Assessing dose metrics in *in vitro* cell assays to improve *in vitro* *in vivo* dose extrapolations

Floris Groothuis, Bsc

Master student Toxicology an Environmental Health, Utrecht University

Student number: 3132099

Thesis project at IRAS, Utrecht University. July-August 2012

Supervisors: Dr. N.I. Kramer

Prof. B. Blaauboer

Abstract

High requirements and challenges formed by legislations like REACH and the 7th amendment of cosmetics have accelerated the development of new alternative toxicology testing methods. *In vitro* cell systems combined with *in silico* methods have been deemed good alternatives for toxicology testing, aiming to reduce or even replace conventional animal toxicity experiments. Unfortunately these methods are not ready to replace animal based toxicity assays yet for several reasons. One of the issues as mainly discussed in this thesis, is the lack of a standardised dose metric for use in dose-response relationships that are to be extrapolated f om *in vitro* cell systems to *in vivo*. Often total or nominal concentrations are used to express *in vitro* derived toxicity while these do not account for possible reductions in bioavailability, thus reducing the biologically effective dose through evaporation, binding to well plastic, serum, cell membranes etc. Differences in exposure media between *in vitro* assays and between *in vitro* and *in vivo*, may cause differences in the biologically effective dose even though nominal concentrations are equal. This can subsequently lead to different predictions of *in vivo* responses. Therefore this thesis discusses numerous alternative dose metrics available that may be used to improve the *in vivo* predictions. It is recommended to first choose the right external dose metric, either nominal or freely dissolved based on the physiological parameters of the test compound. Additionally, a choice can be made for an internal concentration, closer to the target site to improve the extrapolation of *in vitro* effect concentrations to equivalent *in vivo* doses. Finally the chosen metric can be integrated (AUC) or weighted (TWA) in the case of prolonged exposure and depending on the mechanisms of action. Further research needs to focus on whether internal concentrations are truly worth measuring and what cut-off value in bioavailability reduction should be used to choose between either free or nominal concentrations.

Abbreviations: ADI (accepted daily intake), ADME (absorption distribution metabolism and excretion), AUC (area under the curve), TWA (time weighted average), BED (Biologically effective dose), LOAEL (lowest observed effect level), NOAEL (no observed effect level), ITS (Integrated testing strategy), MeOA (mechanism of action), MoOA (mode of action), NRC (US National Research Council), PBPK (pharmacokinetic (modelling)), QSAR (Quantitative structure activity relationship), QIVIVE (quantitative *in vitro in vivo* extrapolation)

Introduction

The regulation of the EU under the name REACH, Registration, Evaluation, Authorisation and Restriction of Chemicals[1] have been estimated by Hartung and Rovida (2009), to require more than 54 million animal vertebrates and over 9 billion Euros in a best case scenario[2]. The estimations calculated for a ten year period, were rejected by the European Chemical Agency in a press release[3]. However, other earlier external estimates also predicted higher requirements than the 9 million animals needed as initially thought[4, 5]. In any case, when considering the probable underestimated costs, the challenge is clearly very large to live up to, and alternative (high throughput) methods will need to be developed for chemical testing and legislation. REACH, together with other regulations like the 7th amendment for cosmetics[6], strengthened the call to reduce traditional animal toxicity testing.

The discussion about the use of non-animal alternatives in toxicology is long on going. In 1959 Russel and Burch presented the concept of the three R's, the Replacement, Reduction and Refinement of animal tests[7], which was mostly based on ethical concerns. Currently, the call for alternatives is not only driven by animal welfare considerations, but also largely by the shortcomings of the conventional methods to cope with the large amount of chemicals that need to be tested. Furthermore, there are some scientific arguments to look for alternatives. As an example, there are uncertainties in toxicity prediction towards low-exposed humans, based on the toxicity observed in a few animals with differing toxicokinetic and toxicodynamic profiles. Such arguments were reiterated in the well-received report by the U.S. National Research Council (NRC) titled "Toxicity testing in the 21st century: A vision and a strategy" [8].

Unfortunately, current alternative methods such as *in vitro* cell assays and physiologically based pharmacokinetic (PBPK) models, or a combination of these, are not ready to replace the animal based testing methods for a number of reasons. One reason is the uncertainty associated with extrapolating *in vitro* effect concentrations to equivalent toxic doses in animals and humans. Regularly, the dose metric used to construct dose-response relationships *in vitro,* is simply the amount of compound added divided by the volume of the exposure medium, often referred to as the nominal concentration. This dose metric may greatly overestimate the actual effect

concentration, because no account has been made for a possible reduction in biologically effective dose (BED) due to chemical binding to e.g. (serum) protein, lipids, cell membrane, and well wall plastic[9-14](figure 1). The BED refers to the bioavailable concentration that causes the toxic effect[15]. This is an important concept to consider because the fraction of a chemical that is bound, is not able to exert its toxic action[14]. In addition, a further reduction of the BED is possible through loss of other processes like evaporation and metabolism[16, 17]. Some chemicals, with pyrene as an example, have been found to bind over 99% in a typical *in vitro* setup with only 5% serum[18].

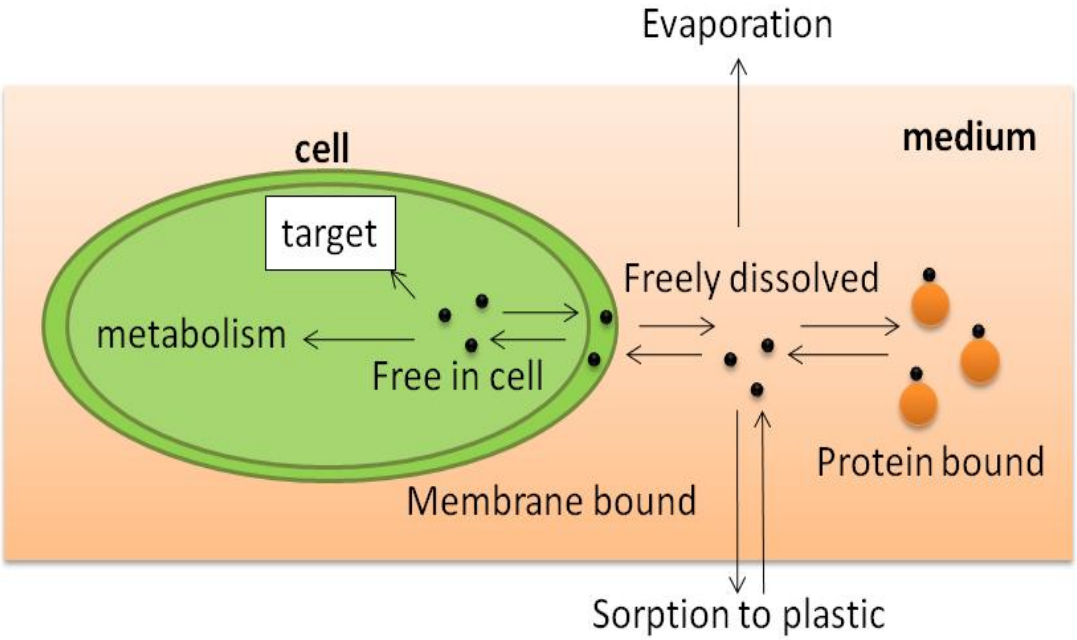

Figure 1. Representation of the processes involved influencing the bioavailability of a test compound. Reproduced with permission from Kramer *et al.*[19]

For the quantitative *in vitro in vivo* dose extrapolations (QIVIVE) required to replace the conventional animal based toxicity testing, a standardised, relevant toxic dose metric will be required for *in vitro* assays. This thesis therefore reviews different dose metrics applied *in vitro* and *in vivo*, and what types should be preferred for quantitative dose extrapolation purposes. For a broader perspective, mechanisms based toxicity testing is specified and conventional and alternative toxicity methods are also discussed.

The new mechanism based toxicity testing

In the final report of the NRC vision about toxicity testing in the 21st century, a paradigm was presented to (1) achieve broad coverage of chemicals, mixtures, outcomes and life stages, (2) reduce the cost and time required for toxicity testing, (3) develop a more robust mechanistically basis for assessing health effects of environmental chemicals, and (4) minimize the use of animals in testing. Furthermore the committee foresaw that an approach using modern tools and methodologies can be used for these four main components: chemical characterisation, toxicity pathways and targeted testing, dose-response relationships and extrapolation modelling, population-based and human exposure data[8].

To implement this new alternative toxicity testing scheme, it is thus believed risk assessment of chemicals should at least in part be based on the mechanisms of toxic action[8, 20, 21]. Modern tools would include Quantitative Structure Activity Relationships (QSARs) to characterize chemicals with their most important mechanisms and modes of action. To be more precise, a toxic mechanism of action (MeOA) refers to the biochemical process or interaction resulting in (adverse) effects[22]. These adverse effects of biological and physiological changes are commonly referred to as the mode of action (MoOA) [22]. One of the most important reasons to use the MeOA for risk assessment is that according to the NRC, knowledge about the mechanisms will be critical for accurate, quantitative prediction of hazardous exposure levels *in vivo*[8]. For example, the mechanisms will give information about the initial location of the adverse effects and the type of reaction, which is also important for dose metric decisions. If limited information is available on the mechanisms, the MoOA describing small, initial effects is possibly the best surrogate for MeOA. In fact, both will be necessary for dose extrapolations because an effect or MoOA is to be predicted based on the location (e.g. organ, cell type) and MeOA.

The range of mechanisms through which a compound can act is very broad. Chemicals can simply partition and accumulate in membranes, causing non-specific baseline toxicity, often referred to as narcosis[22, 23]. Baseline toxicity can be relatively easy predicted for neutral industrial organic chemicals by using their solubility and QSARs[24]. If the chemical exerts a stronger effect than predicted for baseline toxicity,

the chemical also has a more specific MeOA through for example receptor binding, ionic interactions or the formation of covalent bonds. Before use in *in vitro* toxicity testing and predictive computational models, it is important that the MeOA is known because it dictates where in the body the adverse effect occurs. This has an impact on the choice of *in vitro* toxicity assays and the dose metrics that should be used as will be discussed in more detail later.

Alternatives as replacement for animal based toxicological methods.

In vivo methods

In vivo methods still dominate toxicological testing for chemical risk assessment purposes. Animals are used because they contain the complex mixture of tissues, distribution pathways and physiological reactions like humans. The processes and tissues present in animals and humans cannot be fully duplicated with *in vitro* toxicity assays and importantly, a lot of animal data and models are accepted for toxicity testing[25]. As an example, numerous standardized guidelines exist accepted by the Organisation of Economic Cooperation and Development (OECD) to test the potency of a chemical in various *in vivo* endpoints like eye irritation, skin sensitization, carcinogenicity and estrogenicity[26-29]. Most tests result in dose-response relationships from where a no- or lowest observed effect level (NOAEL or LOAEL) is derived. These measures are subsequently used to set safe dose levels after additional safety corrections for e.g. inter and intra-species differences. A good example of a conventional animal toxicity test is the rat uterotrophic assay[29-31]. This test utilises either adult ovariectomized rats or immature female rats that are dosed with a potential estrogenic compound. Dosage can be for example orally, or subcutaneous and is done for three days with 24h intervals. After another 24h the rats are sacrificed and their uteruses are weighed. A dose-response curve can be constructed using the dose and percentage increased uterus weight over controls (not dosed). The percentage increase in uterus weight is a measure for estrogenicity (effect) of a compound and can be compared to references like 17β-estradiol or 17α-ethinylestradiol. Based on such a dose-response, threshold values are established like maximum accepted daily intakes (ADI).

Unfortunately, such conventional *in vivo* methods are costly in terms of laboratory animals, labour and finances, while throughput remains low. Ethical concerns have been the initial driving force since the 1960s to reduce, refine and replace laboratory animals in toxicity testing[7]. As mentioned in the introduction, there are also numerous scientific objections against the use of animal testing. For example, often the animal and human responses show some agreement, but there are many cases where the *in vivo* models fail to predict toxicity effects in humans accurately[32-34]. When making use of inbred strains and only a single gender, there is no proper reflection of natural variation. In addition, statistical power is frequently low due to limited animal numbers and exposure scenarios are in many cases unrealistically high[20, 35]. One of the major issues is that species differences, in absorption, distribution, metabolism, excretion (ADME) and toxicity characteristics, are a strong cause of uncertainty in the predictive outcome towards human toxicity. *In vivo* methods are also inadequate to study the MeOA and interactions of a chemical. This is also evident from the example of the rat uterotrophic assay as described above. The animal can be considered a black box that only gives a specific response as seen in the endpoint. However, the mechanisms occurring towards this specific endpoint are not known.

In vitro as part of alternative methods

Alternative methods for toxicology testing describes a combination of tools that can be utilised as alternative for animal models. Examples of alternatives are *in vitro* cell assays and *in silico* tools like QSARs and PBPK modelling. A small list with examples of alternative methods are presented in table 1. In particular, *in vitro* assays form the centre of alternative methods as they also provide the input for *in silico* tools. *In vitro* refers to non-animal based tests, typically in glass or plastic vessels in the laboratory. An overview of numerous *in vitro* (cell based) culture tests adapted from Bhogal *et al.* can be found in table 2[36].

Unlike most animal based models, *in vitro* tools are well suited to study the underlying mechanisms of toxicity of test chemicals[20, 37]. *In vitro* cell assays can be used to measure disturbances in toxicity pathways, like estrogen receptor activation, and construct dose-response relationships, while other assays can be used to clarify

the chemical properties needed for computational input. As an example, QSAR-predicted physicochemical properties and *in vitro* derived effect concentrations may

Table 1. Examples alternative methods for toxicological risk assessment

Alternative methods	Description
***In vitro* cell and tissue culture (toxicity) assays**	Cell or tissue based *in vitro* assays including reporter gene assays, embryonic stem cell test, 3T3 NRU phototoxicity test, isolated skin etc. See also table 2 for a broader overview. These assays can be used to establish dose-response relationships for relevant cell/tissue types.
***In vitro* characterisation assays**	A term for *in vitro* test systems that measures chemical characteristics like hydrophobicity or metabolism. These are assays to estimate ADME characteristics. An example would be using proteins, microsomes or liver cells to measure intrinsic clearance and metabolite formation. Another example is the octanol-water partitioning test to measure hydrophobicity (Log_{kow}).
Omics	Tests changes in omics e.g. Metabolomics, proteomics and is based on the consensus that a toxicant changes the state of the metabolome, proteome etc. Can be used as sensitive biomarkers for disturbances in toxicological pathways and early detection of (mechanisms) of toxicity
Quantitative Structure Activity Relationship (QSAR)	Models that attempt to relate 3D structure, solubility, molecular volume and intrinsic reactivity to toxic potential in several areas like mutagenic, skin etc. With these models one can find possible MeOA and MoOA.
Phyiologically Based Pharmacokinetic Models (PBPK)	Models the (internal) target dose in relation to external amounts. Basically it models ADME, based on literature and computational results, leading to dose estimations in different compartments of the body. These concentration estimations are then used for risk assessment.
Pharmacodynamic Models	Models the effect over time of a chemical based on for example the mechanisms. These models can e.g. be used to estimate if irreversible or accumulative damage will occur during a certain exposure period leading to adverse effects.

subsequently serve as input into PBPK models to simulate human equivalent toxic doses. In one study, scientists extrapolated dose-response relationships and EC_{10} values for several glycol ethers to rats and humans using a PBPK model[38]. Interestingly, they did not only measure toxicity of the parent compound, but also of the toxic metabolites in the *in vitro* embryonic stem cell test, which served as endpoint for developmental toxicity. The doses at 10% of the maximum effects (EC_{10}) obtained in the *in vitro* embryonic stem cell test were used as input in a PBPK model. These EC_{10} values provided the *in vivo* internal peak dose to predict the related external dose of the parent compound with the PBPK model. Differences between modelled and measured external EC_{10} of the parent compound in rats varied between a factor 0.2-6. As discussed by the authors, one of the weaknesses of the extrapolation was the *in vitro* embryonic stem cell test endpoint of cardiomyocyte differentiation, not reflecting the broader *in vivo* developmental

endpoint accurately. A similar study, based on the same *in vitro* assay also found they could predict *in vivo* effect levels quite well with the exception of one compound[39]. These studies show the potential of QIVIVE and toxicity testing based on alternative methods in the near future.

Table 2. Advantages and Disadvantages of several types of tissue culture systems *in vitro*. Adapted from Bhogal *et al.*[36]

System	Advantages	Disadvantages
Primary cultures	Obtainable from various target tissues; can retain in vivo tissue-specific characteristics	Short in vitro lifespan; progressively lose in vivo properties; prone to contamination
Monolayers and mono-cultures	Can be grown to confluency and subcultured; can be used as barrier models; used to quantify cell proliferation/growth; suitable for genetic manipulation	Limited interactions between cells; absence of other cell types, nervous, immune and endocrine systems
Co-cultures	Involve more than one cell type, so resemble in vivo situation more closely (e.g. blood–brain barrier)	Some cell combinations are incompatible with each other in culture; complicated/conflicting cell culture requirements
Continuous cell lines	Readily available and reproducible source of cells; avoids repeated cell isolation from animals or humans	Tend to lose *in vivo* differentiation and take on new properties induced by culture conditions; enter senescence and decline after a certain number of population doublings
Genetically engineered cell lines	Generated by transforming cells with foreign DNA; DNA can confer cell line stability; DNA might encode structural or functional proteins; used to create polymorphic cell line batteries	Techniques are specialized; methods do not always lead to permanent changes; limited potential for altering cellular features
Immortalized cell lines	Generated from human/animal cells by introducing oncogenes/telomere-controlling DNA; cells have cell line longevity but can retain tissue-type specific features	The immortalization techniques are specialized; there is not always permanent immortalization
Stem cells	Cells are able retain their stem cell capacity and to differentiate into many cell types	Limitations on cell types that can be generated; some animal species/strain limitations; ethical problems when using human embryonic stem cells
Tissue slices	Represents complexity of the organ; cellular contacts retained; useful for inter-species comparisons; many organs from same donor can be obtained; histological and biochemical tests possible; slices from different organs can be co-cultured; regional effects in same organ are particularly useful for metabolism studies	Difficult to produce reproducibly; exposure and activity of cells in slices can vary; limited *in vitro* lifespan
Organotypic cultures	Multi-layered and spatially differentiated; exhibit cellular communication; good retention of in vivo physiology; can be generated from primary/immortalized cells; proprietary models available	Correct culture conditions can be difficult to define; batch variation of propriety models; limited *in vitro* lifespan
Perfused cultures	Applicable to a variety of the systems above; perfusion restores media and removes metabolites; allows cells to grow for extended periods; high cell densities possible; long-term repeat exposure/recovery possible; can be used for whole organs (e.g. kidney)	Technically complex; high risk of contamination; only a small number of samples can be set up; limited *in vitro* lifespan
Reconstructed tissue cultures	Components can be controlled and varied according to purpose	Technically complex
Whole organs	Organ functions modelled closely; different cell types with cellular interactions; particular useful for embryo toxicity studies	Can be difficult to culture; limited culture life; must be freshly isolated; tend to require complex perfusion systems

Several advantages for *In vitro* tests exist over conventional animal methods (see also table 2). For example, they are rarely hampered by ethical concerns except when human embryonic stem cells are used. High throughput setups can increase testing speed and efficiency. Fortunately, there are (cell) models representing almost all tissue types from a range of animal species, including humans. These setups can be miniaturised and automatized, thus requiring very limited test substance, resources and costs. As alternative to *in vivo* toxicity assays, they can provide rapid screening tools, although unfortunately they cannot predict any *in vivo* outcome on their own. However, *In vitro* assays can be incorporated into integrated testing strategies (ITS) similarly as described above, to obtain *In vivo* prediction and reduce the need for animal testing. ITS refers to a tiered approach by which chemicals are tested using the chemical properties, QSARS, *in vitro* tests and other computational models like PBPK and pharmacodynamics modelling[40].

A pilot study, using ITS on ten chemicals showed that prediction of *in vivo* outcomes on several endpoints could be obtained with varying success[41]. Predictions were done using QSARs, PBPK modelling and *in vitro* assays on top of already available data. These predictions were then compared with existing *in vivo* data. The predictions for dermal toxicity and acute fish toxicity, and skin and eye irritation were quite accurate (9/10-7/10 correct). In contrast, acute oral toxicity and repeated dosing toxicity estimations were less precise (5/10 correct). Because only limited animal numbers would be required to confirm the ITS predictions, the approach could potentially reduce animal testing by about 38%. Notably some suggestions were made for future improvement of the method such as (1) incorporation and development of a method for metabolic clearance, (2) better estimation of uptake through the gastro-intestinal tract by either QSAR or *in vitro* methods, (3) incorporate estimations of binding to (plasma) proteins and interestingly, (4) development of a method to relate *in vitro* determined cytotoxic concentration to *in vivo* target tissue concentrations by taking into account intra-cellular concentrations and bioavailability, the point that is being assessed in this thesis.

Unfortunately, *in vitro* cell based assays are not only useful, they have some basic drawbacks as well. A major problem of regular single or e.g. co-culture cell assays is that they resemble the complex tissues and organs in the body poorly. This

is one of the important reasons such tests need to be combined in integrated testing strategies with other assays and computational tools to form a reliable alternative for animal studies. Rotroff *et al.* were able to improve the ToxCast screening program for prioritizing chemicals relatively easy by incorporating some ADME characteristics in the assessment[42]. The ToxCast program was launched to aim the limited resources available for toxicity testing at chemicals showing the strongest *in vitro* potency in a test battery[43]. The researchers selected 35 chemicals tested in the ToxCast program and showed they could improve the classification by incorporating metabolic clearance, plasma protein binding and actual exposure data[42]. Notably, two chemicals were now seen in the top of the prioritisation list while they were not previously. This study illustrates that prioritizing chemicals or risk assessment based on solely *in vitro* assays is risky and insufficient. By incorporation of a few ADME characteristics or more complex QSAR and PBPK models, predictions can be grealy improved. Other *in vitro* methods that can provide better *in vivo* resemblance are tissue slices or e.g. 3D cultures. However, these methods are much more complex and difficult to handle and obtain. Such assays are note very widely used and like single or co-cultures, they still require additional (dose) extrapolation steps.

The most common used *in vitro s*ingle or co-culture cell assays have some additional drawbacks that include their limited resemblance of the tissue they come from because often less than 1% of a tissue actually consists of the cells under study. The other fraction comprises extracellular matrix, other cell types and for example vascular tissue (see also table 2). In a lot of cases, the cells used are immortalised to raise their resilience in culture and ease growth and passaging[44]. Sometimes these immortalised cells resemble the original phenotype only partially[44-46]. In addition, cell cultures are grown in conditions that are different compared to the body on several points. For example, there is exchange of media, depletion and replacement of nutrients and there is optimisation for rapid growth[20]. As mentioned before, 3D cell cultures, tissue slices and other new technologies like flow through systems may be used to overcome a lot of the disadvantages listed. But again such techniques require more effort and knowledge to handle and obtain.

One issue not generally considered is the dose metric used in all types of *in vitro* assays, while it is normally essential for accurate dose-response prediction.

Often nominal concentrations are used as exposure measure and this can lead to significant bias in effect concentrations predicted *in vivo* as is discussed in the following section. To improve the predictive value, different exposure measures should be considered depending on the MeOA.

Dose metrics

In vitro dose metrics: the total and nominal concentrations

Numerous dose metrics have been used in the past for different *in vitro* applications and an overview of the major groups of metrics is listed in table 3. The most commonly known are the total and nominal dose. The nominal dose is defined as the amount of chemical added to the test, divided by the volume of the test solution, while total dose is the concentration measured in the medium. A related exposure measure seen in the nanoparticle toxicology is surface area because it better explains the activity of the nanoparticles. Recently it was shown that the correlation between *in vitro* and *in vivo* improved when using exposed surface area as dose instead of nominal concentration[47]. This is an example of a dose metric derived from one of the major classes that are listed in table 3 because it remains a total concentration measure and can also be used in a freely available concentration. Nominal or total dose is the simplest measure for quantifying dose-response relationships or other toxicity measures, directly explaining why it is so widely used. However, the nominal dose can lead to bias of the actual dose-response relationship (see also figure 2) because it may not be directly related to the biologically effective dose (BED).

Table 3. Overview of major dose metrics

Dose metric	Description
Total concentration	Total concentration measured in the assay
Nominal concentration	Total amount of chemical added to the assay divided by the volume
Freely available or bioavailable concentration	The fraction of the concentration of a test chemical in solution that is unbound or freely dissolved.
Internal cell concentration	The concentration in the cytoplasm, either free or total
Membrane concentration	The concentration in the membrane lipid fraction, either free or total
Area Under the Curve (AUC)	Integrated dosage over time of on of the metrics above
Time Weighted Average (TWA)	Averaged dosage over time of one of the metrics above

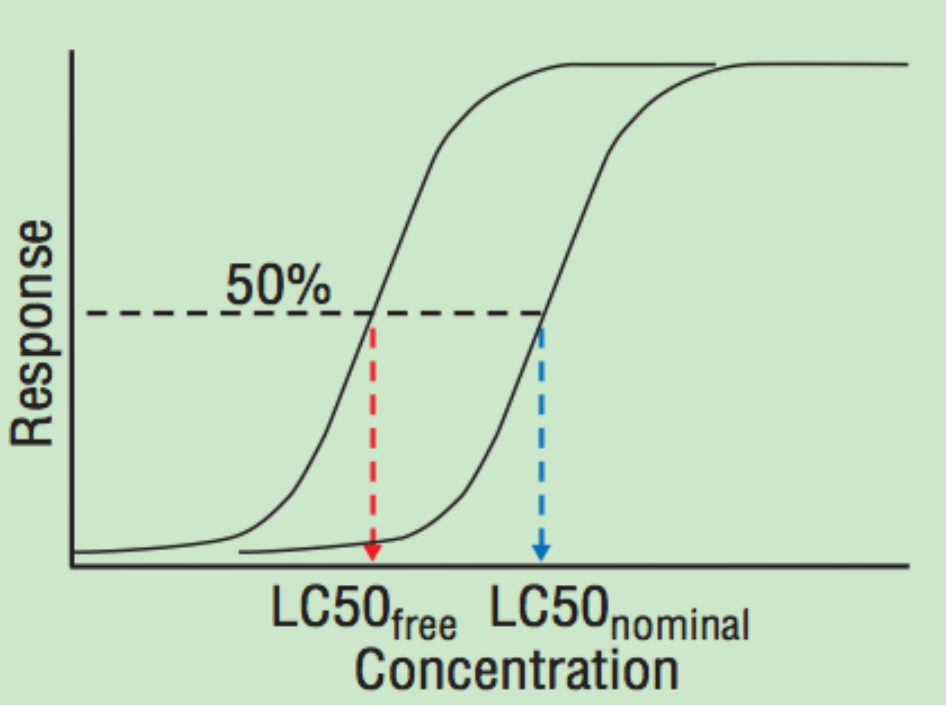

Figure 2. Representation of the difference between nominal and free concentrations and the resulting dose-responses. Adapted with permission from Escher & Hermens[14].

In vitro dose metrics: the freely available or bioavailable concentration

As aforementioned, part of the added dose can be unavailable to exert its toxic effect due to evaporation and binding to serum, cell membranes etc. In *in vitro* assays, the freely available or bioavailable concentration may be a better estimate of the BED. Heringa *et al.* showed for the first time, that the freely available concentration is a better dose metric to use for some hydrophobic chemicals in toxicological dose-response curves[9]. The nominal doses were shown to overestimate the dose-effect relationship, as part of the chemical is unavailable for toxic action (like in figure 2). In short, dose-response relationships based on the freely available concentration are independent of binding parameters to e.g. protein, lipids, cells, well plastic or loss through evaporation.

The importance for *in vitro in vivo* correlation becomes evident from a study by Gülden *et al*.[48] They were able to improve the correlation towards fish by accounting for bioavailability. Generally, fish had proven to be more sensitive than the alternative *in vitro* tests, but the relation was improved when cytotoxic free concentrations were used instead of nominal concentrations. The prediction of *in vivo* EC_{50} values improved as was also seen in the regression slope that became very close to 1 instead of 1.27-1.43. The authors concluded the sensitivity difference was

at least partly explained by bioavailability *in vitro* and *in vivo*. Remaining differences may be explained by species differences (data from different fish species was used), additional binding factors (not just lipids and albumin), and the toxic MoOA in fish that may be somewhat different compared to the *in vitro* cytotoxicity endpoint. Notably, they used a model that was not validated in this study with measurements to calculate the freely available concentration.

When freely available concentrations cannot be measured due to e.g. limited resources, they can alternatively be modelled as becomes evident from the study described above. Recently Kramer *et al.* modelled and measured the freely available concentration of phenanthrene in two cytotoxicity assays based on the processes as schemed in figure 1[19]. The results indicated that the free concentration of phenanthrene, a relatively volatile and hydrophobic compound, is significantly reduced in a typical *in vitro* setup through binding to matrices such as serum protein and well plate plastic. A reduction in free concentration due to increasing serum protein levels is accompanied by an increase in the median effect concentration (EC_{50}) based on nominal concentrations like in figure 2. This effect was modeled without an evaporation parameter, using the partition coefficients of the compound to assay binding components. The model mostly overestimated the effect compared to the measurements by 1-2 fold and was more accurate with a higher serum concentration. One explanation of this bias is that the model assumes a closed system, while there is no airtight system in reality. The study shows modeling is possible, although not always accurate and it remains to be tested on other types of chemicals to check if it is broadly applicable. Thus such an approach will require some additional development before it is truly useful for quantitative predictions. Nonetheless, it may be better to use a modeled free concentration for dose-effect relationships compared to nominal concentrations, where much larger biases can be expected.

Another study by Gülden *et al.* extrapolated *in vitro* EC_{50} values to equivalent *in vivo* concentrations by correcting for binding factors such as albumin using again a model[49]. Unfortunately this attempt could not be validated because of the absence of human data, but the aim was more to increase the understanding of some processes involved determining extrapolated EC_{50} values. They found that for

compounds with low lipophilicity (log_{Kow}<2.0), bioavailability was not affected in both *in vitro* and *in vivo* serum. For more lipophilic compounds, (Log_{kow}>2.0) additional factors like partitioning to lipids have to be taken into account, to successfully estimate equivalent serum concentrations. They also derived some general rules: if the cytotoxic concentration approaches or is higher than the binding capacity of albumin, the impact of protein binding becomes negligible. In other words, the higher the cytotoxic potency with a threshold of EC_{50}<1000 µM, the lower the deviation between direct EC_{50} values and EC_{50} equivalent serum values (corrected for binding factors). Notably no account was made for other binding proteins apart from albumin or other factors that decrease bioavailability such as well plastic binding and evaporation. Additionally the model was again not validated by measuring the freely available concentrations.

Techniques to measure and stabilise the bioavailable concentrations

Although use of the freely available concentration will lead to more accurate dose-effect relations for a lot of chemicals as is evident from the studies described above, it is not widely used due to the difficulty and required effort needed to measure freely dissolved concentrations. As alternative it is also possible to model such concentrations as done in the studies by Gülden *et al.*[48, 49] To measure protein or serum binding, several widely applied techniques exist such as equilibrium dialysis or ultrafiltration. Unfortunately these techniques require significant effort and time to perform and as alternative, a limited number of laboratories apply solid phase micro extraction (SPME) to measure freely available concentrations. This technique uses a small solid phase, often a fibre, to which chemicals can bind in the medium. The amount bound to the solid phase fibre will reach an equilibrium state with the amount remaining freely available in the solution. If the ratio between the fibre concentration and the solution is preliminary determined, one can calculate the freely dissolved concentration from the measured fibre concentration. Because the volume of a typical solid phase fibre is small, it also has the advantage not to alter the concentration in the water. The extraction can therefore take place during the experiment if less than 5% of the chemical present is extracted. This variation of the SPME technique is called negligible SPME or nd-SPME[50].

Freely available concentrations have the additional advantage they can still be used when there is a decline of concentration in the test system due to e.g. evaporation or metabolism. The freely available concentration remains useful in such a scenario because it can be kept constant using a partition controlled dosing system[51]. Such a system works by loading the test chemicals on a solid (polymer) phase that like SPME phases will obtain equilibrium with the test solution. The solid phase can be used as reservoir to replenish the loss of chemicals through e.g. evaporation. In short, this is a useful technique to keep the freely available concentration constant and obtain a more robust dose-response relationship for chemicals that bind, evaporate or are metabolized.

Prolonged exposure metrics

In case of prolonged exposure, the area under the curve (AUC) or sometimes time-weighted average (TWA) is used. Time weighted averages are total added doses divided by the time period of dosing and is often used in carcinogenic risk assessment, as well as dose-response relationships to estimate lifetime risks[15]. The AUC is the integrated dose over time and is also used for prolonged exposures because it is more valuable than a numerical estimate of exposure[15]. AUC and TWA measures can be based on total concentrations, but also on different dose metrics. Gülden *et al*. used AUC measures based on nominal concentrations to quantify exposure of changing H_2O_2 concentrations over time[17]. The aim of the study was to study the impact of different cell concentrations in the assay. While the AUC was kept constant with different tests, the EC_{50} values varied due to different metabolic capacity of varying amount of cells in the tests. More cells meant higher metabolic capacity and lower EC_{50}'s, directly illustrating one factor that can have an influence on the dose-response relationship: cell concentration. It is an important factor that is normally overlooked when using only nominal concentrations, unless specifically investigated. The AUC measure in this study was very effective because the H_2O_2 toxicity can be viewed as mostly irreversible, resulting in an increase in effect over time. Therefore this study also illustrates the effectiveness of the AUC with reactive chemicals for short (<24h) time periods. This is therefore also an example on how the MeOA can be used to determine the right dose metric. When reactive chemicals

are tested causing (irreversible) accumulative damage, an integrated or weighted measure should be used. For other chemicals like baseline toxicants, the AUC measure would be less informative because one can expect an effect only after a peak exposure above a threshold value. In such a case, the exposure history beneath the threshold is not very important meaning that an AUC or TWA measure is not useful.

In vivo dose metrics, target dose and internal concentrations

Like *in vitro* systems, similar problems have been encountered for exposure metrics *in vivo*. The *in vivo* dose metrics are visualised in figure 3 and have analogue processes as illustrated for *in vitro* in figure 1. Just like in the field of *in vitro* toxicology, often external nominal concentrations are used. In other words, the concentration is derived from the amount of compound an animal or human is exposed to divided by its weight. Alternatively, the concentration in external media can be used in e.g. aquatic systems with fish[52]. Analogue to the *in vitro* experiments, this concentration does not account for binding and other loss parameters. In this case these parameters are whole blood binding, lipid binding, metabolism, but also absorption parameters and so on. Thus in fact, the actual concentration at the target site, i.e. the site where the chemical exerts its toxic effect through some mechanism is not known. As alternative for nominal external concentrations it is possible to use internal concentrations or internal residues like blood concentrations or tissue concentrations. The difficulty with tissue residues is that the extraction of tissue for concentration measurements remains delicate work[22]. However, the internal residue concept exists for over a century and can be considered a better dose surrogate for the concentration at the site of toxic action[52]. This relates again to the concept that the concentration at the target site would explain the effects best as there are no factors influencing the relationship like bioavailability, absorption parameters etc. Unfortunately such a concentration is almost always not measurable. Therefore best surrogates are internal residues, as close as possible to the site of toxic action.

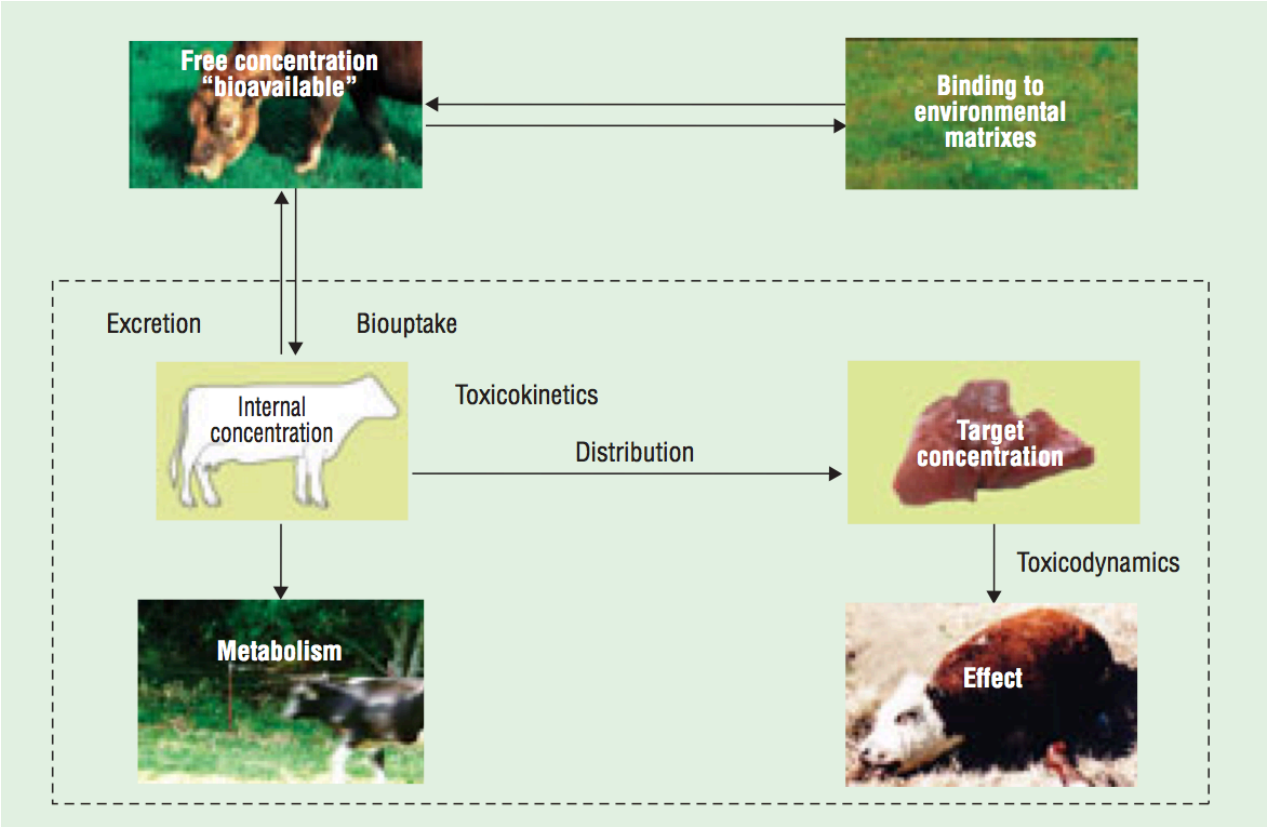

Figure 3. Schematic representation of dose metrics *in vivo*. The figure is a parallel to the dose metrics *in vitro* as can also be seen in figure 1. Figure Adapted from Escher and Hermens[14].

For risk assessment utilising alternative methods based on the mechanisms and small modes of action, it is important to consider these *in vivo* dose metric concepts. In parallel with *in vivo* metrics, freely dissolved concentrations in *in vitro* tests still remain an *in vitro* external dose metric and a surrogate to the concentration at the site of toxic action. Again in theory, the closer the dose metric is to the target site, the stronger the relationship between the dose and effect. For example, if the chemical exerts its toxic effect inside the cell cytoplasm, the freely available concentration in the cytoplasm of the cell should predict the response better due to a lower amount of confounding factors such as cell absorption and binding by proteins inside the cytoplasm. Similarly, the concentration in the cellular membrane would predict effects better if the compound exerts its effect in the cell membrane, as is the case for the baseline toxicants mentioned before. Although possible, the internal concentrations require substantial effort to measure and the best alternative surrogate would still be bioavailable external concentration. This

freely available concentration is also a good alternative if the mechanisms and perhaps small modes of action are unknown.

Conclusion and Recommendations

Where conventional techniques require too much time and effort to cope with the new testing demands of regulations such as REACH, alternatives need to take their place. However, these alternative methods require some standardisation on dose metrics before broad implementation. This thesis aimed to put different dose metrics used in toxicology into perspective for dose extrapolation purposes.

This document explained some of the most important doses used *in vitro* and *in vivo* as they provide insight on the issues and possible solutions. Essentially, the best dose metric would be the concentration at the target site where the most important and sensitive toxic MeOA of the test chemical is located. Because such a concentration is very difficult to obtain, surrogates are available of which internal concentrations would be the closest achievable dose metric. Examples are membrane concentration for baseline toxicants or MeOA in the membrane, and cytoplasm or even nuclear/organelle concentrations for specific MeOA and MoOA inside the cell. However, obtaining an internal concentration for all compounds also requires substantial effort. Another more feasible alternative would be to use an external dose surrogate such as nominal concentrations. Such a dose metric can only be used however, when there is evidence the chemical is not bound significantly to anything in the test system, and has limited evaporation. In case the bioavailability is reduced, freely available concentrations should be used. It must be clear however, that these dose metrics will explain the effect worse with wrong assumptions and with being further away from the target site.

Although it is quite certain internal concentrations will predict the MoOA best, it has not been investigated, and for dose extrapolation purposes it may not be worth it. Additionally, with an internal dose extrapolated, there is still a need to calculate what the concentration outside the cell is for risk assessment purposes. Therefore, internal concentration should not be used on their own but in combination with an external dose metric. In this case, the external concentration will give information about the proportion going inside the cell, while the internal

concentration forms the basis for the dose-effect relationship. Whether it is truly worth to use internal concentrations as dose metric on top of an external exposure measures remains to be investigated.

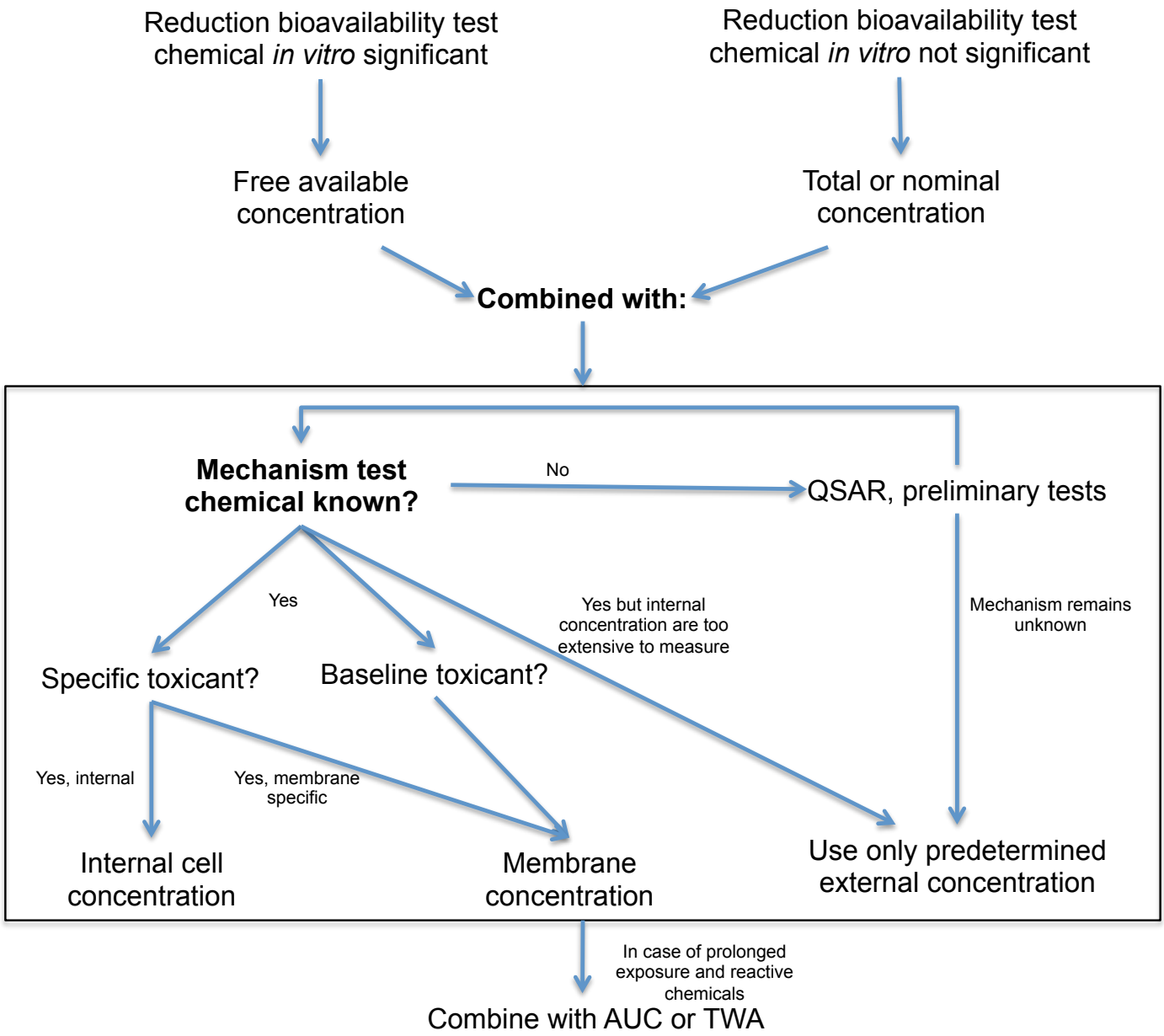

Figure 4. Scheme for dose metric choice. First a choice should be made for an external dose metric on bioavailable or nominal basis. Additionally, the metric can be combined with an internal concentration based on the mechanisms of action if possible. Finally the metric can be integrated or averaged in case of prolonged exposure and e.g. irreversible mechanisms.

A small scheme is proposed here to make the choice for the right dose metric easier (figure 4). Initially a choice should be made for an external dose metric, such as freely available or nominal concentration based on the information available on the physiological parameters of the test compound. Afterwards an internal

concentration can be considered if the tools are available and the MeOA is known. For example, if the chemical is a baseline toxicant, acting through accumulation in the cell membrane causing narcosis, membrane concentrations can be considered. Notably the importance of such an additional measurement remains to be investigated. Finally, the dose metrics can be combined in prolonged metrics such as AUC or TWA if applicable considering the MeOA. As discussed before, AUC or TWA measures should be used in case of reactive or specifically acting chemicals causing (irreversible) accumulative damage.

The threshold in bioavailability reduction is an arbitrary and political choice. In terms of improving the dose effect relation and extrapolation to the best we can, an allowance of 5% reduction in bioavailability could be used. However, in terms of changes in the dose effect prediction *in vivo,* a bioavailability of up to 50% in the *in vitro* test may still have little impact. Gülden *et al.* showed using modelled bioavailable concentrations, that chemicals with a Log_{kow} below 2.0 do not need to be corrected as it has little impact on the dose-response[49]. Furthermore, for high potency chemicals (EC_{50}<1000 µM) binding may become negligible. Thus the threshold value of Log_{kow}2.0 is only a guide and should not be used solely. As another example, the bioavailability of polar test compounds may still be reduced in e.g. an *in vitro* test where metabolism occurs. Furthermore there are other parameters that could have an impact on bioavailability like protein binding (K_d) and pKa or charge of the compound for e.g. cell uptake. Although true threshold values remain to be established, it is clear that possible reductions in bioavailability need to be considered, and if there is any doubt, the freely available concentration should be used.

In conclusion two main factors determine the choice for the right dose metric in an *in vitro* assay. (1) The physiochemical properties of the chemical (e.g. Log_{kow}, K_d, pKa), and (2) the MeOA measured in the particular assay set-ups. Consideration of these factors is important to prevent bias in estimated dose-response relationships and extrapolation attempts. To help choose the right dose metrics for a particular *in vitro* assay with the purpose of establishing dose-response relationships the scheme can be followed shown in figure 4. Whether internal concentrations are truly worth measuring for use in dose extrapolations remains to be investigated. In addition, a

bioavailability cut-off value needs to be established to choose either freely available, or nominal concentrations as external dose metric.

Acknowledgements

I would like to thank Dr. Nynke Kramer for supervising this thesis, her comments and her input in our discussions.

References

1. European Commission, E. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006. concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Official Journal of the European Union* **L136**, 3-280 (2006).

2. Hartung, T. & Rovida, C. Chemical regulators have overreached. *Nature* **460**, 1080-1081 (2009).

3. ECHA. New study on the number of test animals for REACH. *Press Release ECHA/PR/09/11* (2009).

4. Hofer, T. *et al*. Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Arch. Toxicol.* **78**, 549-564 (2004).

5. Breithaupt, H. The costs of REACH. REACH is largely welcomed, but the requirement to test existing chemicals for adverse effects is not good news for all. *EMBO Rep.* **7**, 968-971 (2006).

6. The European Parliament and the Council of the European Union.
Directive 2003/15/EC of the European Parlia- ment and of the council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products.
http://ec.europa.eu.proxy.library.uu.nl/consumers/sectors/cosmetics/documents/directive/ . (2003).

7. Russel, W. S. & Burch, R. L. The principles of humane experimental technique. *Methuen London* (1959).

8. National Research Council, Committee on Toxicity Testing and Assessment of Environmental Agents.
Toxicity testing in the 21st century: A vision and a strategy. *The National Academies Press*, 1-216 (2007).

9. Heringa, M. *et al*. Toward more useful in vitro toxicity data with measured free concentrations. *Environ. Sci. Technol.* **38**, 6263-6270 (2004).

10. Hestermann, E., Stegeman, J. & Hahn, M. Serum alters the uptake and relative potencies of halogenated aromatic hydrocarbons in cell culture bioassays. *Toxicol. Sci.* **53**, 316-325 (2000).

11. Gulden, M., Morchel, S. & Seibert, H. Factors influencing nominal effective concentrations of chemical compounds in vitro: cell concentration. *Toxicol. Vitro* **15**, 233-243 (2001).

12. Seibert, H., Morchel, S. & Gulden, M. Factors influencing nominal effective concentrations of chemical compounds in vitro: medium protein concentration. *Toxicol. Vitro* **16**, 289-297 (2002).

13. Schreiber, R., Altenburger, R., Paschke, A. & Kuester, E. How to deal with lipophilic and volatile organic substances in microtiter plate assays. *Environ. Toxicol. Chem.* **27**, 1676-1682 (2008).

14. Escher, B. & Hermens, J. Internal exposure: Linking bioavailability to effects. *Environ. Sci. Technol.* **38**, 455A-462A (2004).

15. Paustenbach, D. The practice of exposure assessment: A state-of-the-art review (Reprinted from Principles and Methods of Toxicology, 4th edition, 2001). *J. Toxicol. Env. Health-Pt b-Crit. Rev.* **3**, 179-291 (2000).

16. Riedl, J. & Altenburger, R. Physicochemical substance properties as indicators for unreliable exposure in microplate-based bioassays. *Chemosphere* **67**, 2210-2220 (2007).

17. Guelden, M., Jess, A., Kammann, J., Maser, E. & Seibert, H. Cytotoxic potency of H2O2 in cell cultures: Impact of cell concentration and exposure time. *Free Radic. Biol. Med.* **49**, 1298-1305 (2010).

18. Kramer, N. I., van Eijkeren, J. C. H. & Hermens, J. L. M. Influence of albumin on sorption kinetics in solid-phase microextraction: Consequences for chemical analyses and uptake processes. *Anal. Chem.* **79**, 6941-6948 (2007).

19. Kramer, N. I., Krismartina, M., Rico-Rico, A., Blaauboer, B. J. & Hermens, J. L. M. Quantifying Processes Determining the Free Concentration of Phenanthrene in Basal Cytotoxicity Assays. *Chem. Res. Toxicol.* **25**, 436-445 (2012).

20. Hartung, T. From alternative methods to a new toxicology. *Eur. J. Pharm. Biopharm.* **77**, 338-349 (2011).

21. Blaauboer, B. J. Biokinetic Modeling and in Vitro-in Vivo Extrapolations. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews* **13**, 242-252 (2010).

22. Escher, B. I. *et al*. Crucial role of mechanisms and modes of toxic action for understanding tissue residue toxicity and internal effect concentrations of organic chemicals. *Integrated environmental assessment and management* **7**, 28-49 (2011).

23. KONEMANN, H. Quantitative Structure-Activity-Relationships in Fish Toxicity Studies .1. Relationship for 50 Industrial Pollutants. *Toxicology* **19**, 209-221 (1981).

24. Mackay, D. *et al*. The physicochemical basis of QSARs for baseline toxicity. *SAR QSAR Environ. Res.* **20**, 393-414 (2009).

25. Hartung, T. & Daston, G. Are in vitro tests suitable for regulatory use? *Toxicol. Sci.* **111**, 233-237 (2009).

26. OECD. OECD Guideline for testing of chemicals section 4: Health effects. Guideline 405: Eye Irritation. *Organisation for Economic Cooperation and Development, Paris, France* (2002).

27. OECD. OECD Guideline for testing of chemicals section 4: Health effects. Guideline 406: Skin sensitisation. *Organisation for Economic Cooperation and Development, Paris, France* (1992).

28. OECD. OECD Guideline for testing of chemicals section 4: Health effects. Guideline 451: Carcinogenicity Studies. *Organisation for Economic Cooperation and Development, Paris, France* (2009).

29. OECD. OECD Guideline for testing of chemicals section 4: Health effects. Guideline 440: Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties. *Organisation for Economic Cooperation and Development, Paris, France* (2007).

30. Kanno, J. *et al*. The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: Phase 1. *Environ. Health Perspect.* **109**, 785-794 (2001).

31. Kanno, J. *et al*. The OECD program to validate the rat uterotrophic bioassay. Phase 2: Dose-response studies. *Environ. Health Perspect.* **111**, 1530-1549 (2003).

32. Basketter, D., York, M., McFadden, J. & Robinson, M. Determination of skin irritation potential in the human 4-h patch test. *Contact Derm.* **51**, 1-4 (2004).

33. Shanks, N., Greek, R. & Greek, J. Are animal models predictive for humans? *Philos. Ethics Humanit. Med.* **4**, 2 (2009).

34. Olson, H. *et al*. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.* **32**, 56-67 (2000).

35. Rietjens, I. M. C. M. & Alink, G. M. Future of toxicologys - Low-dose toxicology and risk - Benefit analysis. *Chem. Res. Toxicol.* **19**, 977-981 (2006).

36. Bhogal, N., Grindon, C., Combes, R. & Balls, M. Toxicity testing: creating a revolution based on new technologies. *Trends Biotechnol.* **23**, 299-307 (2005).

37. Eisenbrand, G. *et al*. Methods of in vitro toxicology. *Food Chem. Toxicol.* **40**, 193-236 (2002).

38. Louisse, J. *et al*. The Use of In Vitro Toxicity Data and Physiologically Based Kinetic Modeling to Predict Dose-Response Curves for In Vivo Developmental Toxicity of Glycol Ethers in Rat and Man. *Toxicol. Sci.* **118**, 470-484 (2010).

39. Verwei, M., van Burgsteden, J. A., Krul, C. A. M., van de Sandt, J. J. M. & Freidig, A. P. Prediction of in vivo embryotoxic effect levels with a combination of in vitro studies and PBPK modelling. *Toxicol. Lett.* **165**, 79-87 (2006).

40. Blaauboer, B. The applicability of in vitro-derived data in hazard identification and characterisation of chemicals. *Environ. Toxicol. Pharmacol.* **11**, 213-225 (2002).

41. Gubbels-van Hal, W. *et al*. An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul. Toxicol. Pharmacol.* **42**, 284-295 (2005).

42. Rotroff, D. M. *et al*. Incorporating Human Dosimetry and Exposure into High-Throughput In Vitro Toxicity Screening. *Toxicological Sciences* **117**, 348-358 (2010).

43. Dix, D. J. *et al*. The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* **95**, 5-12 (2007).

44. Astashkina, A., Mann, B. & Grainger, D. W. A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol. Ther.* **134**, 82-106 (2012).

45. Donato, M. T., Lahoz, A., Castell, J. V. & Gomez-Lechon, M. J. Cell lines: A tool for in vitro drug metabolism studies. *Curr. Drug Metab.* **9**, 1-11 (2008).

46. Allen, D. *et al*. Cell lines as in vitro models for drug screening and toxicity studies. *Drug Dev. Ind. Pharm.* **31**, 757-768 (2005).

47. Han, X. *et al*. Assessing the relevance of in vitro studies in nanotoxicology by examining correlations between in vitro and in vivo data. *Toxicology* **297**, 1-9 (2012).

48. Gulden, M. & Seibert, H. Impact of bioavailability on the correlation between in vitro cytotoxic and in vivo acute fish toxic concentrations of chemicals. *Aquat. Toxicol.* **72**, 327-337 (2005).

49. Gulden, M. & Seibert, H. In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent to cytotoxic concentrations in vitro. *Toxicology* **189**, 211-222 (2003).

50. Heringa, M. B. & Hermens, J. L. M. Measurement of free concentrations using negligible depletion-solid phase microextraction (nd-SPME). *Trac-Trends in Analytical Chemistry* **22**, 575-587 (2003).

51. Kramer, N. I. *et al*. Development of a Partition-Controlled Dosing System for Cell Assays. *Chem. Res. Toxicol.* **23**, 1806-1814 (2010).

52. McCarty, L. S. *et al*. Advancing environmental toxicology through chemical dosimetry: external exposures versus tissue residues. *Integrated environmental assessment and management* **7**, 7-27 (2011).