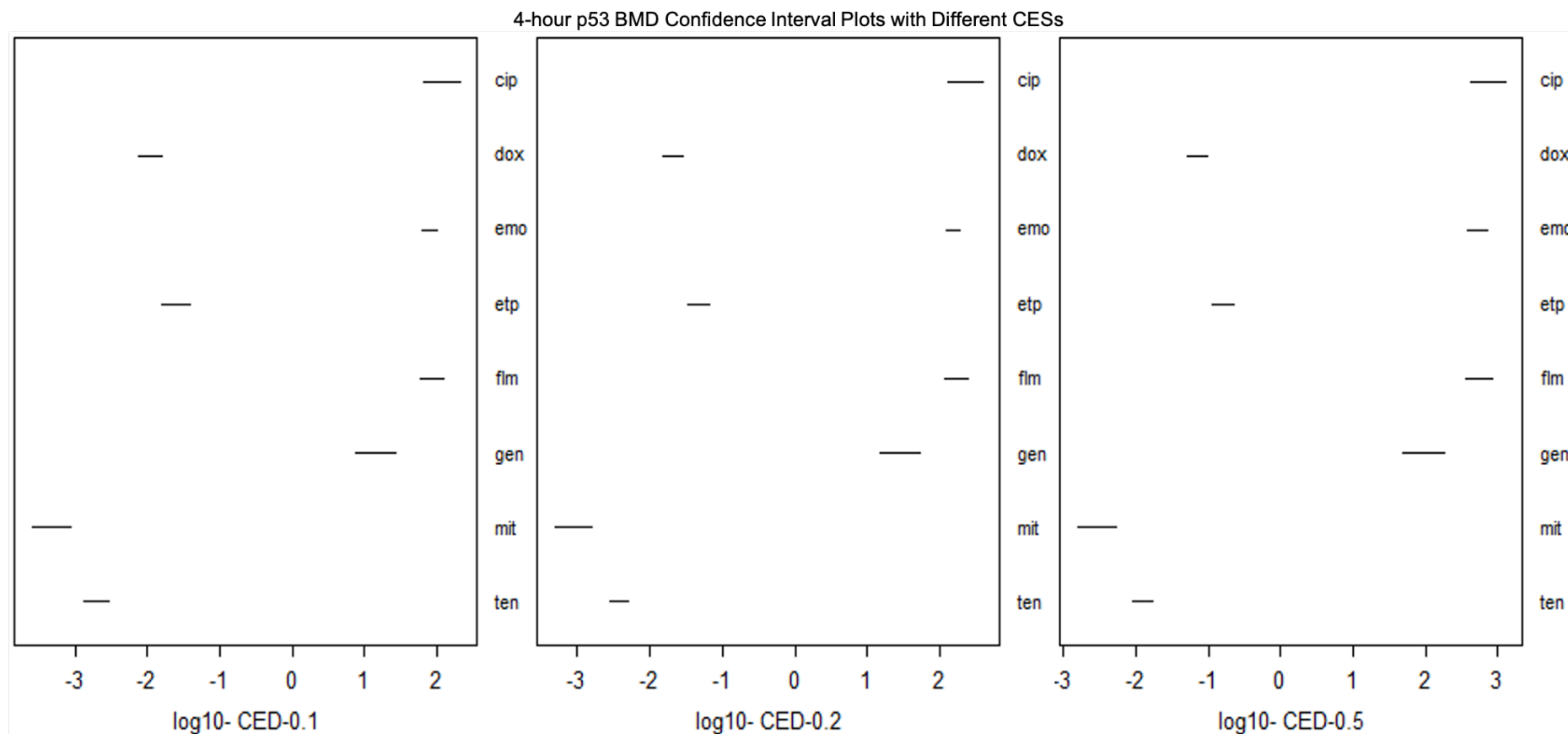
Figure A6.2. BMD Confidence Interval Plots resulting from different CES values. Panel left to right CES 0.1 (10%), 0.2 (20%) and 0.5 (50%) for the 4-hour p53 biomarker responses to Topo-II Poison compounds. Compound abbreviations on the right-hand side of each panel following the abbreviations in **Table 3.1** of the main text. The potency rank order does not differ between the three different CESs.
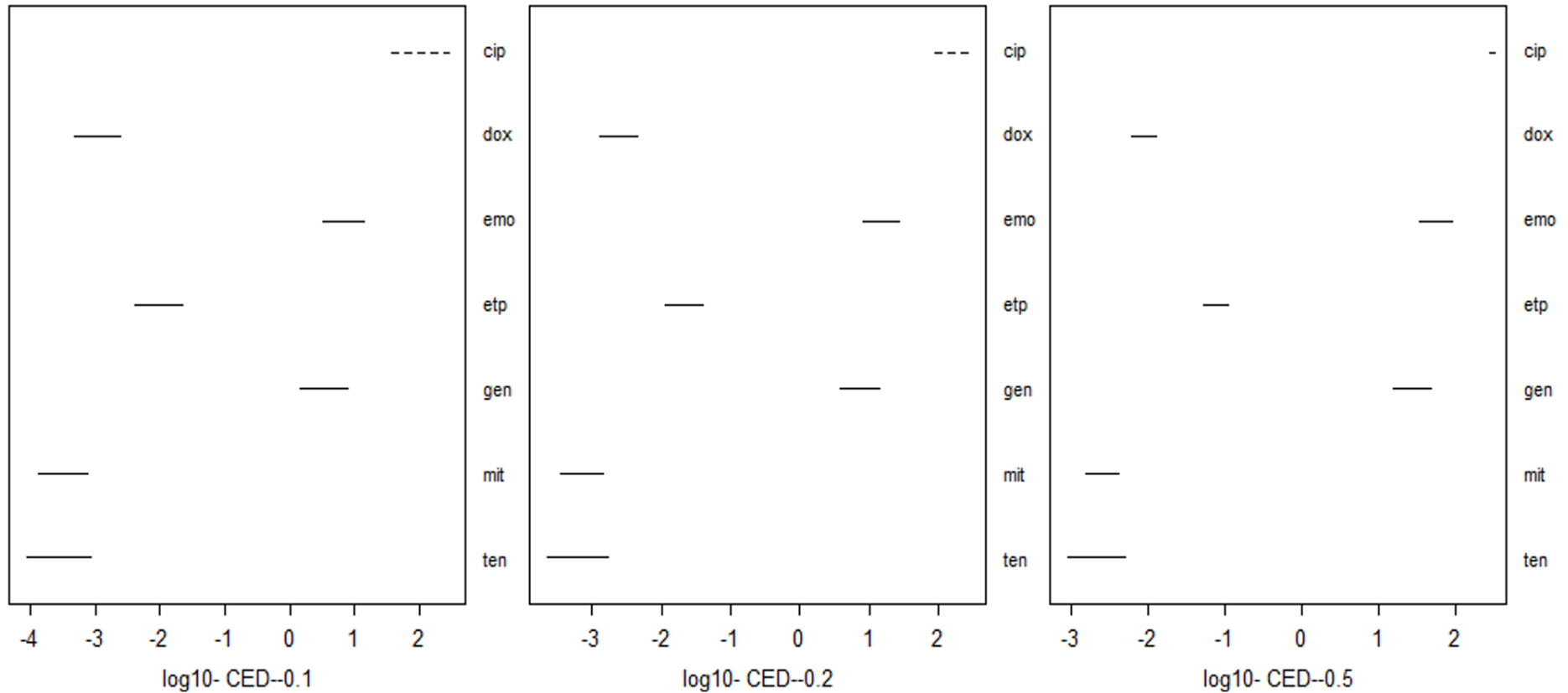
Figure A6.3. BMD Confidence Interval Plots resulting from different CES values. Panel left to right CES 0.1 (10%), 0.2 (20%) and 0.5 (50%) for the 4-hour p-H3 biomarker responses to Topo-II Poison compounds. Compound abbreviations on the right-hand side of each panel following the abbreviations in **Table 3.1** of the main text. The potency rank order does not differ between the three different CESs.

Appendix 7. Individual Dose-Response Curves for the MultiFlow and ToxTracker System Biomarker Responses Resulting from Topo-II Poison Exposure
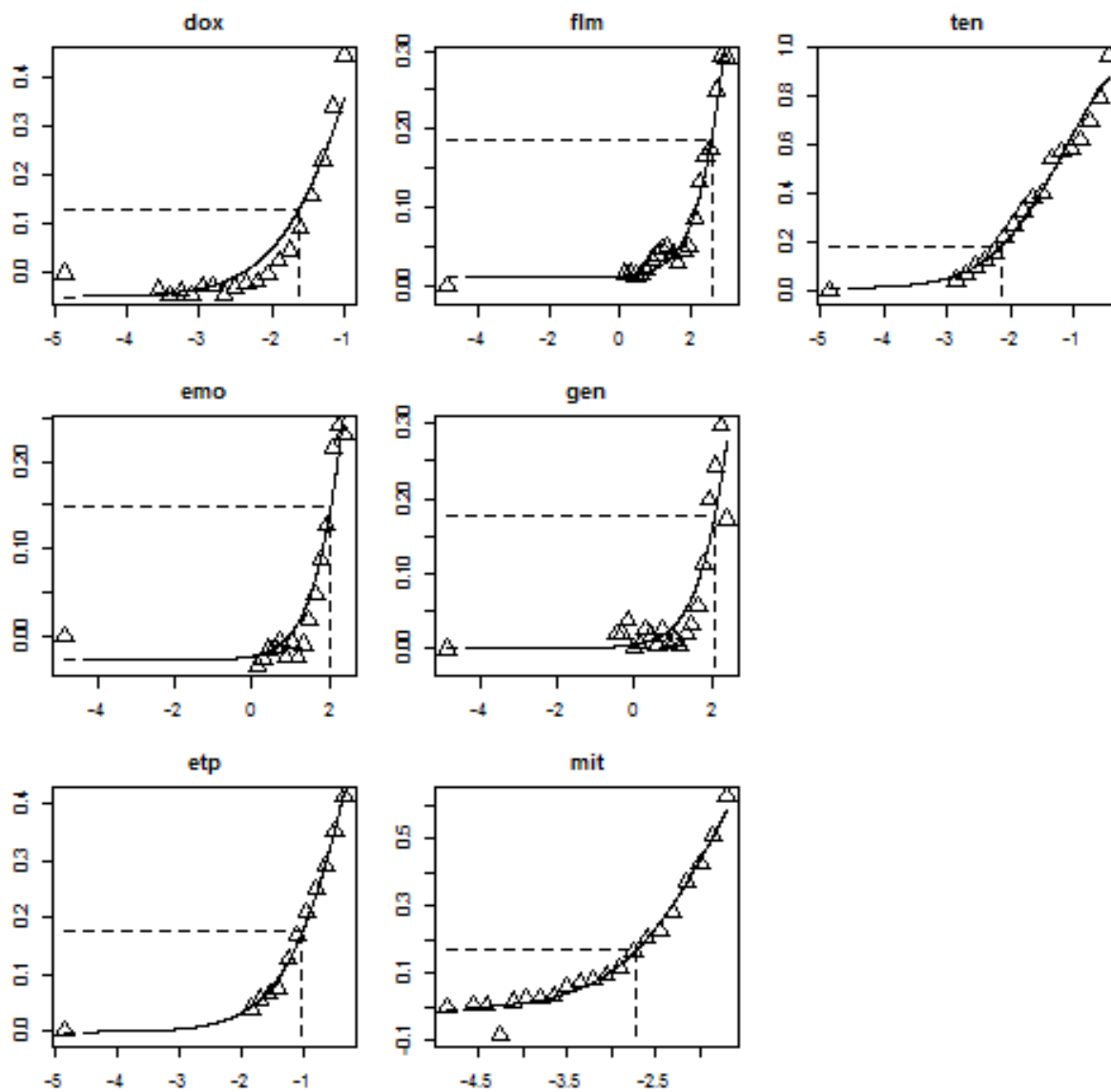
Figure A7.1. Individual dose-response curves for the Topo-II Poisons evaluated with the MultiFlow 4-hour γH2AX biomarker. CES 50%.
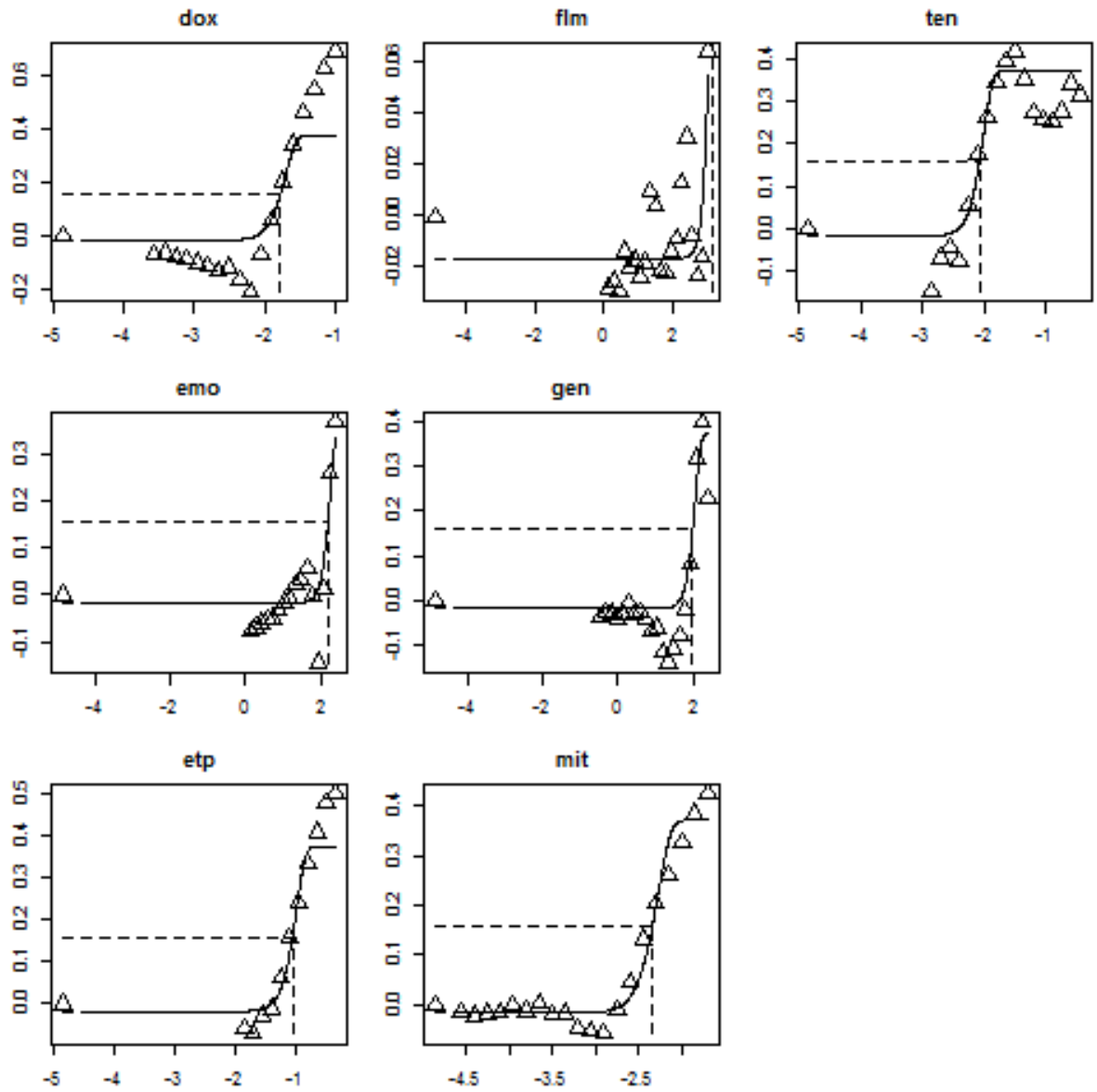
Figure A7.2. Individual dose-response curves for the Topo-II Poisons evaluated with the MultiFlow 24-hour γH2AX biomarker. CES 50%.
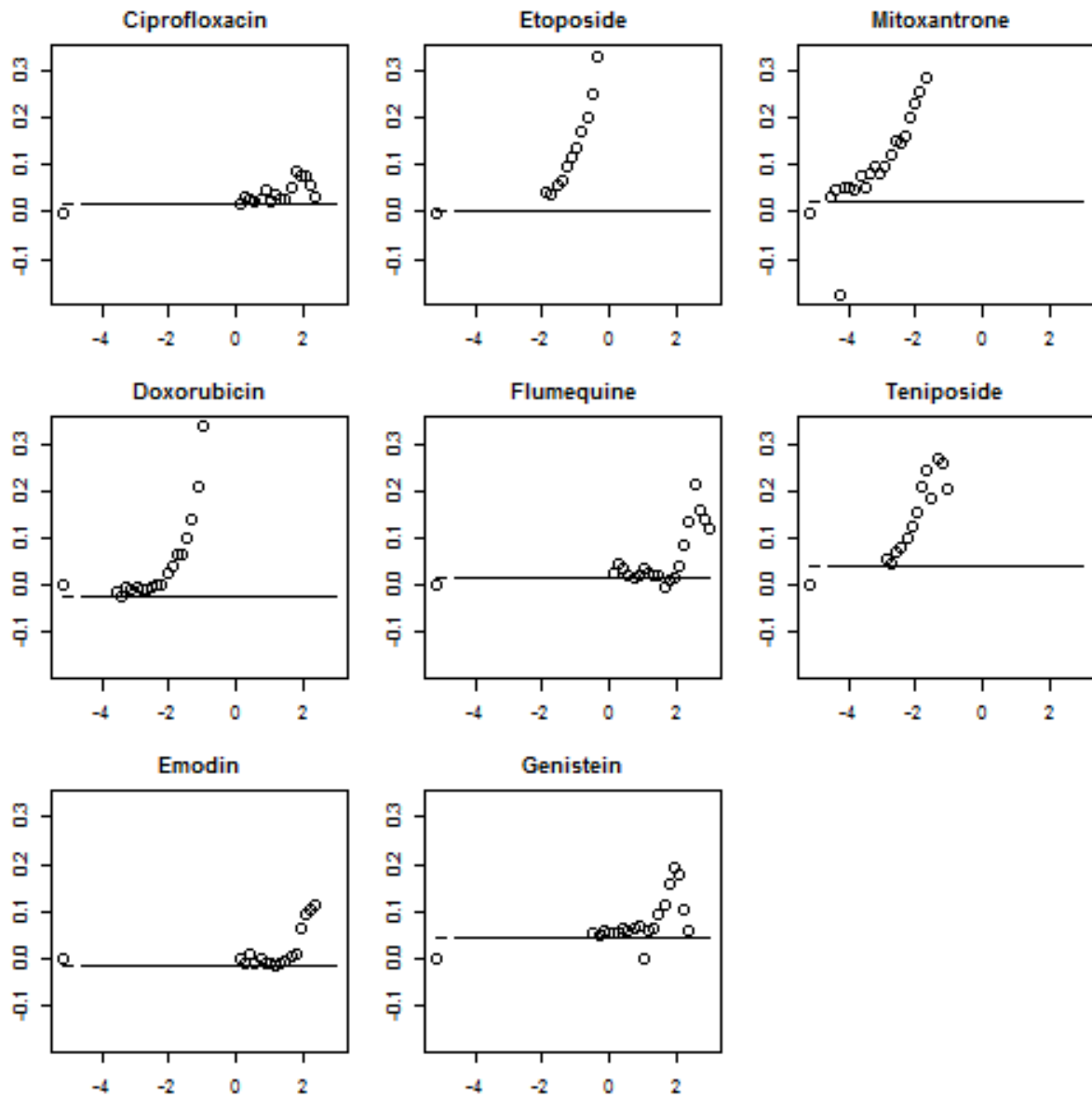
Figure A7.3. Individual dose-response curves for the Topo-II Poisons evaluated with the MultiFlow 4-hour p53 biomarker. CES 50%.
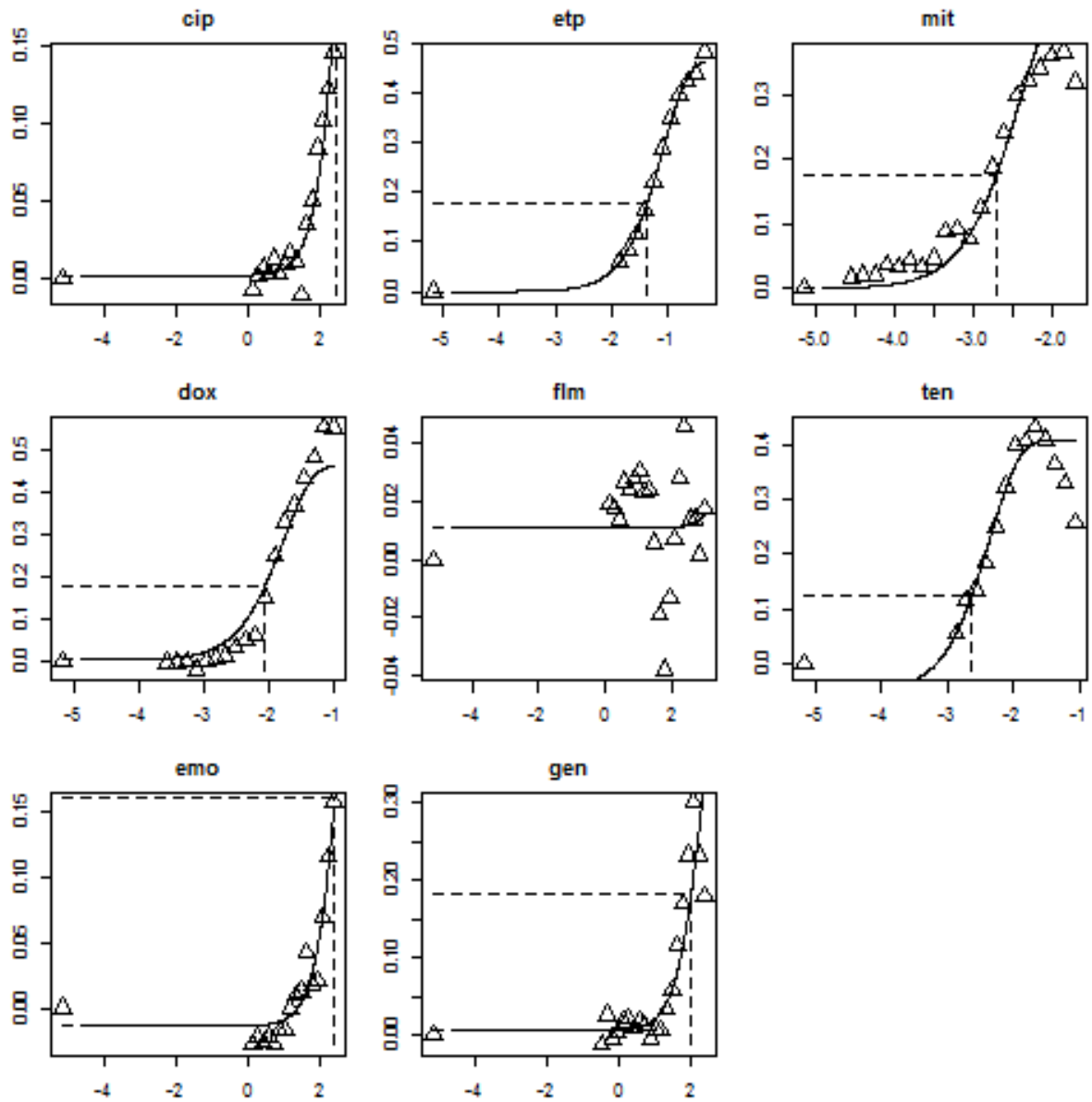
Figure A7.4. Individual dose-response curves for the Topo-II Poisons evaluated with the MultiFlow 24-hour p53 biomarker. CES 50%.
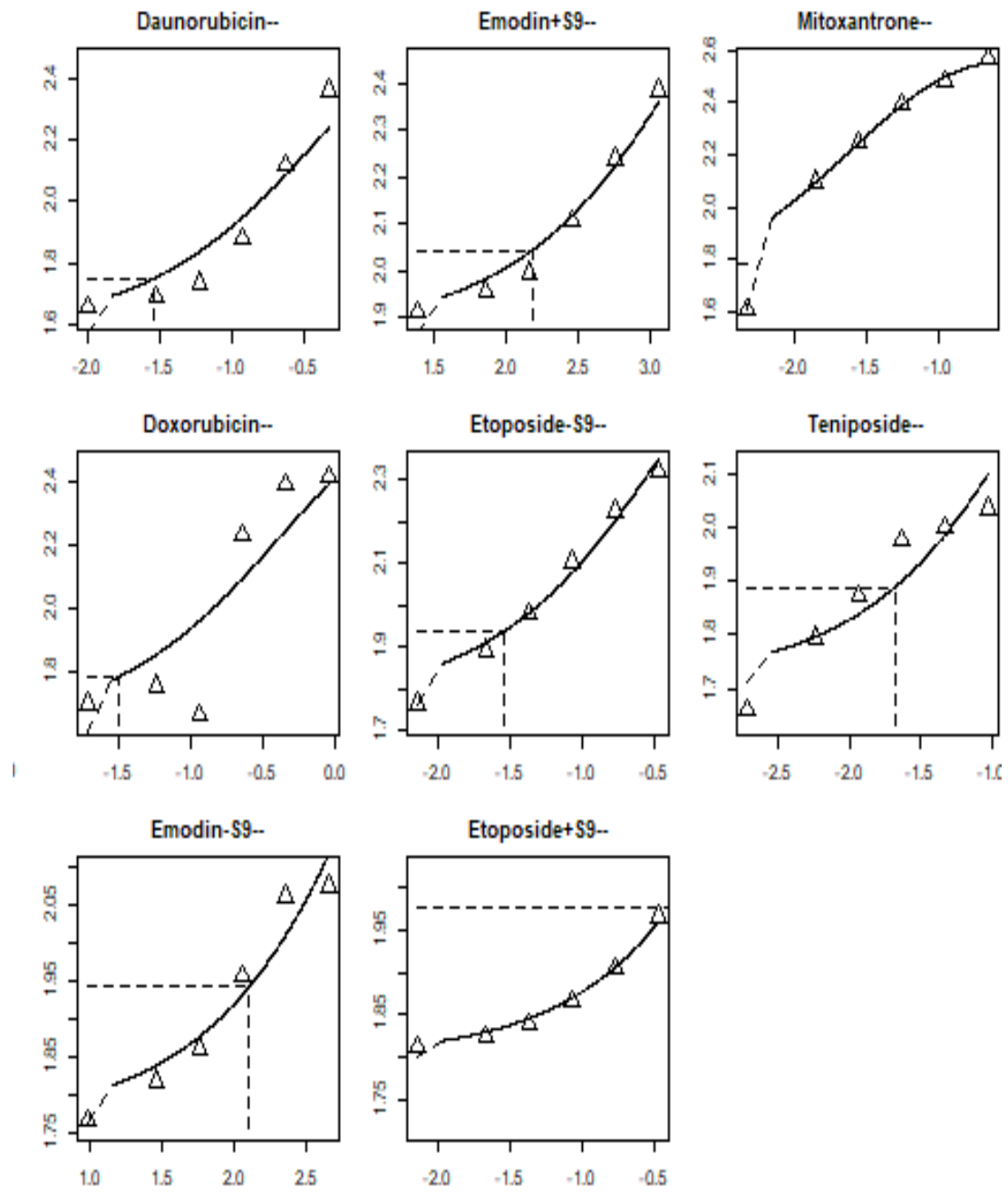
Figure A7.5. Individual dose-response curves for the Topo-II Poisons evaluated with the ToxTracker Bscl2 Biomarker. CES 50%.
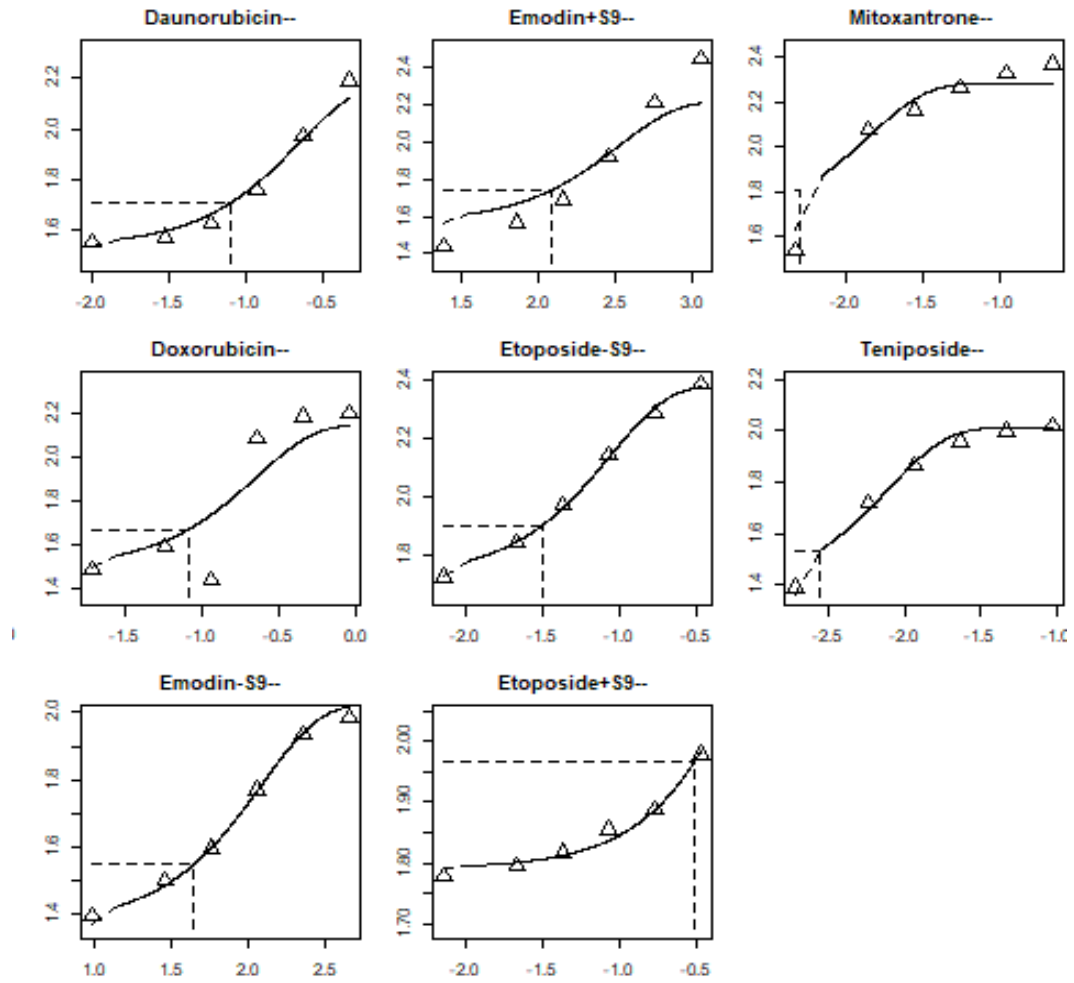
Figure A7.6. Individual dose-response curves for the Topo-II Poisons evaluated with the ToxTracker Btg2 Biomarker. CES 50%.
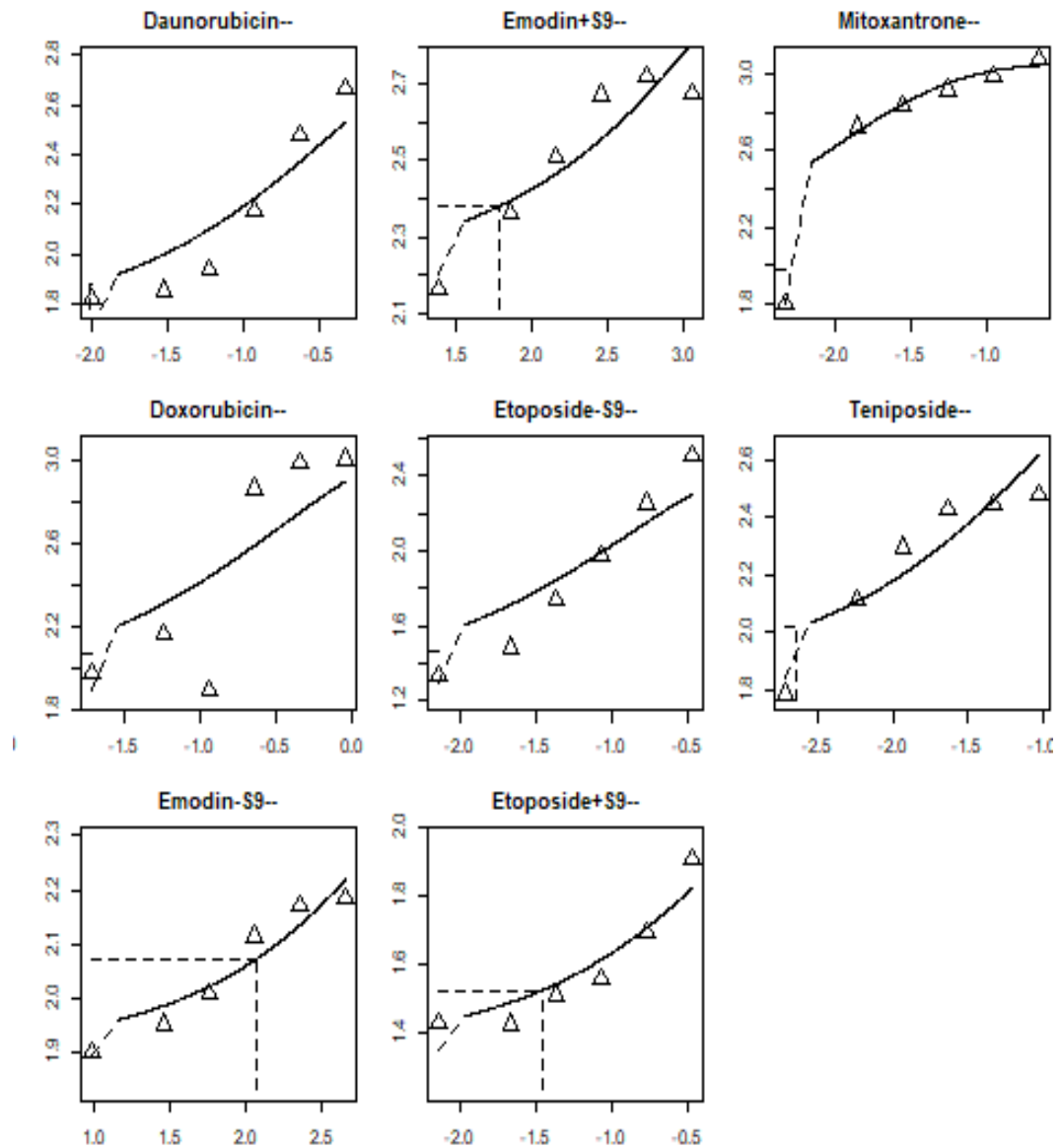
Figure A7.7. Individual dose-response curves for the Topo-II Poisons evaluated with the ToxTracker Btg2 Biomarker. CES 50%.

Appendix 8. Richard Williams Award Abstract

Can Benchmark Dose potency information augment the use of Structural Activity Relationships in Mode of Action-based compound risk assessments?

Ryan Wheeldon[1,2], George Johnson[1]

[1]Institute of Life Science, Swansea University Medical School, Swansea University, SA2 8PP, United Kingdom

[2]Baxter Healthcare, R&D Europe Sprl, Boulevard d'Angleterre 2-4 , 1420 Braine-l'Alleud, Belgium

Dearfield et al. (2017) presented an ambitious next generation testing strategy for the assessment of genomic damage aiming to transform genetic toxicology testing from hazard identification to a quantitative risk-based paradigm. Integral to the next generation approach are the use of Quantitative Structure-Activity Relationships (QSAR) and read-across; modern *in vitro* assays that can rapidly score cells for genomic damage; dose-response analysis of the resulting datasets; and Mode of Action (MOA) information that relates the mechanistic understanding of biological pathways to human health adverse outcomes. To this date, little has been published encompassing all elements of the next generation approach. To put the strategy into practice, we have performed Benchmark Dose (BMD) analyses of multiple *in vitro* multiplexed DNA damage response assay datasets. Firstly, we demonstrated that the response biomarkers p53, γH2AX and phospho-histone H3 (pH3) – for classes of clastogens and aneugens exposed to TK6 cells for 24 hours – mostly correlate with expected responses in the *in vitro* micronucleus gold standard assay for the same cell line. Secondly, to determine relative potency we applied BMD modelling to the same *in vitro* biomarkers' dose-responses resulting from the exposure to eight Topoisomerase II (Topo2) poisons. The patterns observed amongst the eight Topo2 compounds' potencies align with genotoxicity structural alerts of said compounds. We developed an Adverse Outcome Pathway (AOP) focussing on Key Events (KEs) that align with the *in vitro* DNA damage response biomarkers. Relevant Derek Nexus alerts which mention KEs on the Topo2 AOP were mapped to the AOP and had the benefit of connecting compounds to associated biological endpoints and adverse outcomes. Our work supports the next generation genotoxicity testing strategy, and we propose that BMD potency information can augment the use of *in silico* predictions in MOA-based compound risk assessment.

References:

Dearfield KL, Gollapudi BB, Bemis JC, Daniel Benz R, Douglas GR, Elespuru RK, Johnson GE, Kirkland DK, LeBaron MJ, Luijten M et al., 2017. 'Next generation testing strategy for assessment of genomic damage: A conceptual framework and considerations', Environmental and Molecular Mutagenesis, 58: 264-83.

Appendix 9. Topo-II Poison Derek Nexus Network Mapping by Lhasa

Lhasa preliminary network

Genotoxic_or_non_genotoxic	ID	isExcludedFromPaths	isInPath	LEV_ID	name	selected	shared name
U	312	FALSE	FALSE	Tissue	Malignant neoplasm	FALSE	Malignant neoplasm
G	182	FALSE	FALSE	Cell	Chromosome structural damage	FALSE	Chromosome structural damage
		FALSE	FALSE	Cell	Micronucleus formation	FALSE	Micronucleus formation
G	115	FALSE	FALSE	Molecular	DNA double-strand break (DSB)	FALSE	DNA double-strand break (DSB)
G	138	FALSE	FALSE	Molecular	DNA nucleobase reaction with electrophile	FALSE	DNA nucleobase reaction with electrophile
		FALSE	FALSE	Molecular	DNA damage repair	FALSE	DNA damage repair
G	117	FALSE	FALSE	Cell	Inherited DNA mutation	FALSE	Inherited DNA mutation
U	274	FALSE	FALSE	Molecular	Exogenous oxidant increase	FALSE	Exogenous oxidant increase
G	267	FALSE	FALSE	Cell	Mitotic dysregulation	FALSE	Mitotic dysregulation
G	191	FALSE	FALSE	Molecular	Non-covalent interaction with DNA	FALSE	Non-covalent interaction with DNA
G	139	FALSE	FALSE	Molecular	Protein reaction with electrophile	FALSE	Protein reaction with electrophile
G	133	FALSE	FALSE	Molecular	Radical species generation	FALSE	Radical species generation
G	120	FALSE	FALSE	Molecular	Topoisomerase II poisoning	FALSE	Topoisomerase II poisoning
G	141	FALSE	FALSE	Molecular	Tubulin binding	FALSE	Tubulin binding
G	63	FALSE	FALSE	Cell	Uncoupling of oxidative phosphorylation	FALSE	Uncoupling of oxidative phosphorylation
		FALSE	FALSE	Compound	23214-92-8	FALSE	23214-92-8
		FALSE	FALSE	Organism	Mitochondrial dysfunction	FALSE	Mitochondrial dysfunction
		FALSE	FALSE	Compound	29767-20-2	FALSE	29767-20-2
		FALSE	FALSE	Compound	33419-42-0	FALSE	33419-42-0
		FALSE	FALSE	Compound	42835-25-6	FALSE	42835-25-6
		FALSE	FALSE	Molecular	Gyrase inhibition	FALSE	Gyrase inhibition
		FALSE	FALSE	Compound	446-72-0	FALSE	446-72-0
		FALSE	FALSE	Molecular	Estrogen receptor (ER) binding	FALSE	Estrogen receptor (ER) binding
		FALSE	FALSE	Compound	518-82-1	FALSE	518-82-1
		FALSE	FALSE	Compound	65271-80-9	FALSE	65271-80-9
		FALSE	FALSE	Compound	85721-33-1	FALSE	85721-33-1
		FALSE	FALSE	Assay	Ames Test	FALSE	Ames Test
		FALSE	FALSE	Assay	In Vitro Chromosome Aberration Test	FALSE	In Vitro Chromosome Aberration Test
		FALSE	FALSE	Assay	In Vitro Comet Assay	FALSE	In Vitro Comet Assay
		FALSE	FALSE	Assay	In Vitro Hgprt Assay	FALSE	In Vitro Hgprt Assay
		FALSE	FALSE	Assay	In Vitro Micronucleus Test	FALSE	In Vitro Micronucleus Test
		FALSE	FALSE	Assay	In Vitro Uds Assay	FALSE	In Vitro Uds Assay
		FALSE	FALSE	Assay	In Vivo Chromosome Aberration Test	FALSE	In Vivo Chromosome Aberration Test
		FALSE	FALSE	Assay	In Vivo Micronucleus Test	FALSE	In Vivo Micronucleus Test
		FALSE	FALSE	Assay	In Vivo Uds Assay	FALSE	In Vivo Uds Assay
		FALSE	FALSE	Assay	Mouse Lymphoma Assay	FALSE	Mouse Lymphoma Assay
		FALSE	FALSE	Assay	Rodent Carcinogenicity Study	FALSE	Rodent Carcinogenicity Study
		FALSE	FALSE	Assay	Transgenic Rodent Study	FALSE	Transgenic Rodent Study
		FALSE	FALSE	Assay	gamma-H2AX	FALSE	gamma-H2AX
		FALSE	FALSE	Assay	p53	FALSE	p53
		FALSE	FALSE	Cell	Tumour suppressor protein p53 signalling pathway dysregulation	FALSE	Tumour suppressor protein p53 signalling pathway dysregulation
		FALSE	FALSE	Assay	phosphor-histone H3	FALSE	phosphor-histone H3

Appendix 10. Complex Topo-Poison Derek Nexus Mapped AOP Network
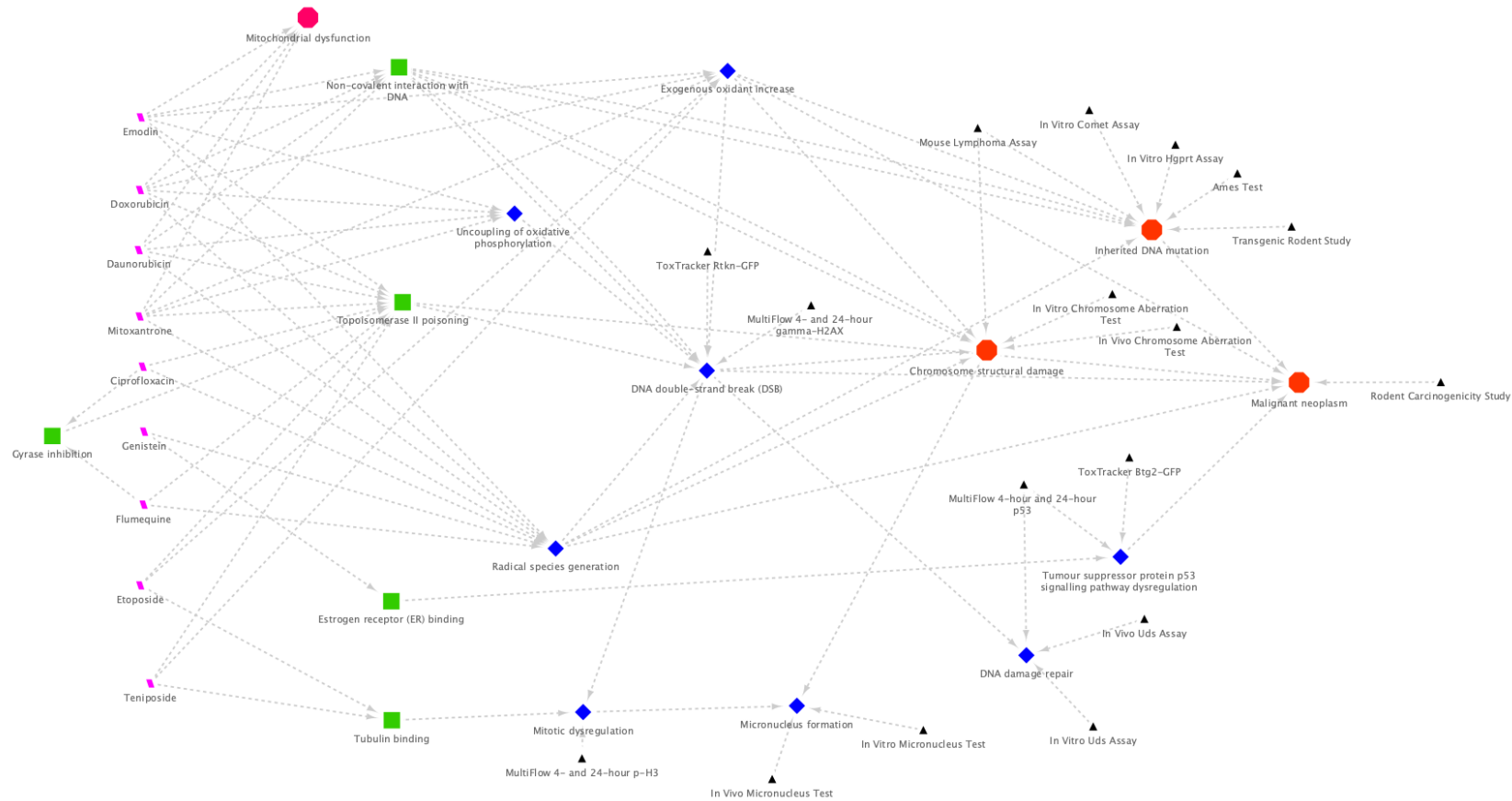
Figure A10.1. Complexity of the Topo-II Derek Nexus mapped AOP network is shown with the "uncoupling of oxidative phosphorylation", "radical species generation", and "exogenous oxidant increase". These three KEs were combined into a single KE named "ROS generation" and mapped to the "Topoisomerase II Poisoning" and "DNA double-strand break (DSB)" KEs **Figure 6.4**.

Glossary

Adverse Outcome Pathway- A model sequence of molecular and cellular events required to produce a toxic effect when an organism is exposed to a substance.

Aneugen- A substance that causes daughter cells to have an abnormal number of chromosomes or aneuploidy.

Assay- A structured laboratory analytical method.

BMD- Benchmark Dose. A dose or concentration that produces a predetermined change in the response rate of an adverse effect.

Clastogen- A mutagenic agent that can cause breaks in chromosomes leading to deletions, additions, or rearrangements of the genetic material.

Confidence Interval- A range of values representing the range in which the true value lies.

DNA Damage- A reversible change in the basic structure of DNA through chemical or physical influences.

Dose-response relationship- A description of the magnitude of the response of an organism, as a function of exposure to a stressor/stimulus after a defined exposure time.

Exponential function- A function of the form $f(x) = ab^x$ where b is a positive number.

H2AX- H2A histone family member X. An important type of protein which forms when double-strand breaks appear.

p53- Tumour protein 53 (Guardian of the Genome). A protein involved in the key responses to genotoxic stress, including cell cycle regulation, apoptosis, and genomic stability.

Phosphorylated histone H3- Histone H3 is heavily phosphorylated during the metaphase of cell cycle division.

Quantitative data type- Value of data in the form of counts, quantity, amount, or range.

S9 fraction- The product of rat liver homogenate used in *in vitro* biological test systems to stimulate metabolic activation of certain compounds.

References

AARDEMA, M. J. 2013. The holy grail in genetic toxicology: Follow-up approaches for positive results in the Ames assay. *Environmental and Molecular Mutagenesis,* 54**,** 617-620.

ABRAHAM, R. T. 2001. Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes & development,* 15**,** 2177-2196.

ADAN, A., ALIZADA, G., KIRAZ, Y., BARAN, Y. & NALBANT, A. 2017. Flow cytometry: basic principles and applications. *Critical Reviews in Biotechnology,* 37**,** 163-176.

ALDRED, K. J., BLOWER, T. R., KERNS, R. J., BERGER, J. M. & OSHEROFF, N. 2016. Fluoroquinolone interactions with Mycobacterium tuberculosis gyrase: Enhancing drug activity against wild-type and resistant gyrase. *Proceedings of the National Academy of Sciences,* 113**,** E839-E846.

ALDRED, K. J., BRELAND, E. J., MCPHERSON, S. A., TURNBOUGH JR, C. L., KERNS, R. J. & OSHEROFF, N. 2014a. Bacillus anthracis GrlAV96A topoisomerase IV, a quinolone resistance mutation that does not affect the water-metal ion bridge. *Antimicrobial agents and chemotherapy,* 58**,** 7182-7187.

ALDRED, K. J., BRELAND, E. J., VLČKOVÁ, V., STRUB, M.-P., NEUMAN, K. C., KERNS, R. J. & OSHEROFF, N. 2014b. Role of the water–metal ion bridge in mediating interactions between quinolones and Escherichia coli topoisomerase IV. *Biochemistry,* 53**,** 5558-5567.

ALDRED, K. J., MCPHERSON, S. A., TURNBOUGH, C. L., JR, KERNS, R. J. & OSHEROFF, N. 2013. Topoisomerase IV-quinolone interactions are mediated through a water-metal ion bridge: mechanistic basis of quinolone resistance. *Nucleic Acids Research,* 41**,** 4628-4639.

ALLEMANG, A., MAHONY, C., LESTER, C. & PFUHLER, S. 2018. Relative potency of fifteen pyrrolizidine alkaloids to induce DNA damage as measured by micronucleus induction in HepaRG human liver cells. *Food Chem Toxicol,* 121**,** 72-81.

ALTMAN, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician,* 46**,** 175-185.

AMBERG, A., BEILKE, L., BERCU, J., BOWER, D., BRIGO, A., CROSS, K. P., CUSTER, L., DOBO, K., DOWDY, E., FORD, K. A., GLOWIENKE, S., VAN GOMPEL, J., HARVEY, J., HASSELGREN, C., HONMA, M., JOLLY, R., KEMPER, R., KENYON, M., KRUHLAK, N., LEAVITT, P., MILLER, S., MUSTER, W., NICOLETTE, J., PLAPER, A., POWLEY, M., QUIGLEY, D. P., REDDY, M. V., SPIRKL, H.-P., STAVITSKAYA, L., TEASDALE, A., WEINER, S., WELCH, D. S., WHITE, A., WICHARD, J. & MYATT, G. J. 2016. Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses. *Regulatory Toxicology and Pharmacology,* 77**,** 13-24.

AMES, B., DURSTON, W., YAMASAKI, E. & LEE, F. 1973. Carcinogens are Mutagens: A Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection. *Proceedings of the National Academy of Sciences of the United States of America,* 70**,** 2281.

ANDROUTSOS, G. 2006. The outstanding British surgeon Percivall Pott (1714-1789) and the first description of an occupational cancer. *Journal of B.U.ON. : official journal of the Balkan Union of Oncology,* 11**,** 533-539.

APLAN, P., CHERVINSKY, D., STANULLA, M. & BURHANS, W. 1996. Site-specific DNA cleavage within the MLL breakpoint cluster region induced by topoisomerase II inhibitors [see comments]. *Blood,* 87**,** 2649-2658.

AU, N., CHEANG, M., HUNTSMAN, D., YORIDA, E., COLDMAN, A., ELLIOTT, W., BEBB, G., FLINT, J., ENGLISH, J., GILKS, C. & GRIMES, H. 2004. Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers. *The Journal of Pathology,* 204**,** 101-109.

AVLASEVICH, S., BRYCE, S., DE BOECK, M., ELHAJOUJI, A., VAN GOETHEM, F., LYNCH, A., NICOLETTE, J., SHI, J. & DERTINGER, S. 2011. Flow cytometric analysis of micronuclei in mammalian cell cultures: past, present and future. *Mutagenesis,* 26**,** 147-52.

AVLASEVICH, S. L., BRYCE, S. M., CAIRNS, S. E. & DERTINGER, S. D. 2006. In vitro micronucleus scoring by flow cytometry: differential staining of micronuclei versus apoptotic and necrotic chromatin enhances assay reliability. *Environ Mol Mutagen,* 47**,** 56-66.

AWASTHI, P., VATSAL, M. & SHARMA, A. 2019. Structural and biological study of synthesized anthraquinone series of compounds with sulfonamide feature. *Journal of Biomolecular Structure and Dynamics,* 37**,** 4465-4480.

BAILEY, S. T., SHIN, H., WESTERLING, T., LIU, X. S. & BROWN, M. 2012. Estrogen receptor prevents p53-dependent apoptosis in breast cancer. *Proceedings of the National Academy of Sciences,* 109**,** 18060-18065.

BALDWIN, E. & OSHEROFF, N. 2005. Etoposide, topoisomerase II and cancer. *Current Medicinal Chemistry-Anti-Cancer Agents,* 5**,** 363-372.

BALL, N., CRONIN, M. T. D., SHEN, J., BLACKBURN, K., BOOTH, E. D., BOUHIFD, M., DONLEY, E., EGNASH, L., HASTINGS, C., JUBERG, D. R., KLEENSANG, A., KLEINSTREUER, N., KROESE, E. D., LEE, A. C., LUECHTEFELD, T., MAERTENS, A., MARTY, S., NACIFF, J. M., PALMER, J., PAMIES, D., PENMAN, M., RICHARZ, A.-N., RUSSO, D. P., STUARD, S. B., PATLEWICZ, G., VAN RAVENZWAAY, B., WU, S., ZHU, H. & HARTUNG, T. 2016. Toward Good Read-Across Practice (GRAP) guidance. *ALTEX,* 33**,** 149-166.

BALL, T., BARBER, C. G., CAYLEY, A., CHILTON, M. L., FOSTER, R., FOWKES, A., HEGHES, C., HILL, E., HILL, N., KANE, S., MACMILLAN, D. S., MYDEN, A., NEWMAN, D., POLIT, A., STALFORD, S. A. & VESSEY, J. D. 2021. Beyond adverse outcome pathways: making toxicity predictions from event networks, SAR models, data and knowledge. *Toxicology Research,* 10**,** 102-122.

BARBER, C., HANSER, T., JUDSON, P. & WILLIAMS, R. 2017. Distinguishing between expert and statistical systems for application under ICH M7. *Regulatory Toxicology and Pharmacology,* 84**,** 124-130.

BAVETSIAS, V. & LINARDOPOULOS, S. 2015. Aurora Kinase Inhibitors: Current Status and Outlook. *Frontiers in Oncology,* 5.

BEMIS, J., WILLS, J., BRYCE, S., TOROUS, D., DERTINGER, S. & SLOB, W. 2016. Comparison of in vitro and in vivo clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. *Mutagenesis,* 31**,** 277-285.

BENIGNI, R. 2008. The Benigni/Bossa rulebase for mutagenicity and carcinogenicity–a module of Toxtree. *JRC Scientific and Technical Reports,* 1**,** 63.

BENIGNI, R. & BOSSA, C. 2018. Data-based review of QSARs for predicting genotoxicity: the state of the art. *Mutagenesis,* 34**,** 17-23.

BENIGNI, R., BOSSA, C., JELIAZKOVA, N., NETZEVA, T. & WORTH, A. 2008. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. European Commission Joint Research Centre.

BERANEK, D. T. 1990. Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 231**,** 11-30.

BERGER, J. M. 1998. Structure of DNA topoisomerases. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1400**,** 3-18.

BERNACKI, D. T., BRYCE, S. M., BEMIS, J. C., KIRKLAND, D. & DERTINGER, S. D. 2016. γH2AX and p53 responses in TK6 cells discriminate promutagens and nongenotoxicants in the presence of rat liver S9. *Environ Mol Mutagen,* 57**,** 546-558.

BOEHDEN, G. S., BAUMANN, C., SIEHLER, S. & WIESMÜLLER, L. 2005. Wild-type p53 stimulates homologous recombination upon sequence-specific binding to the ribosomal gene cluster repeat. *Oncogene,* 24**,** 4183-4192.

BOOBIS, A., BROWN, P., CRONIN, M. T. D., EDWARDS, J., GALLI, C. L., GOODMAN, J., JACOBS, A., KIRKLAND, D., LUIJTEN, M., MARSAUX, C., MARTIN, M., YANG, C. & HOLLNAGEL, H. M. 2017. Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation. *Critical Reviews in Toxicology,* 47**,** 710-732.

BOOBIS, A. R., COHEN, S. M., DELLARCO, V., MCGREGOR, D., MEEK, M. E., VICKERS, C., WILLCOCKS, D. & FARLAND, W. 2006. IPCS Framework for Analyzing the Relevance of a Cancer Mode of Action for Humans. *Critical Reviews in Toxicology,* 36**,** 781-792.

BOOBIS, A. R., DOE, J. E., HEINRICH-HIRSCH, B., MEEK, M. E., MUNN, S., RUCHIRAWAT, M., SCHLATTER, J., SEED, J. & VICKERS, C. 2008. IPCS Framework for Analyzing the Relevance of a Noncancer Mode of Action for Humans. *Critical Reviews in Toxicology,* 38**,** 87-96.

BOS, P. M., BAARS, B. J. & VAN RAAIJ, M. T. 2004. Risk assessment of peak exposure to genotoxic carcinogens: a pragmatic approach. *Toxicol Lett,* 151**,** 43-50.

BRYCE, S. M., AVLASEVICH, S. L., BEMIS, J. C., LUKAMOWICZ, M., ELHAJOUJI, A., VAN GOETHEM, F., DE BOECK, M., BEERENS, D., AERTS, H., VAN GOMPEL, J., COLLINS, J. E., ELLIS, P. C., WHITE, A. T., LYNCH, A. M. & DERTINGER, S. D. 2008. Interlaboratory evaluation of a flow cytometric, high content in vitro micronucleus assay. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 650**,** 181-195.

BRYCE, S. M., AVLASEVICH, S. L., BEMIS, J. C., PHONETHEPSWATH, S. & DERTINGER, S. D. 2010. Miniaturized flow cytometric in vitro micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose–response relationships. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 703**,** 191-199.

BRYCE, S. M., BEMIS, J. C., AVLASEVICH, S. L. & DERTINGER, S. D. 2007. In vitro micronucleus assay scored by flow cytometry provides a comprehensive evaluation of cytogenetic damage and cytotoxicity. *Mutat Res,* 630**,** 78-91.

BRYCE, S. M., BEMIS, J. C., MERENESS, J. A., SPELLMAN, R. A., MOSS, J., DICKINSON, D., SCHULER, M. J. & DERTINGER, S. D. 2014. Interpreting in vitro micronucleus positive results: simple biomarker matrix discriminates clastogens, aneugens, and misleading positive agents. *Environ Mol Mutagen,* 55**,** 542-55.

BRYCE, S. M., BERNACKI, D. T., BEMIS, J. C. & DERTINGER, S. D. 2016. Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. *Environ Mol Mutagen,* 57**,** 171-89.

BRYCE, S. M., BERNACKI, D. T., BEMIS, J. C., SPELLMAN, R. A., ENGEL, M. E., SCHULER, M., LORGE, E., HEIKKINEN, P. T., HEMMANN, U., THYBAUD, V., WILDE, S., QUEISSER, N., SUTTER, A., ZELLER, A., GUÉRARD, M., KIRKLAND, D. & DERTINGER, S. D. 2017. Interlaboratory evaluation of a multiplexed high information content in vitro genotoxicity assay. *Environ Mol Mutagen,* 58**,** 146-161.

BRYCE, S. M., BERNACKI, D. T., SMITH-ROE, S. L., WITT, K. L., BEMIS, J. C. & DERTINGER, S. D. 2018. Investigating the Generalizability of the MultiFlow ® DNA Damage Assay and

Several Companion Machine Learning Models With a Set of 103 Diverse Test Chemicals. *Toxicol Sci,* 162**,** 146-166.

BUENO, C., CATALINA, P., MELEN, G. J., MONTES, R., SÁNCHEZ, L., LIGERO, G., GARCÍA-PÉREZ, J. L. & MENENDEZ, P. 2009. Etoposide induces MLL rearrangements and other chromosomal abnormalities in human embryonic stem cells. *Carcinogenesis,* 30**,** 1628-1637.

BURDEN, D. A. & OSHEROFF, N. 1998. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1400**,** 139-154.

CALABRESE, E. J. 2005. Historical blunders: how toxicology got the dose-response relationship half right. *Cell Mol Biol (Noisy-le-grand),* 51**,** 643-54.

CALABRESE, E. J. 2019. The linear no-threshold (LNT) dose response model: a comprehensive assessment of its historical and scientific foundations. *Chemico-biological interactions,* 301**,** 6-25.

CALABRESE, E. J. & BALDWIN, L. A. 2001. The Frequency of U-Shaped Dose Responses in the Toxicological Literature. *Toxicological Sciences,* 62**,** 330-338.

CALABRESE, E. J. & BLAIN, R. 2005. The occurrence of hormetic dose responses in the toxicological literature, the hormesis database: an overview. *Toxicology and Applied Pharmacology,* 202**,** 289-301.

CAPPELLINI, F., HEDBERG, Y., MCCARRICK, S., HEDBERG, J., DERR, R., HENDRIKS, G., ODNEVALL WALLINDER, I. & KARLSSON, H. L. 2018. Mechanistic insight into reactivity and (geno)toxicity of well-characterized nanoparticles of cobalt metal and oxides. *Nanotoxicology,* 12**,** 602-620.

CÁRDENAS, G. & NOGUEIRA, J. J. 2020. Stacking effects on Anthraquinone/DNA charge-transfer electronically excited states. *Molecules,* 25**,** 5927.

CARPINELLI, P. & MOLL, J. 2008. Aurora Kinases and Their Inhibitors: More Than One Target and One Drug. *In:* COLOTTA, F. & MANTOVANI, A. (eds.) *Targeted Therapies in Cancer: Myth or Reality?* New York, NY: Springer New York.

CHANDRASEKARAN, B., ABED, S. N., AL-ATTRAQCHI, O., KUCHE, K. & TEKADE, R. K. 2018. Chapter 21 - Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties. *In:* TEKADE, R. K. (ed.) *Dosage Form Design Parameters.* Academic Press.

CHILTON, M. L., MACMILLAN, D. S., STEGER-HARTMANN, T., HILLEGASS, J., BELLION, P., VUORINEN, A., ETTER, S., SMITH, B. P. C., WHITE, A., STERCHELE, P., DE SMEDT, A., GLOGOVAC, M., GLOWIENKE, S., O'BRIEN, D. & PARAKHIA, R. 2018. Making reliable negative predictions of human skin sensitisation using an in silico fragmentation approach. *Regulatory Toxicology and Pharmacology,* 95**,** 227-235.

CHOW, K.-C., MACDONALD, T. L. & ROSS, W. E. 1988. DNA binding by epipodophyllotoxins and N-acyl anthracyclines: implications for mechanism of topoisomerase II inhibition. *Molecular Pharmacology,* 34**,** 467-473.

CHUNG, H. W., KANG, S. J. & KIM, S. Y. 2002. A combination of the micronucleus assay and a FISH technique for evaluation of the genotoxicity of 1,2,4-benzenetriol. *Mutat Res,* 516**,** 49-56.

CIMINO, M. C. 2006. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. *Environ Mol Mutagen,* 47**,** 362-90.

CLAYTON, A. L., ROSE, S., BARRATT, M. J. & MAHADEVAN, L. C. 2000. Phosphoacetylation of histone H3 on c-fos-and c-jun-associated nucleosomes upon gene activation. *The EMBO journal,* 19**,** 3714-3726.

COMMITTEE ON MUTAGENICITY OF CHEMICALS IN FOOD, C. P. A. T. E. 2011. Guidance on a Strategy for Genotoxicity Testing of Chemical Substances.

CONTRERA, J. F. 2013. Validation of Toxtree and SciQSAR in silico predictive software using a publicly available benchmark mutagenicity database and their applicability for the qualification of impurities in pharmaceuticals. *Regul Toxicol Pharmacol,* 67**,** 285-93.

COOKE, M. S., EVANS, M. D., DIZDAROGLU, M. & LUNEC, J. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal,* 17**,** 1195-1214.

CORTES, U., MOYRET-LALLE, C., FALETTE, N., DURIEZ, C., GHISSASSI, F. E., BARNAS, C., MOREL, A. P., HAINAUT, P., MAGAUD, J. P. & PUISIEUX, A. 2000. BTG gene

expression in the p53-dependent and-independent cellular response to DNA damage. *Molecular carcinogenesis,* 27**,** 57-64.

CORTÉS-CIRIANO, I., ŠKUTA, C., BENDER, A. & SVOZIL, D. 2020. QSAR-derived affinity fingerprints (part 2): modeling performance for potency prediction. *Journal of Cheminformatics,* 12**,** 41.

CRUMP, K. 1984. A New Method for Determining Allowable Daily Intakes1. *Toxicological Sciences,* 4**,** 854-871.

CRUMP, K. S. 1995. Calculation of Benchmark Doses from Continuous Data. *Risk Analysis,* 15**,** 79-89.

DAL BEN, D., PALUMBO, M., ZAGOTTO, G., CAPRANICO, G. & MORO, S. 2007. DNA topoisomerase II structures and anthracycline activity: insights into ternary complex formation. *Current pharmaceutical design,* 13**,** 2766-2780.

DE LA BARRE, A.-E., GERSON, V., GOUT, S., CREAVEN, M., ALLIS, C. D. & DIMITROV, S. 2000. Core histone N-termini play an essential role in mitotic chromosome condensation. *The EMBO Journal,* 19**,** 379-391.

DEARFIELD, K. L., GOLLAPUDI, B. B., BEMIS, J. C., BENZ, R. D., DOUGLAS, G. R., ELESPURU, R. K., JOHNSON, G. E., KIRKLAND, D. J., LEBARON, M. J., LI, A. P., MARCHETTI, F., POTTENGER, L. H., RORIJE, E., TANIR, J. Y., THYBAUD, V., VAN BENTHEM, J., YAUK, C. L., ZEIGER, E. & LUIJTEN, M. 2017. Next generation testing strategy for assessment of genomic damage: A conceptual framework and considerations. *Environmental and Molecular Mutagenesis,* 58**,** 264-283.

DERTINGER, S., KRAYNAK, A., WHEELDON, R., BERNACKI, D., BRYCE, S., HALL, N., BEMIS, J., GALLOWAY, S., ESCOBAR, P. & JOHNSON, G. 2019. Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered multiflow® assay data analysis strategy. *Environmental and Molecular Mutagenesis,* 60**,** 513-533.

DEWEESE, J. E. & OSHEROFF, N. 2008. The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. *Nucleic Acids Research,* 37**,** 738-748.

DHAMODHARAN, R., JORDAN, M. A., THROWER, D., WILSON, L. & WADSWORTH, P. 1995. Vinblastine suppresses dynamics of individual microtubules in living interphase cells. *Molecular biology of the cell,* 6**,** 1215-1229.

DIMITROV, S. D., DIDERICH, R., SOBANSKI, T., PAVLOV, T. S., CHANKOV, G. V., CHAPKANOV, A. S., KARAKOLEV, Y. H., TEMELKOV, S. G., VASILEV, R. A., GEROVA, K. D., KUSEVA, C. D., TODOROVA, N. D., MEHMED, A. M., RASENBERG, M. & MEKENYAN, O. G. 2016. QSAR Toolbox – workflow and major functionalities. *SAR and QSAR in Environmental Research,* 27**,** 203-219.

DINDA, S., SANCHEZ, A. & MOUDGIL, V. 2002. Estrogen-like effects of thyroid hormone on the regulation of tumor suppressor proteins, p53 and retinoblastoma, in breast cancer cells. *Oncogene,* 21**,** 761-768.

DOAK, S. H., JENKINS, G. J., JOHNSON, G. E., QUICK, E., PARRY, E. M. & PARRY, J. M. 2007. Mechanistic influences for mutation induction curves after exposure to DNA-reactive carcinogens. *Cancer Res,* 67**,** 3904-11.

DOBROVOLSKY, V. N., MIURA, D., HEFLICH, R. H. & DERTINGER, S. D. 2010. The in vivo pig-a gene mutation assay, a potential tool for regulatory safety assessment. *Environmental and Molecular Mutagenesis,* 51**,** 825-835.

EARNSHAW, W. C., HALLIGAN, B., COOKE, C. A., HECK, M. & LIU, L. F. 1985. Topoisomerase II is a structural component of mitotic chromosome scaffolds. *The Journal of cell biology,* 100**,** 1706-1715.

EARNSHAW, W. C. & HECK, M. 1985. Localization of topoisomerase II in mitotic chromosomes. *The Journal of cell biology,* 100**,** 1716-1725.

EFSA 2009. European Food Safety Authority. Guidance of the scientific committee on use of the benchmark dose approach in risk assessment. *EFSA Journal,* 1150**,** 1-72.

EFSA 2017. Update: use of the benchmark dose approach in risk assessment. *EFSA Journal,* 15**,** e04658.

ERTL, P., ROHDE, B. & SELZER, P. 2000. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *Journal of Medicinal Chemistry,* 43**,** 3714-3717.

EVANS, M. D., DIZDAROGLU, M. & COOKE, M. S. 2004. Oxidative DNA damage and disease: induction, repair and significance. *Mutation Research/Reviews in Mutation Research,* 567**,** 1-61.

EVANS-ROBERTS, K. M., MITCHENALL, L. A., WALL, M. K., LEROUX, J., MYLNE, J. S. & MAXWELL, A. 2016. DNA Gyrase Is the Target for the Quinolone Drug Ciprofloxacin in Arabidopsis thaliana. *The Journal of biological chemistry,* 291**,** 3136-3144.

FELIX, C. A. 2001. Leukemias related to treatment with DNA topoisomerase II inhibitors*. *Medical and Pediatric Oncology,* 36**,** 525-535.

FENG, C., WANG, H., LU, N., CHEN, T., HE, H., LU, Y. & TU, X. M. 2014. Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry,* 26**,** 105-109.

FIX, E. 1985. *Discriminatory analysis: nonparametric discrimination, consistency properties*, USAF school of Aviation Medicine.

FOSTER, R. S., FOWKES, A., CAYLEY, A., THRESHER, A., WERNER, A.-L. D., BARBER, C. G., KOCKS, G., TENNANT, R. E., WILLIAMS, R. V., KANE, S. & STALFORD, S. A. 2020. The importance of expert review to clarify ambiguous situations for (Q)SAR predictions under ICH M7. *Genes and Environment,* 42**,** 27.

FUKUSHIMA, S., KINOSHITA, A., PUATANACHOKCHAI, R., KUSHIDA, M., WANIBUCHI, H. & MORIMURA, K. 2005. Hormesis and dose–response-mediated mechanisms in carcinogenesis: evidence for a threshold in carcinogenicity of non-genotoxic carcinogens. *Carcinogenesis,* 26**,** 1835-1845.

GALLOWAY, S. M., MILLER, J. E., ARMSTRONG, M. J., BEAN, C. L., SKOPEK, T. R. & NICHOLS, W. W. 1998. DNA synthesis inhibition as an indirect mechanism of chromosome aberrations: comparison of DNA-reactive and non-DNA-reactive clastogens. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 400**,** 169-186.

GAO, Y., FERGUSON, D. O., XIE, W., MANIS, J. P., SEKIGUCHI, J., FRANK, K. M., CHAUDHURI, J., HORNER, J., DEPINHO, R. A. & ALT, F. W. 2000. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature,* 404**,** 897-900.

GAO, Y., SUN, Y., FRANK, K. M., DIKKES, P., FUJIWARA, Y., SEIDL, K. J., SEKIGUCHI, J. M., RATHBUN, G. A., SWAT, W. & WANG, J. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell,* 95**,** 891-902.

GIRDLER, F., GASCOIGNE, K. E., EYERS, P. A., HARTMUTH, S., CRAFTER, C., FOOTE, K. M., KEEN, N. J. & TAYLOR, S. S. 2006. Validating Aurora B as an anti-cancer drug target. *Journal of Cell Science,* 119**,** 3664-3675.

GOCKE, E., BALLANTYNE, M., WHITWELL, J. & MÜLLER, L. 2009. MNT and MutaMouse studies to define the in vivo dose response relations of the genotoxicity of EMS and ENU. *Toxicol Lett,* 190**,** 286-97.

GOCKE, E. & WALL, M. 2009. In vivo genotoxicity of EMS: statistical assessment of the dose response curves. *Toxicol Lett,* 190**,** 298-302.

GOLD, L. S., MANLEY, N. B., SLONE, T. H., GARFINKEL, G. B., AMES, B. N., ROHRBACH, L., STERN, B. R. & CHOW, K. 1995. Sixth plot of the carcinogenic potency database: results of animal bioassays published in the General Literature 1989 to 1990 and by the National Toxicology Program 1990 to 1993. *Environmental Health Perspectives,* 103**,** 3-122.

GOLD, L. S., SAWYER, C. B., MAGAW, R., BACKMAN, G. M., VECIANA, M. D., LEVINSON, R., HOOPER, N. K., HAVENDER, W. R., BERNSTEIN, L., PETO, R., PIKE, M. C. & AMES, B. N. 1984. A carcinogenic potency database of the standardized results of animal bioassays. *Environmental Health Perspectives,* 58**,** 9-319.

GOLD, L. S., SLONE, T. H. & BERNSTEIN, L. 1989. Summary of carcinogenic potency and positivity for 492 rodent carcinogens in the carcinogenic potency database. *Environmental Health Perspectives,* 79**,** 259-272.

GOLLAPUDI, B., JOHNSON, G., HERNANDEZ, L., POTTENGER, L., DEARFIELD, K., JEFFREY, A., JULIEN, E., KIM, J., LOVELL, D., MACGREGOR, J., MOORE, M., VAN BENTHEM, J., WHITE, P., ZEIGER, E. & THYBAUD, V. 2013a. Quantitative approaches for assessing dose–response relationships in genetic toxicology studies. *Environmental and Molecular Mutagenesis,* 54**,** 8-18.

GOLLAPUDI, B. B., JOHNSON, G. E., HERNANDEZ, L. G., POTTENGER, L. H., DEARFIELD, K. L., JEFFREY, A. M., JULIEN, E., KIM, J. H., LOVELL, D. P., MACGREGOR, J. T., MOORE, M. M., VAN BENTHEM, J., WHITE, P. A., ZEIGER, E. & THYBAUD, V. 2013b. Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. *Environ Mol Mutagen,* 54**,** 8-18.

GOLLAPUDI, B. B., THYBAUD, V., KIM, J. H. & HOLSAPPLE, M. 2011. Strategies for the follow-up of positive results in the in vitro genotoxicity assays--an international collaborative initiative. *Environ Mol Mutagen,* 52**,** 174-6.

GOLLAPUDI, P., BHAT, V. S. & EASTMOND, D. A. 2019. Concentration-response studies of the chromosome-damaging effects of topoisomerase II inhibitors determined in vitro using human TK6 cells. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 841**,** 49-56.

GONÇALVES, D. N. S., GONÇALVES, C. D. M., ASSIS, T. F. D. & SILVA, M. A. D. 2014. Analysis of the Difference between the Euclidean Distance and the Actual Road Distance in Brazil. *Transportation Research Procedia,* 3**,** 876-885.

GRACE, P., GEORGE, H., PRACHI, P. & IMRAN, S. 2017. Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Computational toxicology (Amsterdam, Netherlands),* 3**,** 1-18.

GROMBACHER, T., EICHHORN, U. & KAINA, B. 1998. p53 is involved in regulation of the DNA repair gene O6-methylguanine-DNA methyltransferase (MGMT) by DNA damaging agents. *Oncogene,* 17**,** 845-851.

GUERARD, M., BAUM, M., BITSCH, A., EISENBRAND, G., ELHAJOUJI, A., EPE, B., HABERMEYER, M., KAINA, B., MARTUS, H. & PFUHLER, S. 2015. Assessment of mechanisms driving non-linear dose–response relationships in genotoxicity testing. *Mutation Research/Reviews in Mutation Research,* 763**,** 181-201.

HABER, L. T., DOURSON, M. L., ALLEN, B. C., HERTZBERG, R. C., PARKER, A., VINCENT, M. J., MAIER, A. & BOOBIS, A. R. 2018. Benchmark dose (BMD) modeling: current practice, issues, and challenges. *Critical Reviews in Toxicology,* 48**,** 387-415.

HACKELING, G. 2017. *Mastering Machine Learning with scikit-learn*, Packt Publishing Ltd.

HALLE, W., BAEGER, I., EKWALL, B. & SPIELMANN, H. 1991. Correlation between In Vitro Cytotoxicity and Octanol/Water Partition Coefficient of 29 Substances from the MEIC Programme. *Alternatives to Laboratory Animals,* 19**,** 338-343.

HANDE, K. R. 1998. Etoposide: four decades of development of a topoisomerase II inhibitor. *European Journal of Cancer,* 34**,** 1514-1521.

HANS, F. & DIMITROV, S. 2001. Histone H3 phosphorylation and cell division. *Oncogene,* 20**,** 3021-3027.

HANSER, T., BARBER, C., ROSSER, E., VESSEY, J. D., WEBB, S. J. & WERNER, S. 2014. Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge. *Journal of cheminformatics,* 6**,** 1-21.

HEFLICH, R. H. & NEFT, R. E. 1994. Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutation Research/Reviews in Genetic Toxicology,* 318**,** 73-174.

HENDERSON, L., ALBERTINI, S. & AARDEMA, M. 2000. Thresholds in genotoxicity responses. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 464**,** 123-128.

HENDRIKS, G., ATALLAH, M., MOROLLI, B., CALLÉJA, F., RAS-VERLOOP, N., HUIJSKENS, I., RAAMSMAN, M., VAN DE WATER, B. & VRIELING, H. 2011. The ToxTracker Assay: Novel GFP Reporter Systems that Provide Mechanistic Insight into the Genotoxic Properties of Chemicals. *Toxicological Sciences,* 125**,** 285-298.

HENDRIKS, G., DERR, R. S., MISOVIC, B., MOROLLI, B., CALLÉJA, F. M. G. R. & VRIELING, H. 2015. The Extended ToxTracker Assay Discriminates Between Induction of DNA Damage, Oxidative Stress, and Protein Misfolding. *Toxicological Sciences,* 150**,** 190-203.

HENEL, G. & SCHMITZ, J. L. 2007. Basic Theory and Clinical Applications of Flow Cytometry. *Laboratory Medicine,* 38**,** 428-436.

HENGSTLER, J., BOGDANFFY, M., BOLT, H. & OESCH, F. 2003. Challenging dogma: thresholds for genotoxic carcinogens? The case of vinyl acetate. *Annual review of pharmacology and toxicology,* 43**,** 485-520.

HERNANDEZ, G., SLOB, W., VAN STEEG, H. & VAN BENTHEM, J. 2010. Comparison of Carcinogenic Potency Estimates toIn VivoGenotoxic Potencies from the Micronucleus, Transgenic Rodent Mutation and Comet Assay Using the Benchmark Dose Approach.: 15. *Environmental & Molecular Mutagenesis,* 51.

HERNÁNDEZ, L., JOHNSON, G., POTTENGER, L. & VAN BENTHEM, J. Analysis of Low-Dose Mutagenic Responses and the Applicability of Genotoxicity Tests for Carcinogen Potency Prediction. Environmental and Molecular Mutagenesis, 2011a. WILEY-BLACKWELL COMMERCE PLACE, 350 MAIN ST, MALDEN 02148, MA USA, S26-S26.

HERNÁNDEZ, L. G., SLOB, W., VAN STEEG, H. & VAN BENTHEM, J. 2011b. Can carcinogenic potency be predicted from in vivo genotoxicity data? a meta-analysis of historical data. *Environmental and Molecular Mutagenesis,* 52**,** 518-528.

HERRERO, J., VALENCIA, A. & DOPAZO, J. N. 2001. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics,* 17**,** 126-136.

HIRANO, S., SUN, X. & TSUMOTO, S. 2004. Comparison of clustering methods for clinical databases. *Information Sciences,* 159**,** 155-165.

HOLM, C., GOTO, T., WANG, J. C. & BOTSTEIN, D. 1985. DNA topoisomerase II is required at the time of mitosis in yeast. *Cell,* 41**,** 553-563.

HONMA, M. 2020. An assessment of mutagenicity of chemical substances by (quantitative) structure–activity relationship. *Genes and Environment,* 42**,** 23.

HU, J., WANG, W., ZHU, Z., CHANG, H., PAN, F. & LIN, B. 2007. Quantitative Structure–Activity Relationship Model for Prediction of Genotoxic Potential for Quinolone Antibacterials. *Environmental Science & Technology,* 41**,** 4806-4812.

HURD, C., KHATTREE, N., DINDA, S., ALBAN, P. & MOUDGIL, V. K. 1997. Regulation of tumor suppressor proteins, p53 and retinoblastoma, by estrogen and antiestrogens in breast cancer cells. *Oncogene,* 15**,** 991-995.

ICH 1998. ICH guideline S1B Carcinogenicity: testing for carcinogenicity of pharmaceuticals.

ICH 2008. ICH guideline S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use.

ICH 2013. ICH guideline M7(R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk.

ISAACS, R. J., DAVIES, S. L., SANDRI, M. I., REDWOOD, C., WELLS, N. J. & HICKSON, I. D. 1998. Physiological regulation of eukaryotic topoisomerase II. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 1400**,** 121-137.

JADHAV, A. K. & KARUPPAYIL, S. M. 2017. Molecular docking studies on thirteen fluoroquinolines with human topoisomerase II a and b. *In Silico Pharmacology,* 5**,** 4.

JENKINS, G. J. S., ZAÏR, Z., JOHNSON, G. E. & DOAK, S. H. 2010. Genotoxic thresholds, DNA repair, and susceptibility in human populations. *Toxicology,* 278**,** 305-310.

JMP. 2021. *Distance Method Formulas* [Online]. Available: https://www.jmp.com/support/help/en/16.0/index.shtml#page/jmp/distance-method-formulas.shtml#ww177809 [Accessed 26th May 2021].

JOHANSON, G. 2010. 1.08 - Modeling of Disposition. *In:* MCQUEEN, C. A. (ed.) *Comprehensive Toxicology (Second Edition).* Oxford: Elsevier.

JOHNSON, G., SOETEMAN-HERNÁNDEZ, L., GOLLAPUDI, B., BODGER, O., DEARFIELD, K., HEFLICH, R., HIXON, J., LOVELL, D., MACGREGOR, J. & POTTENGER, L. 2014a. Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. *Environmental and molecular mutagenesis,* 55**,** 609-623.

JOHNSON, G. E., DOAK, S. H., GRIFFITHS, S. M., QUICK, E. L., SKIBINSKI, D. O. F., ZAÏR, Z. M. & JENKINS, G. J. 2009. Non-linear dose–response of DNA-reactive genotoxins: Recommendations for data analysis. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 678**,** 95-100.

JOHNSON, G. E. & PARRY, E. M. 2008. Mechanistic investigations of low dose exposures to the genotoxic compounds bisphenol-A and rotenone. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 651**,** 56-63.

JOHNSON, G. E., SLOB, W., DOAK, S. H., FELLOWS, M. D., GOLLAPUDI, B. B., HEFLICH, R. H., REES, B. J., SOETEMAN-HERNÁNDEZ, L. G., VERMA, J. R., WILLS, J. W., JENKINS, G. J.

S. & WHITE, P. A. 2014b. New approaches to advance the use of genetic toxicology analyses for human health risk assessment. *Toxicology Research,* 4**,** 667-676.

JOHNSON, G. E., YAMAMOTO, M., SUZUKI, Y., ADACHI, H., KYOYA, T., TAKASAWA, H., HORIBATA, K., TSUTSUMI, E., WADA, K., KIKUZUKI, R., YOSHIDA, I., KIMOTO, T., MAEDA, A. & NARUMI, K. 2016. Measuring reproducibility of dose response data for the Pig-a assay using covariate benchmark dose analysis. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 811**,** 135-139.

KAINA, B. & CHRISTMANN, M. 2002. DNA repair in resistance to alkylating anticancer drugs. *International journal of clinical pharmacology and therapeutics,* 40**,** 354-367.

KIRKLAND, D., AARDEMA, M., HENDERSON, L. & MÜLLER, L. 2005. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutat Res,* 584**,** 1-256.

KIRKLAND, D., PFUHLER, S., TWEATS, D., AARDEMA, M., CORVI, R., DARROUDI, F., ELHAJOUJI, A., GLATT, H., HASTWELL, P., HAYASHI, M., KASPER, P., KIRCHNER, S., LYNCH, A., MARZIN, D., MAURICI, D., MEUNIER, J.-R., MÜLLER, L., NOHYNEK, G., PARRY, J., PARRY, E., THYBAUD, V., TICE, R., VAN BENTHEM, J., VANPARYS, P. & WHITE, P. 2007. How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 628**,** 31-55.

KIRKLAND, D., REEVE, L., GATEHOUSE, D. & VANPARYS, P. 2011. A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 721, 27-73.

KIRSCH-VOLDERS, M., AARDEMA, M. & ELHAJOUJI, A. 2000. Concepts of threshold in mutagenesis and carcinogenesis. *Mutat Res,* 464**,** 3-11.

KLAASSEN, C. 2007. *Casarett & Doull's Toxicology: The Basic Science of Poisons, Seventh Edition*, Mcgraw-hill.

KOLLMANNSBERGER, C., BEYER, J., DROZ, J. P., HARSTRICK, A., HARTMANN, J. T., BIRON, P., FLÉCHON, A., SCHÖFFSKI, P., KUCZYK, M., SCHMOLL, H. J., KANZ, L. & BOKEMEYER, C. 1998. Secondary leukemia following high cumulative doses of etoposide in patients treated for advanced germ cell tumors. *Journal of Clinical Oncology,* 16**,** 3386-3391.

KONDURI, S. D., MEDISETTY, R., LIU, W., KAIPPARETTU, B. A., SRIVASTAVA, P., BRAUCH, H., FRITZ, P., SWETZIG, W. M., GARDNER, A. E., KHAN, S. A. & DAS, G. M. 2010. Mechanisms of estrogen receptor antagonism toward p53 and its implications in breast cancer therapeutic response and stem cell regulation. *Proceedings of the National Academy of Sciences,* 107**,** 15081-15086.

KRAMER, O. 2013. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors.* Springer.

LARSEN, A. K., ESCARGUEIL, A. E. & SKLADANOWSKI, A. 2003. Catalytic topoisomerase II inhibitors in cancer therapy. *Pharmacology & Therapeutics,* 99**,** 167-181.

LE HEGARAT, L., ROUDOT, A.-C. & FESSARD, V. 2020. Benchmark dose analyses of γH2AX and pH3 endpoints for quantitative comparison of in vitro genotoxicity potential of lipophilic phycotoxins. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 852**,** 503169.

LEBARON, M., KAN, H., GETER, D., POTTENGER, L., ZHANG, F., SCHISLER, M., BARTELS, M. & GOLLAPUDI, B. Non-linear low dose response in methyl methanesulfonate (MMS)-treated rats as measured by multiple endpoints. 47th Annual Meeting of the Society of Toxicology Abstract, 2008.

LEE, S., ELENBAAS, B., LEVINE, A. & GRIFFITH, J. 1995. p53 and its 14 kDa C-terminal domain recognize primary DNA damage in the form of insertion/deletion mismatches. *Cell,* 81**,** 1013-1020.

LEVINE, A. J., MOMAND, J. & FINLAY, C. A. 1991. The p53 tumour suppressor gene. *Nature,* 351**,** 453-456.

LI, H. H., CHEN, R., HYDUKE, D. R., WILLIAMS, A., FRÖTSCHL, R., ELLINGER-ZIEGELBAUER, H., O'LONE, R., YAUK, C. L., AUBRECHT, J. & FORNACE, A. J., JR. 2017. Development and

validation of a high-throughput transcriptomic biomarker to address 21st century genetic toxicology needs. *Proc Natl Acad Sci U S A,* 114**,** E10881-e10889.

LI, H. H., HYDUKE, D. R., CHEN, R., HEARD, P., YAUK, C. L., AUBRECHT, J. & FORNACE, A. J., JR. 2015. Development of a toxicogenomics signature for genotoxicity using a dose-optimization and informatics strategy in human cells. *Environ Mol Mutagen,* 56**,** 505-19.

LINDAHL, T. & WOOD, R. D. 1999. Quality Control by DNA Repair. *Science,* 286**,** 1897-1905.

LIU, Y. & KULESZ-MARTIN, M. 2001. p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. *Carcinogenesis,* 22**,** 851-860.

LOEB, K. R. & LOEB, L. A. 2000. Significance of multiple mutations in cancer. *Carcinogenesis,* 21**,** 379-385.

LONG, B. H. 1992. Mechanisms of action of teniposide (VM-26) and comparison with etoposide (VP-16). *Seminars in oncology,* 19**,** 3-19.

LORGE, E., HAYASHI, M., ALBERTINI, S. & KIRKLAND, D. 2008. Comparison of different methods for an accurate assessment of cytotoxicity in the in vitro micronucleus test: I. Theoretical aspects. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 655**,** 1-3.

LU, Y., CHEN, J., XIAO, M., LI, W. & MILLER, D. D. 2012. An Overview of Tubulin Inhibitors That Interact with the Colchicine Binding Site. *Pharmaceutical Research,* 29**,** 2943-2971.

LUKAMOWICZ, M., WOODWARD, K., KIRSCH-VOLDERS, M., SUTER, W. & ELHAJOUJI, A. 2011. A flow cytometry based in vitro micronucleus assay in TK6 cells--validation using early stage pharmaceutical development compounds. *Environ Mol Mutagen,* 52**,** 363-72.

LUTZ, W. K. 1998. Dose–response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear–nonlinear curves, practical thresholds, J-shapes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 405**,** 117-124.

LUTZ, W. K. 2009. The Viracept (nelfinavir)--ethyl methanesulfonate case: a threshold risk assessment for human exposure to a genotoxic drug contamination? *Toxicol Lett,* 190**,** 239-42.

LUTZ, W. K. & LUTZ, R. W. 2009. Statistical model to estimate a threshold dose and its confidence limits for the analysis of sublinear dose-response relationships, exemplified for mutagenicity data. *Mutat Res,* 678**,** 118-22.

LYNCH, A., HARVEY, J., AYLOTT, M., NICHOLAS, E., BURMAN, M., SIDDIQUI, A., WALKER, S. & REES, R. 2003. Investigations into the concept of a threshold for topoisomerase inhibitor-induced clastogenicity. *Mutagenesis,* 18**,** 345-353.

MACGREGOR, J., FRÖTSCHL, R., WHITE, P., CRUMP, K., EASTMOND, D., FUKUSHIMA, S., GUÉRARD, M., HAYASHI, M., SOETEMAN-HERNÁNDEZ, L., KASAMATSU, T., LEVY, D., MORITA, T., MÜLLER, L., SCHOENY, R., SCHULER, M. & THYBAUD, V. 2015. IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). *Mutation Research - Genetic Toxicology and Environmental Mutagenesis,* 783**,** 55-65.

MACMILLAN, D. S. & CHILTON, M. L. 2019. A defined approach for predicting skin sensitisation hazard and potency based on the guided integration of in silico, in chemico and in vitro data using exclusion criteria. *Regulatory Toxicology and Pharmacology,* 101**,** 35-47.

MARCHANT, C. A., BRIGGS, K. A. & LONG, A. 2008. In Silico Tools for Sharing Data and Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor, and Vitic. *Toxicology Mechanisms and Methods,* 18**,** 177-187.

MARKOVITS, J., LINASSIER, C., FOSSÉ, P., COUPRIE, J., PIERRE, J., JACQUEMIN-SABLON, A., SAUCIER, J.-M., LE PECQ, J.-B. & LARSEN, A. K. 1989. Inhibitory effects of the tyrosine kinase inhibitor genistein on mammalian DNA topoisomerase II. *Cancer research,* 49**,** 5111-5117.

MATTIONI, B. E., KAUFFMAN, G. W., JURS, P. C., CUSTER, L. L., DURHAM, S. K. & PEARL, G. M. 2003. Predicting the genotoxicity of secondary and aromatic amines using data subsetting to generate a model ensemble. *J Chem Inf Comput Sci,* 43**,** 949-63.

MAUNZ, A., GÜTLEIN, M., RAUTENBERG, M., VORGRIMMLER, D., GEBELE, D. & HELMA, C. 2013. Lazar: a modular predictive toxicology framework. *Frontiers in pharmacology,* 4**,** 38.

MAXWELL, G., MACKAY, C., CUBBERLEY, R., DAVIES, M., GELLATLY, N., GLAVIN, S., GOUIN, T., JACQUOILLEOT, S., MOORE, C., PENDLINGTON, R., SAIB, O., SHEFFIELD, D., STARK, R. & SUMMERFIELD, V. 2014. Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicology in Vitro,* 28**,** 8-12.

MCCLENDON, A. K. & OSHEROFF, N. 2007. DNA topoisomerase II, genotoxity, and cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 623**,** 83-97.

MCELREATH, R. 2018. *Statistical rethinking: A Bayesian course with examples in R and Stan*, Chapman and Hall/CRC.

MEEK, D. W. 2009. Tumour suppression by p53: a role for the DNA damage response? *Nature Reviews Cancer,* 9**,** 714-723.

MERSCH-SUNDERMANN, V., ROSENKRANZ, H. S. & KLOPMAN, G. 1994. The structural basis of the genotoxicity of nitroarenofurans and related compounds. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,* 304**,** 271-284.

MEYER, C., KOWARZ, E., HOFMANN, J., RENNEVILLE, A., ZUNA, J., TRKA, J., BEN ABDELALI, R., MACINTYRE, E., DE BRAEKELEER, E., DE BRAEKELEER, M., DELABESSE, E., DE OLIVEIRA, M. P., CAVÉ, H., CLAPPIER, E., VAN DONGEN, J. J. M., BALGOBIND, B. V., VAN DEN HEUVEL-EIBRINK, M. M., BEVERLOO, H. B., PANZER-GRÜMAYER, R., TEIGLER-SCHLEGEL, A., HARBOTT, J., KJELDSEN, E., SCHNITTGER, S., KOEHL, U., GRUHN, B., HEIDENREICH, O., CHAN, L. C., YIP, S. F., KRZYWINSKI, M., ECKERT, C., MÖRICKE, A., SCHRAPPE, M., ALONSO, C. N., SCHÄFER, B. W., KRAUTER, J., LEE, D. A., ZUR STADT, U., TE KRONNIE, G., SUTTON, R., IZRAELI, S., TRAKHTENBROT, L., LO NIGRO, L., TSAUR, G., FECHINA, L., SZCZEPANSKI, T., STREHL, S., ILENCIKOVA, D., MOLKENTIN, M., BURMEISTER, T., DINGERMANN, T., KLINGEBIEL, T. & MARSCHALEK, R. 2009. New insights to the MLL recombinome of acute leukemias. *Leukemia,* 23**,** 1490-1499.

MILLER, J. A. & MILLER, E. C. 1971. Chemical carcinogenesis: mechanisms and approaches to its control. *Journal of the National Cancer Institute,* 47**,** V-XIV.

MOMBELLI, E. & PANDARD, P. 2021. Evaluation of the OECD QSAR toolbox automatic workflow for the prediction of the acute toxicity of organic chemicals to fathead minnow. *Regulatory Toxicology and Pharmacology,* 122**,** 104893.

MONTECUCCO, A., ZANETTA, F. & BIAMONTI, G. 2015. Molecular mechanisms of etoposide. *EXCLI journal,* 14**,** 95-108.

MORIWAKI, H., TIAN, Y.-S., KAWASHITA, N. & TAKAGI, T. 2018. Mordred: a molecular descriptor calculator. *Journal of Cheminformatics,* 10**,** 4.

MUGGEO, V. M. 2008. Segmented: an R package to fit regression models with broken-line relationships. *R news,* 8**,** 20-25.

MÜLLER, L. & GOCKE, E. 2009. Considerations regarding a permitted daily exposure calculation for ethyl methanesulfonate. *Toxicol Lett,* 190**,** 330-2.

MÜLLER, L., GOCKE, E., LAVÉ, T. & PFISTER, T. 2009. Ethyl methanesulfonate toxicity in Viracept--a comprehensive human risk assessment based on threshold data for genotoxicity. *Toxicol Lett,* 190**,** 317-29.

MÜLLER, L. & KASPER, P. 2000. Human biological relevance and the use of threshold-arguments in regulatory genotoxicity assessment: experience with pharmaceuticals. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 464**,** 19-34.

MÜLLER, L., MAUTHE, R. J., RILEY, C. M., ANDINO, M. M., ANTONIS, D. D., BEELS, C., DEGEORGE, J., DE KNAEP, A. G., ELLISON, D., FAGERLAND, J. A., FRANK, R., FRITSCHEL, B., GALLOWAY, S., HARPUR, E., HUMFREY, C. D., JACKS, A. S., JAGOTA, N., MACKINNON, J., MOHAN, G., NESS, D. K., O'DONOVAN, M. R., SMITH, M. D., VUDATHALA, G. & YOTTI, L. 2006. A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. *Regul Toxicol Pharmacol,* 44**,** 198-211.

MÜLLER, L. & SINGER, T. 2009. EMS in Viracept--the course of events in 2007 and 2008 from the non-clinical safety point of view. *Toxicol Lett,* 190**,** 243-7.

MURTAGH, F. & LEGENDRE, P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification,* 31**,** 274-295.

NELSON, W. G. & KASTAN, M. B. 1994. DNA strand breaks: the DNA template alterations that trigger p53-dependent DNA damage response pathways. *Molecular and Cellular Biology,* 14**,** 1815-1823.

NETTLES, J. H., JENKINS, J. L., BENDER, A., DENG, Z., DAVIES, J. W. & GLICK, M. 2006. Bridging Chemical and Biological Space: "Target Fishing" Using 2D and 3D Molecular Descriptors. *Journal of Medicinal Chemistry,* 49**,** 6802-6810.

NITISS, J. L. 2009. DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer,* 9**,** 327-337.

NOWAK, S. J. & CORCES, V. G. 2000. Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes & development,* 14**,** 3003-3013.

OECD 2016. *Test No. 487: In Vitro Mammalian Cell Micronucleus Test*.

OECD 2018. Users Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways.

OWELLEN, R. J., OWENS JR, A. H. & DONIGIAN, D. W. 1972. The binding of vincristine, vinblastine and colchicine to tubulin. *Biochemical and biophysical research communications,* 47**,** 685-691.

PATLEWICZ, G., KUSEVA, C., KESOVA, A., POPOVA, I., ZHECHEV, T., PAVLOV, T., ROBERTS, D. W. & MEKENYAN, O. 2014. Towards AOP application – Implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regulatory Toxicology and Pharmacology,* 69**,** 529-545.

PELKONEN, O., TERRON, A., HERNANDEZ, A. F., MENENDEZ, P. & BENNEKOU, S. H. 2017. Chemical exposure and infant leukaemia: development of an adverse outcome pathway (AOP) for aetiology and risk assessment research. *Archives of toxicology,* 91**,** 2763-2780.

PENDLETON, M., LINDSEY, R. H., JR., FELIX, C. A., GRIMWADE, D. & OSHEROFF, N. 2014. Topoisomerase II and leukemia. *Annals of the New York Academy of Sciences,* 1310**,** 98-110.

POTTENGER, L. H. & GOLLAPUDI, B. B. 2010. Genotoxicity testing: Moving beyond qualitative "screen and bin" approach towards characterization of dose-response and thresholds. *Environmental and Molecular Mutagenesis,* 51**,** 792-799.

POTTENGER, L. H., SCHISLER, M. R., ZHANG, F., BARTELS, M. J., FONTAINE, D. D., MCFADDEN, L. G. & BHASKAR GOLLAPUDI, B. 2009. Dose–response and operational thresholds/NOAELs for in vitro mutagenic effects from DNA-reactive mutagens, MMS and MNU. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 678**,** 138-147.

REACH 2007. European Parliament and Council Regulation (EC) No 1907/2006 (Corrigendum 29 May 2007) and Directive 2006/121/EC (Corrigendum 29 May 2007). Official Journal of the European Union L 396 of 30 December 2006.

REINHARDT, H. C., JIANG, H., HEMANN, M. T. & YAFFE, M. B. 2009. Exploiting synthetic lethal interactions for targeted cancer therapy. *Cell cycle,* 8**,** 3112-3119.

RIBAUDO, G., ONGARO, A., ZORZAN, M., PEZZANI, R., REDAELLI, M., ZAGOTTO, G., MEMO, M. & GIANONCELLI, A. 2019. Investigation of the molecular reactivity of bioactive oxiranylmethyloxy anthraquinones. *Archiv der Pharmazie,* 352**,** 1900030.

RICHES, L. C., LYNCH, A. M. & GOODERHAM, N. J. 2008. Early events in the mammalian response to DNA double-strand breaks. *Mutagenesis,* 23**,** 331-339.

ROBERTS, D. W. & APTULA, A. O. 2014. Electrophilic Reactivity and Skin Sensitization Potency of SNAr Electrophiles. *Chemical Research in Toxicology,* 27**,** 240-246.

ROBERTS, G., MYATT, G. J., JOHNSON, W. P., CROSS, K. P. & BLOWER, P. E. 2000. LeadScope: Software for Exploring Large Sets of Screening Data. *Journal of Chemical Information and Computer Sciences,* 40**,** 1302-1314.

RODRÍGUEZ-CORTEZ, V. C. & MENÉNDEZ, P. 2020. Genotoxity of permethrin and clorpyriphos on human stem and progenitor cells at different ontogeny stages: implications in leukaemia development. *EFSA Supporting Publications,* 17**,** 1866E.

ROMANOVA, L. Y., WILLERS, H., BLAGOSKLONNY, M. V. & POWELL, S. N. 2004. The interaction of p53 with replication protein A mediates suppression of homologous recombination. *Oncogene,* 23**,** 9025-9033.

ROSS, W., ROWE, T., GLISSON, B., YALOWICH, J. & LIU, L. 1984. Role of Topoisomerase II in Mediating Epipodophyllotoxin-induced DNA Cleavage. *Cancer Research,* 44**,** 5857-5860.

SAITO, H., KOYASU, J., YOSHIDA, K., SHIGEOKA, T. & KOIKE, S. 1993. Cytotoxicity of 109 chemicals to goldfish GFS cells and relationships with 1-octanol/water partition coefficients. *Chemosphere,* 26**,** 1015-1028.

SASAKI, J. C., ALLEMANG, A., BRYCE, S. M., CUSTER, L., DEARFIELD, K. L., DIETZ, Y., ELHAJOUJI, A., ESCOBAR, P. A., FORNACE, A. J., JR., FROETSCHL, R., GALLOWAY, S., HEMMANN, U., HENDRIKS, G., LI, H. H., LUIJTEN, M., OUEDRAOGO, G., PEEL, L., PFUHLER, S., ROBERTS, D. J., THYBAUD, V., VAN BENTHEM, J., YAUK, C. L. & SCHULER, M. 2020. Application of the adverse outcome pathway framework to genotoxic modes of action. *Environ Mol Mutagen,* 61**,** 114-134.

SAYIN, K., KARAKAŞ, D., KARIPER, S. E. & SAYIN, T. A. 2018. Computational study of some fluoroquinolones: Structural, spectral and docking investigations. *Journal of Molecular Structure,* 1156**,** 172-181.

SCHULTZ, T. W., AMCOFF, P., BERGGREN, E., GAUTIER, F., KLARIC, M., KNIGHT, D. J., MAHONY, C., SCHWARZ, M., WHITE, A. & CRONIN, M. T. D. 2015. A strategy for structuring and reporting a read-across prediction of toxicity. *Regulatory Toxicology and Pharmacology,* 72**,** 586-601.

SEGALL, M. D. & BARBER, C. 2014. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discov Today,* 19**,** 688-93.

SEO, J.-E., TRYNDYAK, V., WU, Q., DREVAL, K., POGRIBNY, I., BRYANT, M., ZHOU, T., ROBISON, T. W., MEI, N. & GUO, X. 2019. Quantitative comparison of in vitro genotoxicity between metabolically competent HepaRG cells and HepG2 cells using the high-throughput high-content CometChip assay. *Archives of Toxicology,* 93**,** 1433-1448.

SHAHIN, S., CULLINANE, C. & GRAY, P. J. 2001. Mitochondrial and nuclear DNA damage induced by sulphur mustard in keratinocytes. *Chemico-Biological Interactions,* 138**,** 231-245.

SHAPIRO, A. B. & AUSTIN, C. A. 2014. A high-throughput fluorescence anisotropy-based assay for human topoisomerase II β-catalyzed ATP-dependent supercoiled DNA relaxation. *Analytical Biochemistry,* 448**,** 23-29.

SLOB, W. 1999. Thresholds in Toxicology and Risk Assessment. *International Journal of Toxicology,* 18**,** 259-268.

SLOB, W. 2002. Dose-Response Modeling of Continuous Endpoints. *Toxicological Sciences,* 66**,** 298-312.

SLOB, W. 2017. A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. *Crit Rev Toxicol,* 47**,** 342-351.

SLOB, W. 2018. Joint project on benchmark dose modelling with RIVM. *EFSA Supporting Publications,* 15**,** 1497E.

SLOB, W. & SETZER, R. 2014. Shape and steepness of toxicological dose-response relationships of continuous endpoints. *Crit Rev Toxicol,* 44**,** 270-97.

SNYDER, R. D. & DIEHL, M. S. 2000. The bleomycin amplification assay in V79 cells predicts frameshift mutagenicity of intercalative agents. *Mutagenesis,* 15**,** 203-205.

SNYDER, R. D., EWING, D. E. & HENDRY, L. B. 2004. Evaluation of DNA intercalation potential of pharmaceuticals and other chemicals by cell-based and three-dimensional computational approaches. *Environmental and Molecular Mutagenesis,* 44**,** 163-173.

SNYDER, R. D. & STREKOWSKI, L. 1999. Enhancement of bleomycin-induced micronucleus formation in V79 cells as a rapid and sensitive screen for non-covalent DNA-binding compounds. *Mutation research,* 444**,** 181-192.

SOETEMAN-HERNÁNDEZ, L. G., FELLOWS, M. D., JOHNSON, G. E. & SLOB, W. 2015. Correlation of In Vivo Versus In Vitro Benchmark Doses (BMDs) Derived From Micronucleus Test Data: A Proof of Concept Study. *Toxicological Sciences,* 148**,** 355-367.

SOETEMAN-HERNÁNDEZ, L. G., JOHNSON, G. E. & SLOB, W. 2016. Estimating the carcinogenic potency of chemicals from the in vivo micronucleus test. *Mutagenesis,* 31**,** 347-358.

SOLARY, E., BERTRAND, R. & POMMIER, Y. 1994. Apoptosis Induced by DNA Topoisomerase I and II Inhibitors in Human Leukemic HL-60 Cells. *Leukemia & Lymphoma,* 15**,** 21-32.

SPEIGHT, J. G. 2018. Chapter 9 - Molecular Interactions, Partitioning, and Thermodynamics. *In:* SPEIGHT, J. G. (ed.) *Reaction Mechanisms in Environmental Engineering.* Butterworth-Heinemann.

SPENCER, N. H. 2013. *Essentials of multivariate data analysis*, CRC press.

SPINOZZI, F., PAGLIACCI, M. C., MIGLIORATI, G., MORACA, R., GRIGNANI, F., RICCARDI, C. & NICOLETTI, I. 1994. The natural tyrosine kinase inhibitor genistein produces cell cycle arrest and apoptosis in Jurkat T-leukemia cells. *Leukemia research,* 18**,** 431-439.

STANULLA, M., WANG, J., CHERVINSKY, D. S. & APLAN, P. D. 1997. Topoisomerase II inhibitors induce DNA double-strand breaks at a specific site within the AML1 locus. *Leukemia,* 11**,** 490-496.

STURLA, S. J. 2018. Point of Departure. *Chemical Research in Toxicology,* 31**,** 2-3.

SUN, H., WANG, Y., CHEFF, D. M., HALL, M. D. & SHEN, M. 2020. Predictive models for estimating cytotoxicity on the basis of chemical structures. *Bioorganic & Medicinal Chemistry,* 28**,** 115422.

SWAIN, C. G. & SCOTT, C. B. 1953. Quantitative Correlation of Relative Rates. Comparison of Hydroxide Ion with Other Nucleophilic Reagents toward Alkyl Halides, Esters, Epoxides and Acyl Halides1. *Journal of the American Chemical Society,* 75**,** 141-147.

TADDY, M. 2019. *Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions*, McGraw Hill Professional.

THUMKEO, D., WATANABE, S. & NARUMIYA, S. 2013. Physiological roles of Rho and Rho effectors in mammals. *European journal of cell biology,* 92**,** 303-315.

THYBAUD, V., AARDEMA, M., CLEMENTS, J., DEARFIELD, K., GALLOWAY, S., HAYASHI, M., JACOBSON-KRAM, D., KIRKLAND, D., MACGREGOR, J. T., MARZIN, D., OHYAMA, W., SCHULER, M., SUZUKI, H. & ZEIGER, E. 2007. Strategy for genotoxicity testing: hazard identification and risk assessment in relation to in vitro testing. *Mutat Res,* 627**,** 41-58.

TIAN, S., CYR, A., ZEISE, K., BRYCE, S., HALL, N., BEMIS, J. & DERTINGER, S. 2020. 3Rs-friendly approach to exogenous metabolic activation that supports high-throughput genetic toxicology testing. *Environmental and molecular mutagenesis,* 61**,** 408.

TOLLEFSEN, K. E., SCHOLZ, S., CRONIN, M. T., EDWARDS, S. W., DE KNECHT, J., CROFTON, K., GARCIA-REYERO, N., HARTUNG, T., WORTH, A. & PATLEWICZ, G. 2014. Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory Toxicology and Pharmacology,* 70**,** 629-640.

TRIPATHI, K. 2011. A review on knowledge-based expert system: Concept and architecture. *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications,* 4**,** 19-23.

TWEATS, D. J., BLAKEY, D., HEFLICH, R. H., JACOBS, A., JACOBSEN, S. D., MORITA, T., NOHMI, T., O'DONOVAN, M. R., SASAKI, Y. F., SOFUNI, T. & TICE, R. 2007. Report of the IWGT working group on strategies and interpretation of regulatory in vivo tests: I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis,* 627**,** 78-91.

VAN HOOSER, A., GOODRICH, D. W., ALLIS, C. D., BRINKLEY, B. & MANCINI, M. A. 1998. Histone H3 phosphorylation is required for the initiation, but not maintenance, of mammalian chromosome condensation. *Journal of cell science,* 111**,** 3497-3506.

VERMA, J. R., REES, B. J., WILDE, E. C., THORNTON, C. A., JENKINS, G. J. S., DOAK, S. H. & JOHNSON, G. E. 2017. Evaluation of the automated MicroFlow® and Metafer™ platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. *Archives of Toxicology,* 91**,** 2689-2698.

VILLENEUVE, D. L., CRUMP, D., GARCIA-REYERO, N., HECKER, M., HUTCHINSON, T. H., LALONE, C. A., LANDESMANN, B., LETTIERI, T., MUNN, S., NEPELSKA, M., OTTINGER, M. A., VERGAUWEN, L. & WHELAN, M. 2014. Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicological Sciences,* 142**,** 312-320.

VOTANO, J. R., PARHAM, M., HALL, L. H., KIER, L. B., OLOFF, S., TROPSHA, A., XIE, Q. & TONG, W. 2004. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis,* 19**,** 365-377.

WANG, H., MAO, Y., CHEN, A. Y., ZHOU, N., LAVOIE, E. J. & LIU, L. F. 2001. Stimulation of Topoisomerase II-Mediated DNA Damage via a Mechanism Involving Protein Thiolation. *Biochemistry,* 40**,** 3316-3323.

WARD, J. H. & HOOK, M. E. 1963. Application of an Hierarchical Grouping Procedure to a Problem of Grouping Profiles. *Educational and Psychological Measurement,* 23**,** 69-81.

WHEELDON, R., BERNACKI, D., DERTINGER, S., BRYCE, S., BEMIS, J. & JOHNSON, G. 2020. Benchmark Dose Analysis of DNA Damage Biomarker Responses Provides Compound Potency and Adverse Outcome Pathway Information for the Topoisomerase II Inhibitor Class of Compounds. *Environmental and Molecular Mutagenesis,* 61**,** 396-407.

WHEELDON, R. P., DERTINGER, S. D., BRYCE, S. M., BEMIS, J. C. & JOHNSON, G. E. 2021. The use of benchmark dose uncertainty measurements for robust comparative potency analyses. *Environmental and Molecular Mutagenesis,* 62**,** 203-215.

WHITE, P. A., LONG, A. S. & JOHNSON, G. E. 2020. Quantitative Interpretation of Genetic Toxicity Dose-Response Data for Risk Assessment and Regulatory Decision-Making: Current Status and Emerging Priorities. *Environmental and Molecular Mutagenesis,* 61**,** 66-83.

WILLS, J., JOHNSON, G., DOAK, S., SOETEMAN-HERNÁNDEZ, L., SLOB, W. & WHITE, P. 2016a. Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. *Mutagenesis,* 31**,** 255-263.

WILLS, J. W., JOHNSON, G. E., BATTAION, H. L., SLOB, W. & WHITE, P. A. 2017. Comparing BMD-derived genotoxic potency estimations across variants of the transgenic rodent gene mutation assay. *Environmental and Molecular Mutagenesis,* 58**,** 632-643.

WILLS, J. W., LONG, A. S., JOHNSON, G. E., BEMIS, J. C., DERTINGER, S. D., SLOB, W. & WHITE, P. A. 2016b. Empirical analysis of BMD metrics in genetic toxicology part II: in vivo potency comparisons to promote reductions in the use of experimental animals for genetic toxicity assessment. *Mutagenesis,* 31**,** 265-275.

WILSTERMANN, A. & OSHEROFF, N. 2003. Stabilization of Eukaryotic Topoisomerase II-DNA Cleavage Complexes. *Current Topics in Medicinal Chemistry,* 3**,** 321-338.

WINICK, N. J., MCKENNA, R. W., SHUSTER, J. J., SCHNEIDER, N. R., BOROWITZ, M. J., BOWMAN, W. P., JACARUSO, D., KAMEN, B. A. & BUCHANAN, G. R. 1993. Secondary acute myeloid leukemia in children with acute lymphoblastic leukemia treated with etoposide. *Journal of Clinical Oncology,* 11**,** 209-217.

WOOD, D. A. 1969. New concepts in cancer control. Preventable and avoidable cancer. *Calif Med,* 111**,** 313-6.

WOOD, R. D., MITCHELL, M., SGOUROS, J. & LINDAHL, T. 2001. Human DNA Repair Genes. *Science,* 291**,** 1284-1289.

YAN, G.-R., ZOU, F.-Y., DANG, B.-L., ZHANG, Y., YU, G., LIU, X. & HE, Q.-Y. 2012. Genistein-induced mitotic arrest of gastric cancer cells by downregulating KIF20A, a proteomics study. *PROTEOMICS,* 12**,** 2391-2399.

YAN, Y., SU, X., LIANG, Y., ZHANG, J., SHI, C., LU, Y., GU, L. & FU, L. 2008. Emodin azide methyl anthraquinone derivative triggers mitochondrial-dependent cell apoptosis involving in caspase-8-mediated Bid cleavage. *Molecular cancer therapeutics,* 7**,** 1688-1697.

YAUK, C. L., LAMBERT, I. B., MEEK, M. E., DOUGLAS, G. R. & 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. *Environmental and Molecular Mutagenesis,* 56**,** 724-750.

YOON, D., WANG, Y., STAPLEFORD, K., WIESMÜLLER, L. & CHEN, J. 2004. p53 Inhibits Strand Exchange and Replication Fork Regression Promoted by Human Rad51. *Journal of Molecular Biology,* 336**,** 639-654.

ZELLER, A., TANG, L., DERTINGER, S. D., FUNK, J., DURAN-PACHECO, G. & GUÉRARD, M. 2016. A proposal for a novel rationale for critical effect size in dose-response analysis based on a multi-endpoint in vivo study with methyl methanesulfonate. *Mutagenesis,* 31**,** 239-53.