Erschienen in: Alternatives to animal experimentation : ALTEX ; 29 (2012), 4. - S. 411-425

The Use of Biomarkers of Toxicity for Integrating *In Vitro* Hazard Estimates Into Risk Assessment for Humans

Bas J. Blaauboer [1], Kim Boekelheide [2§], Harvey J. Clewell [3], Mardas Daneshian [4], Milou M. L. Dingemans [1], Alan M. Goldberg [5], Marjoke Heneweer [6], Joanna Jaworska [7], Nynke I. Kramer [1], Marcel Leist [4,8], Hasso Seibert [9], Emanuela Testai [10], Rob J. Vandebriel [11], James D. Yager [12], and Joanne Zurlo [5 12]

[1]Institute for Risk Assessment Sciences, Division of Toxicology, Utrecht University, Utrecht, The Netherlands; [2]Department of Pathology and Laboratory Medicine, Brown University, Providence, USA; [3]The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina, USA; [4]Center for Alternatives to Animal Testing-Europe, University of Konstanz, Konstanz, Germany; [5]Center for Alternatives to Animal Testing, Johns Hopkins University, Baltimore, USA; [6]Shell International B.V., The Hague, The Netherlands; [7]Procter & Gamble, Central Product Safety, Strombeek-Bever, Belgium; [8]Doerenkamp-Zbinden Chair for in vitro toxicology and biomedicine, Faculty of Sciences and Mathematics, University of Konstanz, Konstanz, Germany; [9]Institute for Toxicology and Pharmacology for Natural Scientists, University Medical School Schleswig-Holstein, Kiel, Germany; [10]Istituto Superiore di Sanità, Environment and Primary Prevention Department, Mechanism of Toxicity Unit, Rome, Italy; [11]Laboratory for Health Protection Research, National Institute for Public Health and the Environment, Bilthoven, The Netherlands; [12]Johns Hopkins Bloomberg School of Public Health, Department of Environmental Health Sciences, Baltimore, MD, USA

Summary

The role that in vitro systems can play in toxicological risk assessment is determined by the appropriateness of the chosen methods, with respect to the way in which in vitro data can be extrapolated to the in vivo situation. This report presents the results of a workshop aimed at better defining the use of in vitro-derived biomarkers of toxicity (BoT) and determining the place these data can have in human risk assessment. As a result, a conceptual framework is presented for the incorporation of in vitro-derived toxicity data into the risk assessment process. The selection of BoT takes into account that they need to distinguish adverse and adaptive changes in cells. The framework defines the place of in vitro systems in the context of data on exposure, structural and physico-chemical properties, and toxicodynamic and biokinetic modeling. It outlines the determination of a proper point-of-departure (PoD) for in vitro-in vivo extrapolation, allowing implementation in risk assessment procedures. A BoT will need to take into account both the dynamics and the kinetics of the compound in the in vitro systems. For the implementation of the proposed framework it will be necessary to collect and collate data from existing literature and new in vitro test systems, as well as to categorize biomarkers of toxicity and their relation to pathways-of-toxicity. Moreover, data selection and integration need to be driven by their usefulness in a quantitative in vitro-in vivo extrapolation (QIVIVE).

Keywords: biomarker of toxicity, integrated testing strategies, quantitative in vitro-in vivo extrapolations

* a report of t4 the transatlantic think tank for toxicology a collaboration of the toxicologically oriented chairs in Baltimore Konstanz and Utrecht sponsored by the Doerenkamp Zbinden Foundation The opinions expressed in this report are those of the participants as individuals and do not necessarily reflect the opinions of the organizations they are affiliated with

§ Competing nterests Declaration K Boekelheide is an occasional expert consultant for chemical and pharmaceutical companies

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-256471

1 Introduction

This is the report of a workshop organized to identify the possible next steps in incorporating the use of *in vitro*, *in silico*, and other non-animal-based methodologies into the process of toxicological risk assessment. The workshop was organized by the Transatlantic Think Tank for Toxicology, a group of scientists[1] that promotes recent changes in the paradigm of toxicity testing (Daneshian et al., 2010). A general outline of the new approach to toxicity testing is presented in a number of documents produced by the Dutch Health Council (HCN, 2001), ILSI-Europe (Eisenbrand et al., 2002), and the National Research Council report *Toxicity Testing in the 21st Century* (NRC, 2007). The event was hosted by the Institute for Risk Assessment Sciences of Utrecht University and was held in Utrecht in January 2011. This report has been updated with references through 2012. For explanation of terminology used in this document, see Table 1.

The workshop was designed to further define the use of biomarkers obtained for *in vitro* systems (BoT) and to clarify their role in toxicological risk assessment. Discussion was driven by a question formed at the beginning of the meeting: "How can *in vitro*-derived biomarkers (BoT) be used as input in the risk assessment procedure?"

2 Background

Current practice in toxicological risk assessment of health or environmental risk associated with chemical exposure is most commonly based on clinical or histopathological endpoints determined in animal models. Apart from ethical objections to the use of animals (Russell and Burch, 1959), there is also a scientific motivation for re-evaluating these models. The use of animal data to predict the biological activities of compounds in humans is always prone to some degree of uncertainty due to the differences in kinetics and dynamics between the animal models and humans (Renwick and Lazarus, 1998). In addition, the apical clinical endpoints do not identify mechanisms of toxicity.

A shift in research practices has taken place over the last decades. New approaches seek to elucidate the mechanisms of toxicity (Hartung, 2011), based on the understanding that a chemical can interact with relevant sites or processes in a living organism. Mechanism-of-action is defined here as the primary chemico-biological interaction between the compound and a structural moiety in the biological system (*viz*. in or on a cell, a tissue, or an organ). The functional and structural changes that subsequently occur within a biological system, including the resulting clinically observable changes in the organism, are then collectively referred to as the toxicologically relevant mode of action (MoA) (Blaauboer and Andersen, 2007).

The above considerations have resulted in a (re)definition of the paradigm of toxicology; rather than relying on apical endpoints of toxicity as determined in animal models, the toxicity of a compound can be determined by its effect – or the effect of a bioactivated metabolite – on a critical target in the biological system. This effect, in turn, is governed by the concentration of the compound or its metabolite, and the change therein over time at the site of action. Depending on the nature of the interaction, this dose metric can either be described in terms of the area under the curve (AUC), by a peak concentration, or by a concentration above a certain threshold, *inter alia*. These three elements: comprehensive information on the active site concentration, critical compound (*viz*. parent or metabolite), and critical site of action, should be the basis of our understanding of the toxicity of a chemical, together with information about the physiological and toxicological relevance of these interactions, i.e., a chemical-induced adverse effect (Krewski et al., 2009; Blaauboer, 2010; Bhattacharya et al., 2011).

Precise data on the mechanisms and modes of action cannot easily be obtained by studying the apical endpoints in animal studies. This has led to the development of *in vitro* methods for toxicity testing focusing more specifically on mechanisms and modes of action. Over the last decades, test systems for evaluating the possible toxicological hazard of chemical compounds have been developed that make use of biological systems on a lower level of organization: isolated organs, cell cultures, and subcellular systems (Worth and Balls, 2002; Basketter et al., 2012; Tralau et al., 2012). These *in vitro* systems have been very useful for screening purposes, particularly in studying the mechanism(s) of toxic action of potentially harmful chemicals (Eisenbrand et al, 2002; Adler et al., 2011; Bouvier d'Yvoir et al., 2012). In addition, important developments have occurred that allow the prediction of biological reactivity based on physico-chemical properties such as structure, molecular size, reactive groups, etc. One application of this knowledge is in the construction of structure-activity relationships (SARs), although they are limited to specific groups of chemicals, depending on the applicability domain of the model used, that ideally, correlate a quantifiable property to a quantifiable biological activity (QSARs) (Ellison et al., 2011; Demchuk et al., 2011).

Despite the great potential these developments offer for chemical risk assessment, the use of *in vitro* toxicity data is highly dependent on the physiological relevance of the *in vitro*-derived data and its potential use in an *in vitro-in vivo* extrapolation (IVIVE) (Blaauboer, 2008). Because many *in vitro* systems lack specific biokinetic relevance, extrapolation using these data would be particularly difficult (Gülden and Seibert, 2006; Blaauboer, 2010).

Selection of the appropriate *in vitro* system and relevant biological parameters to be measured is critical to ensuring useful data for analysis. For some parameters, it is possible to predict the most relevant physico-chemical features, toxicological modes of action (e.g., mutagenicity) or (bioactivated or deactivated) metabolites on the basis of the compound's structure. Such methods make use of systems such as DEREK, HazardExpert, TOPKAT, METEOR and MultiCase. This approach is not successful for all classes of chemicals, nor is it easily quantifiable (Ellison et al., 2011). It may, however, allow a bet-

[1] n alphabetical order Bas Blaauboer Alan Goldberg Thomas Hartung and Marcel Leist

Tab. 1: Glossary of terms used in the context of this paper

b omarker of tox c ty	A parameter that prov des quant tat ve nformat on that s mechan st ca y re evant to and pred ct ve of an adverse effect (Boeke he de and Schuppe-Ko st nen, 2012)
b omarker of effect	A parameter that prov des quant tat ve nformat on for an effect but does not necessar y d scr m nate between adverse and non-adverse effects
b omarker of exposure	A parameter that prov des quant tat ve nformat on on exposure (*in vitro*: of the ce u ar system; *in vivo*: of organ sms)
in vitro b omarker of tox c ty (BoT)	An *in vitro* der ved parameter that prov des quant tat ve nformat on that s mechan st ca y re evant to and pred ct ve of an adverse effect *in vivo*
endpo nt	The b o og ca or chem ca process response of effects assessed by a test method (Le st and Karreman, 2010; Crofton et a ., 2011)
ap ca endpo nt	An emp r ca y ver fiab e outcome of exposure, assessed n an ntact organ sm (Krewsk et a ., 2010)
mechan sm of act on	The pr mary chem co-b o og ca nteract on between the compound and a structura mo ety n the b o og ca system (B aauboer and Andersen, 2007)
mode of act on (MoA)	Funct ona and structura changes that occur subsequent to the pr mary chem co-b o og ca nteract on w th n a b o og ca system, nc ud ng the resu t ng c n ca y observab e changes n the organ sm (B aauboer and Andersen, 2007)
pathway of tox c ty (PoT)	A ce u ar response pathway that, when suffic ent y perturbed, s expected to resu t n an adverse hea th effect (NRC, 2007)
adverse outcome pathway (AOP)	A pathway of events, start ng w th a mo ecu ar n t at ng event n wh ch a chem ca nteracts w th a b o og ca target, ead ng to a sequent a ser es of h gher order effects to produce an adverse outcome w th d rect re evance to a g ven r sk assessment context (Ank ey et a ., 2010)
dos metry	An est mat on of the externa or nterna dose n organ sms or t ssues resu t ng from the exposure to a chem ca (exposure be ng a funct on of dose and t me)
reverse dos metry	The process of ca cu at ng the dose to wh ch an organ sm wou d be exposed to produce a concentrat on n t ssues that s equ va ent to a concentrat on measured n an *in vitro* system
nom na concentrat on	The amount of a compound added to the cu ture med um of an *in vitro* test system d v ded by the vo ume thereof
po nt of departure (PoD)	The concentrat on or dose of a compound that s taken from a concentrat on- or dose-effect re at onsh p n a test system and s used as a start ng po nt for extrapo at ons n a r sk assessment
quant tat ve *in vitro-in vivo* extrapo at on (QIVIVE)	The process of est mat ng the env ronmenta exposures to a chem ca that cou d produce target t ssue exposures n humans equ va ent to those assoc ated w th effects n an *in vitro* tox c ty test. Th s ca cu at on s done based on an *in vitro* concentrat on-effect re at onsh p and phys o og ca y-based k net c mode ng (Yoon et a ., 2012).
QSAR	Quant tat ve structure act v ty re at onsh p
Area-under-the curve (AUC)	The ntegra of the curve n a concentrat on-t me or dose-t me d agram
DEREK (Lhasa Ltd)	A know edge-based system that dent fies structura a erts for a w de var ety of tox c t es and target organs
TOPKAT (Acce rys)	A QSAR-based tox c ty pred ct on system that conta ns mode s for carc nogen c ty, deve opmenta tox c ty, sk n sens t zat on and var ous effect eve s (e.g., the chron c LOAEL)
METEOR (Lhasa Ltd)	A know edge-based system for metabo te pred ct on, nked to DEREK
HazardExpert (CompuDrug Ltd)	A know edge-based system that dent fies structura a erts for a w de var ety of tox c t es and target organs
Mu t case (Mu t Case Inc)	A system des gned to dent fy automat ca y a the mo ecu ar fragments that may ex st w th n a set of d verse chem ca s tested under a common protoco for any k nd of endpo nt

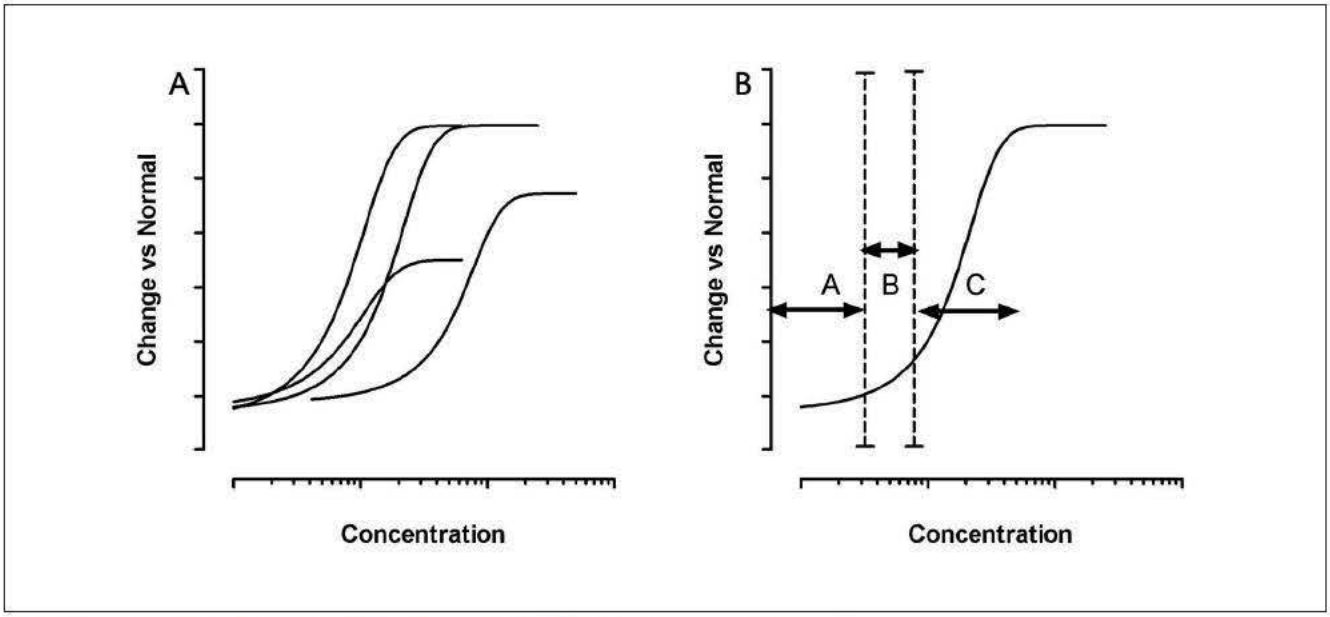

Fig. 1: Interpretation of concentration-effect relationships for *in vitro* experiments
(A) Examples for the many possible relationships between a compound's concentration and endpoint changes within one give experimental system are shown. For instance, different endpoints react at different compound concentrations. Some of these changes will be a reflection of adaptation or they may be unrelated to the eventual cell fate. Some will be related to adversity or they reflect a pathway of toxicity (PoT) relevant for cell fate and for *in vivo* toxicity prediction. Note that also the factor time will have an effect on the shape of the curves: duration of exposure, timing of (short-term) exposure within a more extended experimental protocol, and timing of measurement. Choices will have to be made for selecting the most relevant of these endpoints as BoT.
(B) If a choice has been made for one or more of the relationships in A to be used as BoT, the next step is to define concentration thresholds related to adversity. For each BoT, ranges of compound concentrations can be observed that do not affect the biomarker (region A). In other concentration ranges (region B) the BoT changes significantly from its baseline, but this effect does not predict adversity. In a third concentration range (region C) the change of the BoT is related to adversity

ter choice of relevant test systems and BoT for an initial evaluation of a compound's toxicological profile.

3 Biomarkers

Progress in the field of alternative methods depends on our ability to establish relevant *in vitro* systems (or batteries of systems) for the different domains of risk evaluation. In this context, it is necessary to improve our ability to select the most appropriate (functional or structural) parameters to be used in each of the new systems. This consideration is particularly important at a time when high-throughput chemical testing (HTS) is needed for analysis (Benford et al., 2000; Sipes et al., 2011b; Dix et al., 2012; Judson et al., 2012). Not all simple endpoints with technical advantages for HTS also qualify automatically as relevant biomarkers of toxicity. Therefore, a clear operational definition of a BoT and the distinction of BoT from other concepts brought forward in the field of *in vitro* toxicology become important at this point.

Using *in vitro* systems, early cellular responses can be studied that may help predict toxic responses *in vivo*. Examples of early cellular responses include: oxidative stress and glutathione homeostasis, cellular stress responses, changes in enzyme activities, and cytokine responses, among others (Eisenbrand et al., 2002; Pöltl et al., 2012). Measurement of such biomarkers of effect often is complemented by high-throughput approaches such as genomics, transcriptomics, and proteomics. These methods provide high-content information on the behavior of *in vitro* test systems, but their interpretation also requires advances in bioinformatics and systems biology (Adler et al., 2011; Van Summeren et al., 2012; Basketter et al., 2012). The different omics methods measure a multitude of endpoints, but not every endpoint qualifies as a BoT. In other words: not every parameter that changes is relevant and predictive for hazardous effects *in vivo* (see also Fig. 1). This is an important distinction between a simple test endpoint (Leist et al., 2010; Crofton et al., 2011; Boekelheide and Schuppe-Koistinen, 2012) and a BoT, which is the focus of this review.

Many parameters may technically qualify as a test system endpoint. However, the definition of BoT additionally includes a conceptual element, linked to toxicological predictivity and to the relevance of the parameter with respect to prediction of (human) hazard (Fig. 1). Thus, the concept of a BoT goes beyond the rather technical definition of an assay "endpoint." In that sense, BoT are related to the concepts of pathways-of-toxicity (PoT) (NRC, 2007) as explored within the human toxome project (Hartung and McBride, 2011) and to adverse outcome

pathways (AOP), as explored by the OECD and other regulatory agencies (Ankley et al., 2010). In simple terms, BoT, PoT, and AOP are related, but they differ mainly in scale. A PoT is a chain of events triggered by a chemical and leading to a hazardous outcome for the cell (Hartung and McBride, 2011; Perkel, 2012). A BoT could be regarded as an important component of a PoT, particularly useful for quantification in an *in vitro* assay. AOPs were originally used in environmental toxicology to describe the chain of events starting from molecular interaction of a chemical with a target (mechanism-of-action) and ending at effects on the organism and even its population. In the last two years the concept has been more broadly used to link toxicant effects on many levels of toxicity. The intention is to link initial mechanistic knowledge to the prediction of hazard for humans (Ankley et al., 2010; Sipes et al., 2011; Watanabe et al., 2011). Thus, an AOP provides the rationale for the use of one or the other BoT by showing how the changes measured by the BoT relate to the prediction of human hazard.

How to *define* "biomarkers of toxicity," specifically as relevant to *in vitro* systems, was the topic of an extensive discussion during the workshop. Since relevance of the chosen *in vitro* approaches greatly determines their ability to be extrapolated to an *in vivo* context, the choice of what to measure (i.e., the biomarkers) is also of high importance. Moreover, to define the distinction between terms was also considered essential, e.g., between "biomarkers of effect" and "biomarkers of exposure." Furthermore, the relationship between a "biomarker of effect," the primary mechanism of action, the MoA as defined above, adaptive responses versus adverse responses, etc., were discussed. These issues will be treated in detail below. A number of terms are included in Table 1, also referring to earlier published definitions (Ankley et al., 2010; Leist et al., 2010; Crofton et al., 2011).

During the discussion, the following biomarker-defining questions helped to create a broad definition for biomarkers *in vitro*:

- Is it a measureable variable?
- Is it quantifiable?
- Does it represent a chemico-biological interaction?
- Is it predictive of the most sensitive (rate limiting) toxic processes?
- Is it representative of a toxic pathway?
- Does it have one or a set of measurable endpoints (fingerprint)?
- Is it a parameter that represents or mirrors a toxic response *in vivo*?
- Does it provide information on the rate, magnitude and reversibility of a parameter?

After ample discussion we agreed upon the following definition:

> An *in vitro* biomarker of toxicity (BoT) provides quantitative information that is mechanistically relevant to and predictive of an adverse effect *in vivo*.

4 Biokinetic[2] considerations

Proper interpretation of *in vitro* data, particularly for their relevance in a toxicological risk evaluation for intact organisms, requires the consideration of kinetic aspects of each system (Blaauboer, 2010; Caldwell et al., 2012). Knowledge of the biokinetic behavior of the chemical is required in two areas: first, the kinetics of the compound in the *in vitro* system ("biokinetics *in vitro*"), second, the use of kinetic models in extrapolating the *in vitro* dose metrics to the *in vivo* situation.

The first deals with the determination of the actual biological exposure. Toxic effects, or biotransformation rates, for *in vitro* models usually are related to the concentrations of the compound added to the medium. These nominal concentrations can deviate from the actual free concentration of the compound in the system, and they change over time (due to binding to proteins in the medium, adsorption to the plastic devices, evaporation, or uptake in the cells). Since the freely available concentration usually is the driving force for kinetic processes, as well as toxic reactions on the (sub-) cellular level, these processes will influence the free concentration and thus the effect (Gülden et al., 2002). It is therefore necessary to estimate or measure this free concentration, especially when it is expected that the free concentration will differ from the nominal concentration (on the basis of known physico-chemical properties such as lipophilicity) (Gülden and Seibert, 2003; Heringa et al., 2004; Kramer et al., 2010, 2012).

Several techniques exist to estimate the free concentration of chemicals in an *in vitro* assay medium, including equilibrium dialysis, ultracentrifugation, and ultrafiltration (Oravcová et al., 1996). A more recent technique uses the simultaneous extraction and sampling of the unbound chemical from culture medium with solid-phase micro-extraction (SPME) devices and to analyze the compound (Vaes et al., 1997; Kramer et al., 2007; Broeders et al., 2011). These devices consist of small rods covered with material that absorbs the compound in equilibrium with its free concentration. This technique allows the identification of processes that influence the free concentration, which in turn enables modeling of the *in vitro* system. The application of these techniques has shown that, for some compounds, the free concentration can differ up to two orders of magnitude from the nominal concentration, emphasizing the importance of understanding, measuring, and modeling the biokinetics *in vitro* (Gülden et al., 2006; Kramer et al., 2012). Moreover, the cellular concentration can differ from the medium concentration by several orders of magnitude (Zimmer et al., 2011; Kramer et al., 2012).

Biokinetic considerations are equally important when designing the technical set-up of an *in vitro* experiment, particularly on the relationship between the amount of the compound present in the *in vitro* system and the number of cells. If these conditions are different from those expected *in vivo*, the relevance of the *in vitro*-derived toxicity data may be diminished. If the number

[2] Quote from Clewell et al 2008 "The time course of drugs in biological systems has traditionally been referred to as pharmacokinetics On the other hand it has become popular to use the term toxicokinetics when dealing with chemicals that are toxic This of course ignores the wisdom of Paracelsus only the dose differentiates a poison and a remedy To avoid this false distinction the term biokinetic will be used in this paper"

of cells in the system is changed, the amount of test compound available for the individual cells in the system also will change (Gülden et al., 2001, 2006). In addition to the experimental setup itself, the compound's dynamics also can influence the system's kinetics: compounds with a high reactivity can react with a cellular component, causing an immediate effect on or in the cells and thereby leading to a decrease in the compound's concentration (Gülden et al., 2010).

Because biokinetic considerations are critical to accurately interpreting *in vitro* data (Blaauboer, 2010; Adler et al., 2011; Coecke et al., 2012) the use of physiologically-based biokinetic (PBBK) models has become critical in translating the concentration-effect relationships found in relevant *in vitro* models to dose-effect relationships *in vivo*. In essence, the kinetic models are used to estimate the external exposure that would result in effective concentrations at relevant targets. In these so-called "reverse dosimetry" calculations, it is assumed that:

1) the *in vitro* toxicity data reflect the relevant toxicity parameters for the *in vivo* situation (see also the next section: *in vitro* effects battery);
2) the *in vitro* effective concentrations are representative of effective concentrations *in vivo*; and
3) the appropriate parameters for constructing an adequate PBBK model are available.

Ideally, these parameters also are derived from non-animal studies (Adler et al., 2011; Basketter et al., 2012; Coecke et al., 2012). For a recent review of these "Quantitative *In Vitro-In Vivo* Extrapolations," (QIVIVE) we refer to Yoon et al. (2012).

5 Adversity versus adaptation

The use of cell culture in toxicity testing of chemicals has the potential to provide a detailed picture of the changes of many parameters at once. Even if these changes show a clear concentration-effect relationship, care must be taken in interpreting the results in view of their relevance to the compound's toxicity. Most likely, the sensitivity of these detailed studies will be much higher than what can be derived from the interpretation of apical endpoints in an animal study, e.g., due to the lack of compensatory/homeostatic processes, usually working *in vivo*. The question is then: *when is a change related to an adverse effect, and when should a change be interpreted as falling within the boundary of the physiologically "normal" adaptive range?*

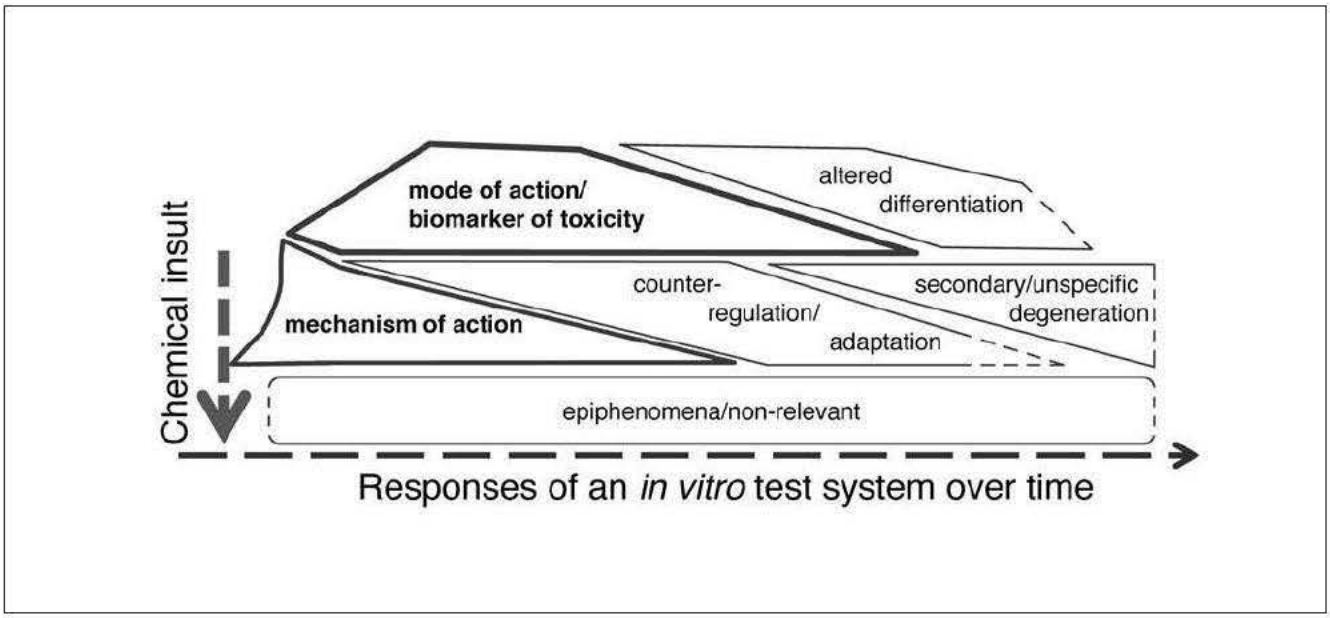

Fig. 2: Responses of an *in vitro* test system over time

The test system s character zed by a mu t tude of parameters that are n t a y w th n the r homeostat c range. After a chem ca nsu t (nd cated by an arrow h tt ng the hor zonta t me ax s) many of these parameters (e.g., metabo tes, transcr pts or ce organe e funct ons) w change n a t me-dependent manner. For the se ect on of re evant BoT, these parameters may be grouped accord ng to the r re at on to ce fate and hazard of the chem ca . The first group reflects the mmed ate mechan sm of act on of the chem ca (e.g., b nd ng to an enzyme). As chem ca s may have mu t p e targets, the pred ct v ty of changes of one of these ear y parameters s often ow, but t can be usefu as BoT, espec a y for pharmaceut ca s. The second group reflects the downstream mode of act on (MoA) of the chem ca and often s very su tab e as BoT. Some parameters change w thout hav ng a pred ct ve va ue for the fate of the ce (ep phenomena) or they are ce u ar counter-regu at ons of the n t a nsu t. They are not su tab e as BoT. In ate phases, there s a strong change of parameters, e.g., re ated to ce death. These appear usefu at first s ght, but they are often unspec fic, and often on y reflect a genera breakdown of homeostas s. A comp cated group of changes s re ated to a tered ce u ar d fferent at on. They reflect a new form of homeostas s and are d fficu t to nterpret. They can be usefu n the fie d of deve opmenta tox c ty, but the r use as BoT requ res great care and va dat on. The gaps n some box out nes symbo ze that such changes phase n and phase out at d fferent t me po nts that cannot be sharp y defined.

In analyzing *in vitro* toxicity data it is important, then, to distinguish between adaptive changes and adversity. Within one given experimental system many possible relationships between a compound's concentration and endpoint changes can be envisioned (Fig. 1A). For instance, different endpoints react at different compound concentrations. Some of these changes will be a reflection of adaptation or they may be unrelated to the eventual cell fate, while others will be related to adversity or they reflect a pathway-of-toxicity (PoT) relevant for cell fate and for prediction of *in vivo* toxicity. Note that the time factor will have an effect on the shape of the curves: duration of exposure, timing of (short-term) exposure, and timing of measurement. Moreover, since the different processes may have different dynamics and dynamic ranges, the types of phenomena observed also will change with time (Fig. 2).

For each chosen BoT there will be a range of concentrations at which there is a measurable effect, which is within the normal physiological range and not related to the adverse effect that will occur at higher concentrations (range B in Fig. 1B). For example, if the chosen BoT is the inhibition of an enzyme activity, a relatively small inhibition would not result in cellular dysfunction, while higher levels of inhibition would do so. These considerations need to be taken into account when selecting a BoT and using it to determine point-of-departure (PoD) for evaluation of human risk.

One caveat in the use of *in vitro* systems is the absence of integrative systems occurring in more complex tissues, whole organs, or the total organism, so it is important that mechanistically-based BoT derived from non-animal systems are predictive for the adverse effect in the whole, integrated organism. It will be a challenge to select those BoT and their relevant values to take both the inherent high sensitivity *in vitro* and the possible feedback loops present at higher levels of biological complexity into account (Aldridge et al., 2006; Boekelheide and Andersen, 2010). The use of *in vitro* methods is therefore complicated by a potential lack of interactions (i.e., between compounds and cells) that otherwise would be present at higher levels of biological integration (Kadereit et al., 2012; van Thriel et al., 2012). These feedback mechanisms should be considered when interpreting the results of *in vitro* toxicity testing for risk assessment. Organ slices have the capability to show the interaction of cells in their *in situ* tissue coherence, e.g., the hepatocytes with Kupffer cells in liver slices (van de Bovenkamp et al., 2005; Catania et al., 2007). Examples exist where these interactions are

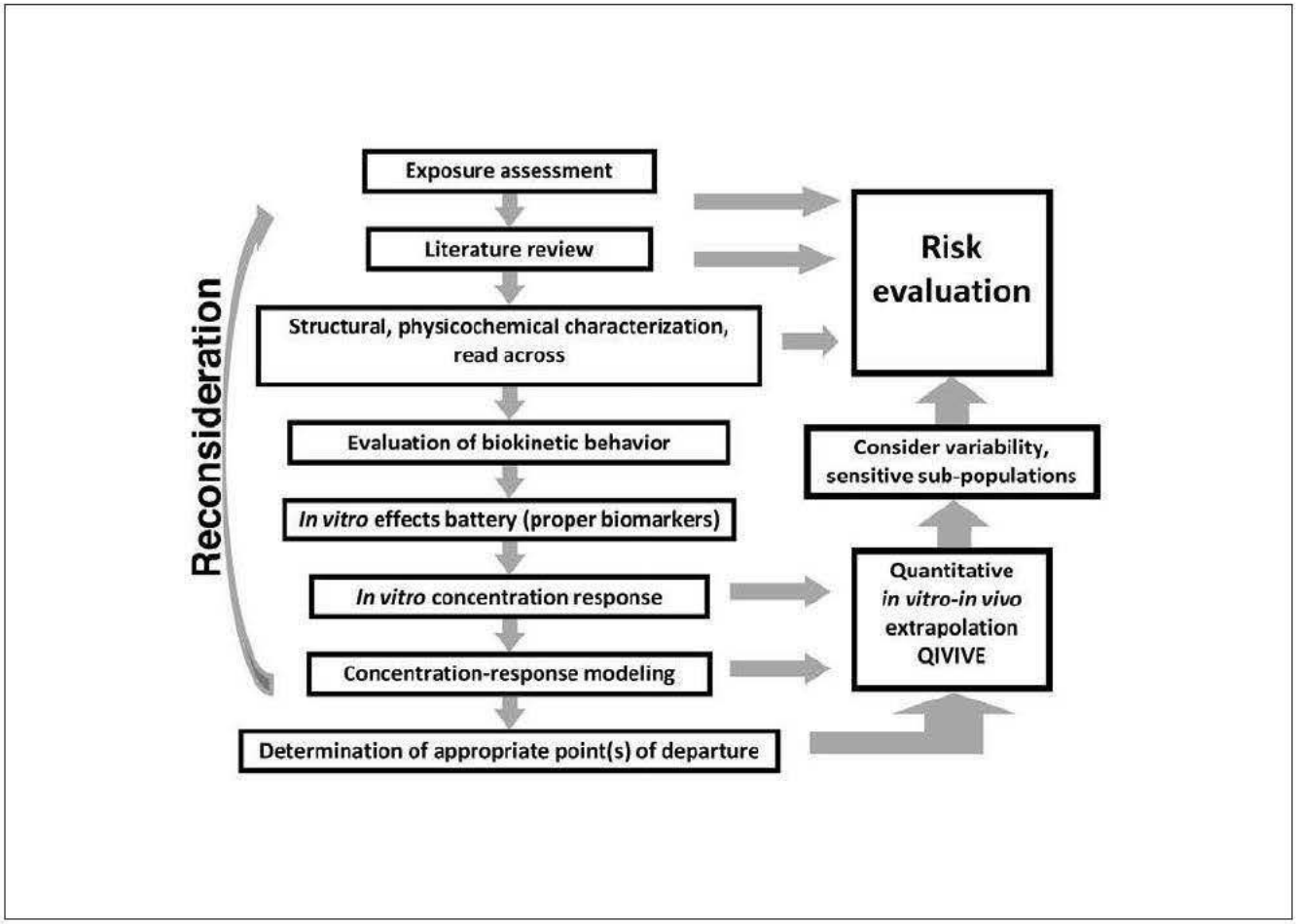

Fig. 3: Scheme for the incorporation of *in vitro* biomarker-derived toxicity data in the process of chemical risk assessment
For further exp anat on see text, sect on 6.

studied by employing *in vitro* co-cultures of the relevant cells (Heneweer et al., 2005; Hallier-Vanuxeem et al., 2009; Henn et al., 2009; Li et al., 2012; Leite et al., 2011; Schildknecht et al., 2009, 2011, 2012). The human- or organ-on-a-chip techniques provide another example where different cell cultures can be employed in the same system, offering a more integrated *in vitro* system (van Midwoud et al., 2010; Hartung and Zurlo, 2012; Prot and LeClerc, 2012). New and developing methods allow these integrative effects to model the whole organism (Bosgra et al., 2009). The examples listed above show that the integration of kinetic and dynamic models is adding crucial power to these approaches (e.g., see DeJongh et al., 1999; Bushnell et al., 2005; Forsby and Blaauboer, 2007; Paini et al., 2010).

There are many different options for studying toxicologically relevant effects *in vitro*. However, the interpretation of data with regard to the difference between adversity and adaptation is still a challenge: to address it would make *in vitro* data more applicable for assessing risks. The conceptual framework described in the next paragraph highlights the most urgent issues.

6 Conceptual framework

Taking the above kinetic and mechanistic considerations into account, a conceptual scheme is proposed for the integration of *in vitro*-derived biomarkers into the process of risk assessment (Fig. 3). A number of schemes that modernize the process of chemical risk assessment, e.g., the one developed by the Health Council of the Netherlands (HCN, 2001), have been presented in the literature. The specific purpose of the scheme presented here is to place the proper use of *in vitro*-derived biomarkers into the perspective of the risk evaluation of chemicals. In this respect, we build on earlier reports on integrated testing schemes (Blaauboer et al., 1999; Jaworska and Hoffmann, 2010; Kinsner-Ovaskainen et al., 2012).

Exposure

In this scheme, a risk evaluation begins by considering the probable exposure scenarios for a given chemical. In cases where all relevant exposures will be low, i.e., below the threshold of toxicological concern (TTC) (Kroes et al., 2007), a risk evaluation for that chemical could be initiated without any need for testing.

For many chemicals, some toxicity data are available in the literature. The next step, therefore, would be a proper evaluation of available data using an evidence-based approach (e.g., Hartung, 2009) and thus further testing may also be unnecessary. *In vitro* testing could provide additional mechanistic insights, and this could be a reason to continue experimental work, as proposed in the scheme.

Structural properties

After the evaluation of potential exposure scenarios, a next step is evaluation of the structural properties of the chemical and/or its active metabolites. Knowledge of specific physicochemical characteristics, e.g., a high reactivity towards biomacromolecules, can then form a starting basis for risk evaluation (Ellison et al., 2011). Adverse effects of chemicals may be evaluated using a read-across approach based on known data from similar compounds (Vink et al., 2010; Schüürmann et al., 2011). Since such knowledge can be useful to predict possible toxicological targets, structural and physicochemical properties of compounds can be the basis for selecting a proper *in vitro* test battery. However, selection of appropriate cellular systems also should involve *biokinetic* considerations. For example, there is no need for the determination of any systemic effects if a topically applied compound has very low or no internal exposure due to a minimal level of dermal absorption; this situation most likely suggests that the *internal* threshold of toxicological concern is not exceeded. In that case only local toxicity would have to be assessed, for which the appropriate *in vitro* models should be selected.

As mentioned above, the structural properties of a compound can help guide the selection of an appropriate cell culture system. A number of software systems are available for making these *in silico* predictions of toxicity, either employing knowledge-based data sets or QSAR-based models. An example of the former is DEREK, which identifies structural alerts for a variety of toxicological endpoints (Marchant et al., 2008). Examples of QSAR-based approaches are TOPKAT and the OECD Toolbox. TOPKAT is a commercial QSAR-based toxicity prediction system that contains models for carcinogenicity, developmental toxicity, skin sensitization and various effect levels (e.g., the chronic Lowest Observed Adverse Effect Level (LOAEL); Venkatapathy et al., 2004). The freely available OECD Toolbox (van Leeuwen et al., 2009) identifies the potential for macromolecular interactions (DNA binding, protein binding, estrogen receptor binding) based on the physico-chemical properties of the compound.

On the basis of these data an initial selection of the appropriate cell culture systems may be determined. As an example, if the systems find structural properties that indicate a possible or probable interaction with a certain target tissue, this may guide the choice of the most appropriate *in vitro* systems to study a concentration-effect relationship.

Evaluation of biokinetic behavior

The importance of biokinetics in the interpretation of *in vitro* data for risk assessment was discussed earlier. It remains only to explain why the evaluation of biokinetic behavior should be placed prior to the *in vitro* test battery in Figure 3. The answer again comes from the importance of using the appropriate biomarker – in this case the biomarker of exposure. It has already been discussed that the use of nominal concentration as the measure of exposure in an *in vitro* system overlooks a number of factors that may lead to the free concentration of chemical being different from the nominal. However, there is also a second concern that must be considered: that the toxicity of a chemical may result from the action of one or more of its metabolites rather than from the chemical itself. *In vitro* toxicity tests will inevitably possess differing capabilities for metabolic transformation (Coecke et al., 2006). It is therefore critical to know whether metabolism needs to be considered during the design and interpretation of the *in vitro* tests for a particular chemical and, if necessary, its metabolites (NIEHS, 2001).

Similarly, when extrapolating *in vitro* test results to the equivalent *in vivo* exposures the comparison must be made on the basis of the correct biomarker of exposure (Yoon et al., 2012). For direct chemical toxicity the appropriate quantity to measure would usually be the area under the concentration curve (AUC) or average concentration (AUC divided by duration of exposure) of the parent chemical (Andersen et al., 1987b); however, for a chemical whose toxicity results from a metabolite, the appropriate dose metric would be related to the concentration of the metabolite rather than that of the parent (Andersen et al., 1987a; Clewell et al., 2002). Whereas the average concentration of the parent is proportional to the ratio of dose to parent clearance, the metabolite concentration is proportional to the ratio of parent clearance to the clearance of the metabolite (Andersen, 1987). Further, in the case of a highly reactive metabolite, where its disappearance is due to chemical reactivity rather than enzyme mediated clearance, the appropriate biomarker of exposure is the rate of formation of the metabolite divided by the volume (media or target tissue) into which it is generated (Andersen et al., 1987a). To ensure that the correct biomarker of exposure is measured in the *in vitro* assays, it is necessary to identify those cases where the toxicity of a chemical may be due to a metabolite prior to conducting the *in vitro* effects battery.

In vitro effects battery

An *in vitro* effects battery for the new toxicity testing paradigm needs to be designed to efficiently detect biomarkers of toxicity. This test battery will depend upon a thorough systems biology understanding of cellular function, and will use a variety of test platforms, including reporters for stress pathways, omics approaches (transcriptomics, proteomics, and metabolomics) (Adler et al., 2011; Kienhuis et al., 2011; Van Summeren et al., 2012), and high-content analysis imaging platforms (Zanella et al., 2010; Stiegler et al., 2011). Many of these technical approaches are likely to provide complementary information, and only through experience and inter-laboratory validations will the most sensitive and robust tests and platforms be identified.

The development of the *in vitro* effects battery will be an iterative process, likely beginning with established cell lines that are well understood and well characterized, and building on lessons learnt (Boekelheide and Andersen, 2010; Basketter et al., 2012). The ideal test system will display all of the differentiated features and cellular functions found in intact organisms of various life stages, disease states, and conditions. Potential models for a cellular test system could use human embryonic stem cells (hESCs), or other types of stem/progenitor cells, in conjunction with protocols that allow these cells to differentiate along numerous organ-specific pathways (Leist et al., 2008; Kuegler et al., 2010; Wobus and Löser, 2011; Zimmer et al., 2011, 2012; Balmer et al., 2012; Meganathan et al., 2012).

By incorporating reporters that mark differentiated functions into these cells, toxicant-induced perturbation of organ-specific attributes could be examined and deduced. Further, the use of three-dimensional and heterogeneous cellular aggregates may provide additional insight into cell-cell interactions and the disruption of paracrine signaling processes by toxicant exposure (Heneweer et al., 2005; Cantòn et al., 2010). The broad goal is to design a fit-for-purpose set of optimized *in vitro* cellular systems that provide maximal coverage of human functionality while minimizing cost, complexity, and testing time.

In vitro effects batteries can also be used to model the variability of human susceptibility due to a genetic background or environmental factors. The parallel use of several cell lines from different donors for the same assay and endpoint can model different human genotypes (Lock et al., 2012). Alternatively, cells may be tested in different situations, e.g., after preconditioning, in inflammatory situations, and at different metabolic situations and ages/passage numbers (Latta et al., 2000; Falsig et al., 2004; Lotharius et al., 2005; Henn et al., 2011).

Combining data from different, complementary platforms and assays into a coherent testing package that appropriately weights and evaluates the different data sources will be a challenging task. An important part of this integration will be the development of visualization tools that display the combined data in an easily understood format. In addition, the development of tiered testing strategies is likely to provide an efficient means of identifying stopping points when sufficient data is available for decision making (HCN, 2001; Combes and Balls, 2011). However, such strategies should not be too rigid (Jaworska et al., 2010, 2011).

With respect to the type of BoT used in test batteries, different directions are being followed. A traditional approach is to use a single, relatively complex endpoint. This may be neurite growth, cell proliferation, or the change of reporters that respond to oxidative stress or inflammatory stimuli. This approach has the advantage that the BoT reflects different types of primary mechanisms of action, and it can be related to the MoA and hence to adverse effects *in vivo*. Therefore, this will most likely play an important role in the near future. Other approaches use multiple endpoints. Low numbers and high complexity of endpoints is typical for high-content imaging. High numbers (tens of thousands) of endpoints are tested in many "omics" approaches. A considerable amount of future work will be required to extract the most meaningful information from these approaches. An opposite type of development uses single endpoints and highly simplified test systems. In extreme cases, these may only consist of an isolated enzyme or receptor. Instead, very large test batteries are used (e.g., in the ToxCast program: Judson et al., 2012; Dix et al., 2012; Sipes et al., 2011a). Machine learning approaches are being developed to correlate the pattern of changes in such test batteries to *in vivo* data, and to use knowledge of such correlations for future predictions. Possibly, these three types of approaches (use of many single simple endpoints; use of few single complex endpoints; use of multiple endpoints) will be used in the future to define the best BoT and to provide predictions on chemical hazard.

Two major alternative approaches to the design of the *in vitro* test battery are proposed. In one approach, the *in vitro*-to-*in vivo* extrapolation occurs from the analysis of systems biology information after the execution of a common test platform. This approach depends on the development of a broad-based, *de novo*, holistic, and self-contained test system that is predictive of adverse effects based on alterations within components of

the interacting pathways that contribute to overall function. This approach demands that a broad range of differentiated characteristics of cells are represented, and that effects on this broad range of targets can be evaluated. Bioinformatics and systems biology approaches could be used to further extrapolate these results to understand the possible responses in individual organ systems.

In an alternative approach, the test system itself would be compartmentalized by the different types of biology inherent in the *in vivo* endpoints of concern. Development of test system modules would then be based on the apical endpoints of interest (Maxwell et al., 2008). In this approach, the specialized biology inherent to each apical endpoint would be emphasized in the development of each module, optimizing the tests within the module for sensitive detection of the specific biological areas of concern. Examples of such distinct modules might include a general screening test battery, with more specific test systems for reproductive and developmental effects, (developmental) neurotoxicity, hepatic toxicity, etc. The interpretation of the combined result of a high-throughput test battery was discussed by Judson et al. (2011), who used the lowest "biological pathway altering concentrations", together with probability distributions of kinetic and dynamic parameters in selecting a PoD.

Concentration-effect data

An important outcome of any *in vitro* toxicity test is the adequate evaluation of the concentration-dependent effects for the relevant parameters. As mentioned above, it is important to assess relevant concentration of the compound driving the toxicity (this might be the active metabolite(s)), taking into account their possible losses due to absorption to plastic, binding to proteins or chemical instability in the medium, or by evaporation (Seibert et al., 2002; Kramer et al., 2007), as well as biotransformation to innocuous metabolites. It is also important to consider the appropriate metric for the effective concentration, which can either be the peak concentration, or a peak concentration above a certain threshold level, or an area under the curve. Alternatively, the amount of the compound present in the cells ("cell burden") or even subcellular distribution may be the determining factor for the observed effect (Gülden et al., 2010).

Modeling and determination of points-of-departure for further evaluation of risk

Once reliable concentration-effect relationships have been established, these data need to be interpreted for their usefulness in determining risk. The above-mentioned notions regarding "adaptation vs. adversity" should be considered. Furthermore, a proper quantification of the results will help in determining an appropriate PoD for inclusion in risk evaluation. The application of modeling the concentration-effect relationship derived in a relevant *in vitro* system by means of the benchmark approach may be considered (Crump and Teeguarden, 2009; Sand et al., 2012). This process could then help identify a possible PoD for the next step of evaluation. One example is the use of the BMCL10: the benchmark concentration-lower limit of confidence for 10% of the maximal response (Fig. 4).

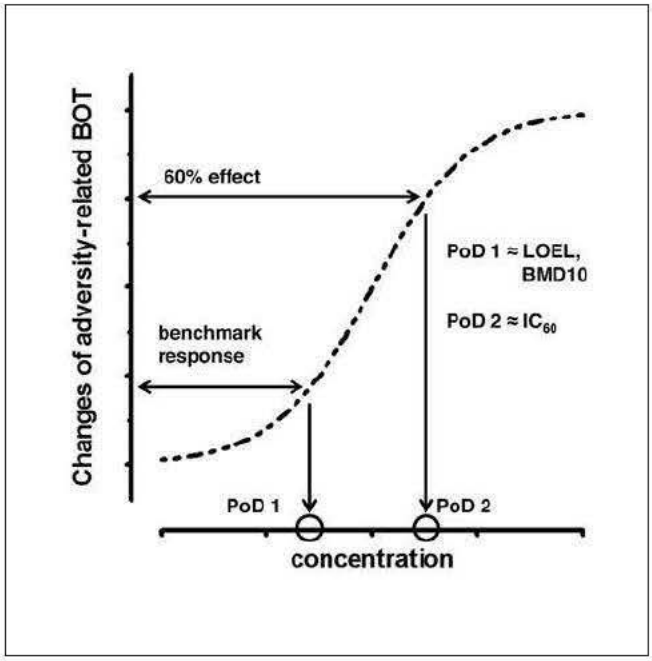

Fig. 4: Modeling and determination of points of departure (PoD)

The concentrat on-dependent change of a BoT, wh ch s re ated to the hazard of the respect ve chem ca tested, s shown. Such curves, obta ned from exper menta data and mode ng, are used to determ ne the appropr ate PoD for quant tat ve *in vitro*→*in vivo* extrapo at ons (QIVIVE). The PoD s named after ts ro e as start ng po nt for QIVIVE mode ng and ca cu at ons, and t s expressed as a concentrat on. It s NOT pre-defined wh ch part of the curve s most appropr ate for determ nat on of the PoD. Th s nformat on depends on the under y ng b o ogy and tox co og ca re evance. For nstance IC_{50}, IC_{90} or owest observed effect eve (LOEL) may be used, f th s s tox co og ca y/ b o og ca y just fied. Two examp es are g ven: PoD1 s defined by the BMD_{10}, .e., the owest concentrat on that resu ts n a 10% change of the base ne; PoD2 s defined as the IC_{60} of the curve. Concentrat ons sma er than the PoD wou d be cons dered non-s gn ficant w th respect to hazard-pred ct on. D fferent stat st ca and mode ng approaches are ava ab e to define such PoD.

The rules for choice of the PoD

Depending on the features of the test system and the nature of the BoT chosen, the concentration used for QIVIVE may differ (Fig. 4). For instance, the minimal significant effect concentrations corresponding to the lowest observed effect level (LOEL) of *in vivo* toxicity are relevant if mutations are chosen as BoT. In many cases the EC_{50} values may be a good choice as this parameter is the mathematically most robust datapoint to determine. Furthermore, the EC_{60} or EC_{90} value could be determined for instance in cases when cells have large reserve/buffering capacity, which is relevant, e.g., for glutathione depletion or ATP depletion, as biomarker.

Quantitative in vitro-in vivo extrapolations (QIVIVE)
The essence of QIVIVE is the "translation" of a concentration-effect relationship (or any point in this relation, e.g., a $BMCL_{10}$, or an EC_{50} or EC_{10}) to a dose-effect relationship in an intact organism. This process, also referred to as "reverse dosimetry", implies the consideration of a compound's kinetic behavior in the organism and results in the estimation of a dose (or any appropriate exposure scenario) *in vivo* that would result in the effective concentration at the site of toxic action. This subject is discussed in a recent review (Yoon et al., 2012).

Implementation in risk assessment
The final step of the presented procedure is the assessment of risk in the classical way: what is the risk of a certain exposure scenario to the health of an organism? This assessment should include the uncertainties related to the procedure described above, and also take into account the variability in sensitivity of sub-populations of individuals (HCN, 2008). Experience with these types of non-animal-based procedure is still limited, making evaluation of these uncertainties challenging.

Reconsideration
Each step in the procedure may require reconsideration of earlier steps in the scheme, for example, the outcome of *in vitro* biomarker-derived concentration-effect data may lead to reconsideration of the test battery.

7 Future research needs, conclusions, and recommendations

It is clear that much more experience with these test systems is needed for full integration of *in vitro* biomarker-derived data into the process of risk assessment (Punt et al., 2011; Gabbert and Benighaus, 2012). However, we can make use of past experiences, such as in the area of pharmaceutical development, where the use of *in vitro* methods in screening and selecting compounds for their efficacy, and to a lesser extent for their toxicity, has a longer history. Despite this, there is a need to construct more rigorous testing schemes for non-animal based risk assessments (Leist et al., 2010).

A number of studies provide some proof of concept for schemes similar to those proposed here. The first one was on the neurotoxicity of acrylamide (DeJongh et al., 1999), which used a sensitive and specific BoT *in vitro*, i.e., the inhibition of neurite formation in a neuroblastoma cell line. The study combined this BoT with a kinetic-dynamic model to perform reverse dosimetry and accurately predicted *in vivo* neurotoxicity in rodents. Later studies used other *in vitro* endpoints, including some complicated ones: developmental toxicity-related effects, by combining the embryonic stem cell test with PBBK modeling to predict the embryotoxic effects of glycol ethers (Louisse et al., 2010).

From these examples it is clear that it will be difficult to define one single approach that is "fit-for-all". However, any approach should be as simple as possible but as complex as necessary (Basketter et al., 2012). Therefore, there is an urgent need to put more research effort into building experience with chemical risk assessment using a scheme like the one proposed here. The identification of a "catalogue" of appropriate biomarkers for important pathways that are good indicators of adversity would be useful (Zimmer et al., 2012). Such pathways of toxicity will have to be biologically relevant and clearly related to toxicological endpoints. If a certain biomarker is a good indicator of adversity, (e.g., hepatotoxicity), the appropriate pathways of toxicity should be identified, and the *in vitro* systems selected should be able to identify the related biomarkers and pathways (OECD, 2012).

Although many *in vitro* toxicity data exist in the literature, a systematic overview of these data (which endpoints, which pathways, which biomarkers) is lacking. Data mining of the literature and the development of a monitored open-access database is recommended (Leist et al., 2008a). Such a database could also include data on other essential parts of the scheme proposed above, including the interaction of chemicals with biomacromolecules (proteins, lipids, DNA, etc.). These data could also be derived from computational toxicology techniques as these are further developed (Krewski et al., 2010).

An important conclusion from this workshop is that the integration of *in silico* and *in vitro* data in a risk assessment stands and falls with proper quantification – for biokinetics as well as for effect parameters. The study of the behavior of a chemical *in vitro* by measuring concentration (free concentration, dose in cells) or modeling biokinetics needs more toxicological emphasis. The same applies to proper quantification of the toxicolog - cal read-outs. Furthermore, the development and application of new tools or integrated strategies to evaluate the risk on a weight-of-evidence approach will also require adequate training for future risk assessors (Daneshian et al., 2011; Håkansson et al., 2011).

References

Adler, S., Basketter, D., Creton, S., et al. (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects-2010. *Arch Toxicol 85*, 367-485.

Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. *Nat Cell Biol 8*, 1195-1203.

Andersen, M. E. (1987). Tissue dosimetry in risk assessment, or what's the problem here anyway? *Drinking Water and Health* (8-23). Vol. 8. Washington, DC: National Research Council.

Andersen, M. E., Clewell, H. J., Gargas, M. L., et al. (1987a). Physiologically based pharmacokinetics and the risk assessment process for methylene chloride. *Toxicol Appl Pharmacol 87*, 185-205.

Andersen, M. E., MacNaughton, M. G., Clewell, H. J., and Paustenbach, D. J. (1987b). Adjusting exposure limits for long and short exposure periods using a physiological pharmacokinetic model. *Am Indus Hyg Ass J 48*, 335-343.

Ankley, G. T., Bennett, R. S., Erickson, R. J., et al. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem 29*, 730-741.

Balmer, N. V., Weng, M. K., Zimmer, B., et al. (2012). Epigenetic changes and disturbed neural development in a human embryonic stem cell-based model relating to the fetal valproate syndrome. *Human Molec Genet 21*, 4104-4114.

Basketter, D. A., Clewell, H., Kimber, I., et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX 29*, 3-91.

Benford, D. J., Hanley, A. B., Bottrill, K., et al. (2000). Biomarkers as predictive tools in toxicity testing. *Altern Lab Anim 28*, 119-131.

Bhattacharya, S., Zhang, Q., Carmichael, P. L., et al. (2011). Toxicity testing in the 21st century: Defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *PLoS ONE 6*, e20887.

Blaauboer, B. J., Barratt, M. D., and Houston, J. B. (1999). The integrated use of alternative methods in toxicological risk evaluation: ECVAM Integrated Testing Strategies Task Force Report I. *Altern Lab Anim 27*, 229-237.

Blaauboer, B. J. and Andersen, M. E. (2007). The need for a new toxicity testing and risk analysis paradigm to implement REACH or any other large scale testing initiative. *Arch Toxicol 81*, 385-387.

Blaauboer, B. J. (2008). The contribution of in vitro toxicity data in hazard and risk assessment: Current limitations and future perspectives. *Toxicol Lett 180*, 81-84.

Blaauboer, B. J. (2010). Biokinetic modelling and in vitro-in vivo extrapolations. J Toxicol Environ Health B 13, 242-252.

Boekelheide, K. and Andersen, M. E. (2010). A mechanistic redefinition of adverse effects – A key step in the toxicity testing paradigm shift. *ALTEX 27*, 243-252.

Boekelheide, K. and Schuppe-Koistinen, I. (2012). SOT/EUROTOX debate: Biomarkers from blood and urine will replace traditional histopathological evaluation to determine adverse responses. *Toxicol Sci 129*, 249-255.

Bosgra, S., van Eijkeren, J. C. H., van der Schans, M. J., et al. (2009). Toxicodynamic analysis of the inhibition of isolated human acetylcholinesterase by combinations of methamidophos and methomyl in vitro. *Toxicol Appl Pharmacol 236*, 1-8.

Bouvier d'Yvoire, M., Bremer, S., Casati, S., et al. (2012). ECVAM and new technologies for toxicity testing. *Adv Exp Med Biol 745*, 154-180.

Broeders, J. J. W., Blaauboer, B. J., and Hermens, J. L. M. (2011). Development of a negligible depletion-solid phase microextraction method to determine the free concentration of chlorpromazine in aqueous samples containing albumin. *J Chromatogr A 1218*, 8529-8535.

Bushnell, P. J., Shafer, T. J., Bale, A. S., et al. (2005). Developing an exposure-dose-response model for the acute neurotoxicity of organic solvents: Overview and progress on in vitro models and dosimetry. *Environ Toxicol Pharmacol 19*, 607-614.

Caldwell, J. C., Evans, M. V., and Krishnan, K. (2012). Cutting edge PBPK models and analyses: Providing the basis for future modeling efforts and bridges to emerging toxicology paradigms. *J Toxicol*, 852384.

Cantòn, I., Cole, D. M., Kemp, E. H., et al. (2010). Development of a 3D human in vitro skin co-culture model for detecting irritants in real-time. *Biotech Bioengin 106*, 794-803.

Catania, J. M., Pershing, A. M., and Gandolfi, A. J. (2007). Precision-cut tissue chips as an in vitro toxicology system. *Toxicol in Vitro 21*, 956-961.

Clewell, H. J., Andersen, M. E., and Barton, H. A. (2002). A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. *Environ Health Perspect 110*, 85-93.

Clewell, H. J., Andersen, M. E., and Blaauboer, B. J. (2008). On the incorporation of chemical-specific information in risk assessment. *Toxicol Letters 180*, 100-109.

Coecke, S., Ahr, H., Blaauboer, B. J., et al. (2006). Metabolism: A bottleneck in in vitro toxicological test development. The report and recommendations of ECVAM Workshop 54. *Altern Lab Anim 34*, 49-84.

Coecke, S., Pelkonen, O., Leite, S. B., et al. (2012). Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. *Toxicol In Vitro*, in press. http://dx.doi.org/10.1016/j.tiv.2012.06.012

Combes, R. D. and Balls, M. (2011). Integrated testing strategies for toxicity employing new and existing technologies. *Altern Lab Anim 39*, 213-225.

Crofton, K. M., Mundy, W. R., Lein, P. J., et al. (2011). Developmental neurotoxicity testing: Recommendations for developing alternative methods for the screening and prioritization of chemicals. *ALTEX 28*, 9-15.

Crump, K. S. and Teeguarden, J. G. (2009). Benchmark calculations from summarized data: an example. *Environ Ecol Stat 16*, 13-24.

Daneshian, M., Leist, M., and Hartung, T. (2010). The Center for Alternatives to Animal Testing – Europe (CAAT-EU): a transatlantic bridge for the paradigm shift in toxicology. *ALTEX 27*, 63-69.

Daneshian, M., Akbarsha, M. A., Blaauboer, B., et al. (2011). A framework program for the teaching of alternative methods (replacement, reduction, refinement) to animal experimentation. *ALTEX 28*, 341-352.

DeJongh, J., Nordin-Andersson, M., Ploeger, B. A., and Forsby, A. (1999). Estimation of systemic toxicity of acrylamide by integration of in vitro toxicity data with kinetic simulations. *Toxicol Appl Pharmacol 158*, 261-268.

Demchuk, E., Ruiz, P., Chou, S., and Fowler, B. A. (2011). SAR/QSAR methods in public health practice. *Toxicol Appl Pharmacol 254*, 192-197.

Dix, D. J., Houck, K. A., Judson, R. S., et al. (2012). Incorporating biological, chemical and toxicological knowledge into predictive models of toxicity. *Toxicol Sci*, in press. doi:10.1093/toxsci/kfs281

Eisenbrand, G., Pool-Zobel, B., Baker, V., et al. (2002). Methods of in vitro toxicology. *Food Chem Toxicol 40*, 193-236.

Ellison, C. M., Enoch, S. J., and Cronin, M. T. (2011). A review of the use of in silico methods to predict the chemistry of molecular initiating events related to drug toxicity. *Expert Opinion Drug Metabol Toxicol 7*, 1481-1495.

Falsig, J., Latta, M., and Leist, M. (2004). Defined inflammatory states in astrocyte cultures: Correlation with susceptibility towards CD95-driven apoptosis. *J Neurochem 88*, 181-193.

Forsby, A., Blaauboer, B. (2007). Integration of in vitro neurotoxicity data with biokinetic modelling for the estimation of in vivo neurotoxicity. *Hum Exp Toxicol 26*, 333–338.

Gabbert, S. and Benighaus, C. (2012). Quo vadis integrated testing strategies? Experiences and observations from the work floor. *J Risk Research 15*, 583-599.

Gülden, M., Mörchel, S., and Seibert, H. (2001). Factors influencing nominal effective concentrations of chemical compounds in vitro: Cell concentration. *Toxicol In Vitro 15*, 233-243.

Gülden, M., Mörchel, S., Tahan, S., and Seibert, H. (2002). Impact of protein binding on the availability and cytotoxic potency of organochlorine pesticides and chlorophenols in vitro. *Toxicology 175*, 201-213.

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

Gülden, M. and Seibert. H. (2006). In vitro-in vivo extrapolation of toxic potencies for hazard and risk assessment – problems and new developments. *ALTEX 23*, 218-225.

Gülden, M., Dierickx, P., and Seibert, H. (2006). Validation of a prediction model for estimating serum concentrations of chemicals which are equivalent to toxic concentrations in vitro. *Toxicol In Vitro 20*, 1114-1124.

Gülden, M., Jess, A., Kammann, J., et al. (2010). Cytotoxic potency of H_2O_2 in cell cultures: Impact of cell concentration and exposure time. *Free Radical Biol Med 49*, 1298-1305.

Håkansson, H., Zilliacus, J., Hanberg, A., et al. (2011). Training of risk assessors in Europe. *Toxicol Letters 205, Suppl*, S262.

Hallier-Vanuxeem, D., Prieto, P., Culot, M., et al. (2009). New strategy for alerting central nervous system toxicity: Integration of blood-brain barrier toxicity and permeability in neurotoxicity assessment. *Toxicol In Vitro 23*, 447-453.

Hartung, T. (2009). Food for thought … on evidence-based toxicology. *ALTEX 26*, 75-82.

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

Hartung, T. and McBride, M. (2011). Food for thought ... on mapping the human toxome. *ALTEX 28*, 83-93.

Hartung, T. and Zurlo, J. (2012). Alternative approaches for medical countermeasures to biological and chemical terrorism and warfare. *ALTEX 29*, 251-260.

HCN – Health Council of the Netherlands (2001). *Toxicity testing: a more efficient approach*. The Hague, The Netherlands: Health Council of the Netherlands. Publication no. 2001/24E, ISBN 90-5549-415-1.

HCN (2008). *Uncertainty factors in risk assessment*. The Hague, The Netherlands: Health Council of the Netherlands. Publication no. 2008/13, ISBN 978-90-5549-720-1.

Heneweer, M., Muusse, M., Dingemans, M., et al. (2005). Co-culture of primary human mammary fibroblasts and MCF-7 cells as an in vitro breast cancer model. *Toxicol Sci 83*, 257-263.

Henn, A., Lund, S., Hedtjärn, M., et al. (2009). The suitability of BV2 cells as alternative model system for primary microglia cultures or for animal experiments examining brain inflammation. *ALTEX 26*, 83-94.

Henn, A., Kirner, S., and Leist, M. (2011). TLR2 hypersensitivity of astrocytes as functional consequence of previous inflammatory episodes. *J Immunol 186*, 3237-3247.

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

Jaworska, J. and Hoffmann, S. (2010). Integrated Testing Strategy (ITS) – Opportunities to better use existing data and guide future testing in toxicology. *ALTEX 27*, 231-242.

Jaworska, J., Gabbert, S., and Aldenberg, T. (2010). Towards optimization of chemical testing under REACH: A Bayesian network approach to Integrated Testing Strategies. *Regul Toxicol Pharmacol 57*, 157-167.

Jaworska, J., Harol, A., Kern, P. S., and Gerberick, F. G. (2011). Integrating non-animal test information into an adaptive testing strategy. Skin sensitization proof of concept case. *ALTEX 28*, 211-225.

Judson, R. S., Kavlock, R. J., Setzer, R. W., et al. (2011). Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem Res Toxicol 18*, 451-462.

Judson, R. S., Martin, M. T., Egeghy, P., et al. (2012). Aggregating data for computational toxicology applications: The U.S. environmental protection agency (EPA) Aggregated Computational Toxicology Resource (ACToR) system. *Int J Molec Sci 13*, 1805-1831.

Kadereit, S., Zimmer, B., van Thriel, C., et al. (2012). Compound selection for in vitro modeling of developmental neurotoxicity. *Front Biosci 17*, 2442-2460.

Kienhuis, A. S., Bessems, J. G. M., Pennings, J. L. A., et al. (2011). Application of toxicogenomics in hepatic systems toxicology for risk assessment: Acetaminophen as a case study. *Toxicol Appl Pharmacol 250*, 96-107.

Kinsner-Ovaskainen, A., Maxwell, G., Kreysa, J., et al. (2012). Report of the EPAA-ECVAM Workshop on the validation of Integrated Testing Strategies (ITS). *Altern Lab Anim 40*, 175-181.

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

Kramer, N. I., Busser, F. J. M., Oosterwijk, M. T. T., et al. (2010). Development of a partition-controlled dosing system for cell assays. *Chem Res Toxicol 23*, 1806-1814.

Kramer, N. I., Krismartina, M., Rico-Rico, Á., et al. (2012). Quantifying processes determining the free concentration of phenanthrene in basal cytotoxicity assays. *Chem Res Toxicol 25*, 436-445.

Krewski, D., Andersen, M. E., Mantus, E., and Zeise, L. (2009). Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Anal 29*, 474-479.

Krewski, D., Acosta, D., Andersen, M., et al. (2010). Toxicity testing in the 21st century: A vision and a strategy. *J Toxicol Environ Health B 13*, 51-138.

Kroes, R., Renwick, A. G., Feron, V., et al. (2007). Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem Toxicol 45*, 2533-2562.

Kuegler, P. B., Zimmer, B., Waldmann, T., et al. (2010). Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing. *ALTEX 27*, 17-42.

Latta, M., Künstle, G., Leist, M., and Wendel, A. (2000). Metabolic depletion of ATP by fructose inversely controls cd95- and tumor necrosis factor receptor 1-mediated hepatic apoptosis. *J Exp Med 191*, 1975-1985.

Leist, M., Hartung, T., and Nicotera, P. (2008a). The dawning of a new age of toxicology. *ALTEX 25*, 103-114.

Leist, M., Bremer, S., Brundin, P., et al. (2008b). The biological and ethical basis of the use of human embryonic stem cells for in vitro test systems or cell therapy. *ALTEX 25*, 163-190.

Leist, M., Efremova, L., and Karreman, C. (2010). Food for thought ... Considerations and guidelines for basic test method descriptions in toxicology. *ALTEX 27*, 309-317.

Leite, S. B., Teixeira, A. P., Miranda, J. P., et al. (2011). Merging bioreactor technology with 3D hepatocyte-fibroblast culturing approaches: Improved in vitro models for toxicological applications. *Toxicol In Vitro 25*, 825-832.

Li, A. P., Uzgare, A., and Laforge, Y. S. (2012). Definition of metabolism-dependent xenobiotic toxicity with co-cultures of human hepatocytes and mouse 3T3 fibroblasts in the novel integrated discrete multiple organ co-culture (IdMOC) experimental system: Results with model toxicants aflatoxin B1, cyclophosphamide and tamoxifen. *Chem-Biol Int 199*, 1-8.

Lock, E. F., Abdo, N., Huang, R., et al. (2012). Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol Sci 126*, 578-588.

Lotharius, J., Falsig, J., Van Beek, J., et al. (2005). Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. *J Neurosci 25*, 6329-6342.

Louisse, J., Bai, Y., Verwei, M., et al. (2010). Decrease of intracellular pH as possible mechanism of embryotoxicity of glycol ether alkoxyacetic acid metabolites. *Toxicol Appl Pharmacol 245*, 236-243.

Marchant, C. A., Briggs, K. A., and Long, A. (2008). In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic. *Toxicol Mech Meth 18*, 177-187.

Maxwell, G., Aleksic, M., Aptula, A., et al. (2008). Assuring consumer safety without animal testing: A feasibility case study for skin sensitization. *Altern Lab Anim 36*, 557-568.

Meganathan, K., Jagtap, S., Wagh, V., et al. (2012). Identification of thalidomide-specific transcriptomics and proteomics signatures during differentiation of human embryonic stem cells. *PLoS ONE 7*, e44228.

NIEHS (2001). National Institute of Environmental Health Sciences Report of the International Workshop on in vitro methods for assessing acute systemic toxicity. NIH Publication 01-4499. NIEHS, Research Triangle Park, North Carolina, USA.

NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and A Strategy*. Washington, D.C.: National Academy Press.

Oravcová, J., Böhs, B., and Lindner, W. (1996). Drug-protein binding studies new trends in analytical and experimental methodology. *J Chromat B: Biomed Appl 677*, 1-28.

OECD (2012). The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. OECD environment, health and safety publications series on testing and assessment. No 168. OECD, Paris.

Paini, A., Punt, A., Viton, F., et al. (2010). A physiologically based biodynamic (PBBD) model for estragole DNA binding in rat liver based on in vitro kinetic data and estragole DNA adduct formation in primary hepatocytes. *Toxicol Appl Pharmacol 245*, 57-66.

Perkel, J. M. (2012). Animal-free toxicology: sometimes, in vitro is better. *Science 335*, 1122-1125.

Pöltl, D., Schildknecht, S., Karreman, C., and Leist, M. (2012). Uncoupling of ATP-depletion and cell death in human dopaminergic neurons. *Neurotoxicol 33*, 769-779.

Prot, J. M. and Leclerc, E. (2012). The current status of alternatives to animal testing and predictive toxicology methods using liver microfluidic biochips. *Ann Biomed Engin 40*, 1228-1243.

Punt, A., Schiffelers, M. J. W. A., Jean Horbach, G., et al. (2011). Evaluation of research activities and research needs to increase the impact and applicability of alternative testing strategies in risk assessment practice. *Regul Toxicol Pharmacol 61*, 105-114.

Renwick, A. G. and Lazarus, N. R. (1998). Human variability and noncancer risk assessment – An analysis of the default uncertainty factor. *Regul Toxicol Pharmacol 27*, 3-20.

Russell, W. S and Burch, R. L. (1959). *The principles of humane experimental technique*. London, UK: Methuen and Co Ltd.

Sand, S., Ringblom, J., Håkansson, H., and Öberg, M. (2012). The point of transition on the dose-effect curve as a reference point in the evaluation of in vitro toxicity data. *J Appl Toxicol*, in press. doi: 10.1002/jat.2757

Schildknecht, S., Pöltl, D., Nagel, D. M., et al. (2009). Requirement of a dopaminergic neuronal phenotype for toxicity of low concentrations of 1-methyl-4-phenylpyridinium to human cells. *Toxicol Appl Pharmacol 241*, 23-35.

Schildknecht, S., Pape, R., Müller, N., et al. (2011). Neuroprotection by minocycline caused by direct and specific scavenging of peroxynitrite. *J Biol Chem 286*, 4991-5002.

Schildknecht, S., Kirner, S., Henn, A., et al. (2012). Characterization of mouse cell line IMA 2.1 as a potential model system to study astrocyte functions. *ALTEX 29*, 261-74.

Schüürmann, G., Ebert, R., and Kühne, R. (2011). Quantitative read-across for predicting the acute fish toxicity of organic compounds. *Environl Sci Technol 45*, 4616-4622.

Seibert, H., Mörchel, S., and Gülden, M. (2002). Factors influencing nominal effective concentrations of chemical compounds in vitro: medium protein concentration. *Toxicol In Vitro 16*, 289-297.

Sipes, N. S., Martin, M. T., Reif, D. M., et al. (2011a). Predictive models of prenatal developmental toxicity from toxcast high-throughput screening data. *Toxicol Sci 124*, 109-127.

Sipes, N. S., Padilla, S., and Knudsen, T. B. (2011b). Zebrafish

– as an integrative model for twenty-first century toxicity testing. *Birth Defects Research Part C – Embryo Today: Reviews 93*, 256-267.

Stiegler, N. V., Krug, A. K., Matt, F., and Leist, M. (2011). Assessment of chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures. *Toxicol Sci 121*, 73-87.

Tralau, T., Riebeling, C., Pirow, R., et al. (2012). Wind of change challenges toxicological regulators. *Environ Health Perspect*, in press. PMID 22871563.

Vaes, W. H. J., Ramos, E. U., Hamwijk, C., et al. (1997). Solid phase microextraction as a tool to determine membrane/water partition coefficients and bioavailable concentrations in in vitro systems. *Chem Res Toxicol 10*, 1067-1072.

van de Bovenkamp, M., Groothuis, G. M. M., Draaisma, A. L., et al. (2005). Precision-cut liver slices as a new model to study toxicity-induced hepatic stellate cell activation in a physiologic milieu. *Toxicol Sci 85*, 632-638.

van Leeuwen, K., Schultz, T. W., Henry, T., et al. (2009). Using chemical categories to fill data gaps in hazard assessment. *SAR QSAR Environ Res 20*, 207-220.

van Midwoud, P. M., Groothuis, G. M. M., Merema, M. T., and Verpoorte, E. (2010). Microfluidic biochip for the perifusion of precision-cut rat liver slices for metabolism and toxicology studies. *Biotechnol Bioeng 105*, 184-194.

Van Summeren, A., Renes, J., Van Delft, J. H. M., et al. (2012). Proteomics in the search for mechanisms and biomarkers of drug-induced hepatotoxicity. *Toxicol In Vitro 26*, 373-385.

van Thriel, C., Westerink, R. H. S., Beste, C., et al. (2012). Translating neurobehavioural endpoints of developmental neurotoxicity tests into in vitro assays and readouts. *Neurotoxicol 33*, 911-924.

Venkatapathy, R., Moudgal, C. J., and Bruce, R. M. (2004). Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. *J Chem Inform Computer Sci 44*, 1623-1629.

Vink, S. R., Mikkers, J., Bouwman, T., et al. (2010). Use of read-across and tiered exposure assessment in risk assessment under REACH – a case study on a phase-in substance. *Regul Toxicol Pharmacol 58*, 64-71.

Watanabe, K. H., Andersen, M. E., Basu, M., et al. (2011). Defining and modeling known adverse outcome pathways: Domoic acid and neuronal signaling as a case study. *Environ Toxicol Chem 30*, 9-21.

Wobus, A. M. and Löser, P. (2011). Present state and future perspectives of using pluripotent stem cells in toxicology research. *Arch Toxicol 85*, 79-117.

Worth, A. P. and Balls, M. (eds.) (2002). Alternative (non-animal) methods for chemicals testing: current status and future perspectives. *Altern Lab Anim 30, Suppl 1*, 1-125.

Yoon, M., Campbell, J. L., Andersen, M. E., and Clewell, H. J. (2012). Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. *Crit Rev Toxicol 42*, 633-652.

Zanella, F., Lorens, J. B., and Link, W. (2010). High content screening: Seeing is believing. *Trends Biotechnol 28*, 237-245.

Zimmer, B., Kuegler, P. B., Baudis, B., et al. (2011). Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing. *Cell Death Diff 18*, 383-395.

Zimmer, B., Lee, G., Balmer, N. V., et al. (2012). Evaluation of developmental toxicants and signaling pathways in a functional test based on the migration of human neural crest cells. *Environ Health Perspect 120*, 1116-1122.

Acknowledgements

We gratefully acknowledge the Doerenkamp-Zbinden Foundation and CAAT-Europe for their support of this workshop.

Correspondence to

Bas J. Blaauboer, PhD
Doerenkamp-Zbinden Chair on Alternatives to Animal Testing in Toxicological Risk Assessment
Institute for Risk Assessment Sciences
Division of Toxicology
Utrecht University
PO Box 80.177
3508 TD Utrecht
The Netherlands
Phone: +31 30 2535330
e-mail: b.blaauboer@uu.nl