altex-29-411-Workshop Report*

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

Bas J. Blaauboer¹, Kim Boekelheide², Harvey J. Clewell³, Mardas Daneshian⁴, Milou M. L. Dingemans¹, Alan M. Goldberg⁵, Marjoke Heneweer⁶, Joanna Jaworska⁷, Nynke I. Kramer¹, Marcel Leisì⁴,⁸, Hasso Seibert⁹, Emanuela Testai¹⁰, Rob J. Vandebriel¹¹, James D. Yager¹², and Joanne Zurlo⁵,¹²

¹Institute for Risk Assessment Sciences, Division of Toxicology, Utrecht University, Utrecht, the Netherlands; ²Department of Pathology and Laboratory Medicine, Brown University, Providence, USA; ³The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina, USA; ⁴Center for Alternatives to Animal Testing—Europe, University of Konstanz, Konstanz, Germany; ⁵Center for Alternatives to Animal Testing, Johns Hopkins University, Baltimore, USA; ⁶Shell International B.V., The Hague, the Netherlands; ⁷Procter & Gamble, Central Product Safety, Strombeek-Bever, Belgium; ⁸Doerenkamp-Zbinden Chair for in vitro toxicology and biomedicine, Faculty of Sciences and Mathematics, University of Konstanz, Konstanz, Germany; ⁹Institute for Toxicology and Pharmacology for Natural Scientists, University Medical School Schleswig-Holstein, Kiel, Germany; ¹⁰Instituto Superiore di Sanità, Environment and Primary Prevention Department, Mechanism of toxicity Unit, Rome, Italy; ¹¹Laboratory for Health Protection Research, National Institute for Public Health and the Environment, Bilthoven, The Netherlands; ¹²Johns 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 Interests Declaration: K. Boekelheide is an occasional expert consultant for chemical and pharmaceutical companies.
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 scientists1 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, HazardenExpert, 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 In alphabetical order: Bas Blaauboer, Alan Goldberg, Thomas Hartung, and Marcel Leist.
### Tab. 1: Glossary of terms used in the context of this paper

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>biomarker of toxicity</td>
<td>A parameter that provides quantitative information that is mechanistically relevant to and predictive of an adverse effect (Boekelheide and Schuppe-Koistinen, 2012)</td>
</tr>
<tr>
<td>biomarker of effect</td>
<td>A parameter that provides quantitative information for an effect but does not necessarily discriminate between adverse and non-adverse effects</td>
</tr>
<tr>
<td>biomarker of exposure</td>
<td>A parameter that provides quantitative information on exposure (<em>in vitro</em>: of the cellular system; <em>in vivo</em>: of organisms)</td>
</tr>
<tr>
<td><em>in vitro</em> biomarker of toxicity (BoT)</td>
<td>An <em>in vitro</em> derived parameter that provides quantitative information that is mechanistically relevant to and predictive of an adverse effect in vivo</td>
</tr>
<tr>
<td>endpoint</td>
<td>The biological or chemical process response of effects assessed by a test method (Leist and Karreman, 2010; Crofton et al., 2011)</td>
</tr>
<tr>
<td>apical endpoint</td>
<td>An empirically verifiable outcome of exposure, assessed in an intact organism (Krewski et al., 2010)</td>
</tr>
<tr>
<td>mechanism of action</td>
<td>The primary chemico-biological interaction between the compound and a structural moiety in the biological system (Blaauboer and Andersen, 2007)</td>
</tr>
<tr>
<td>mode of action (MoA)</td>
<td>Functional and structural changes that occur subsequent to the primary chemico-biological interaction within a biological system, including the resulting clinically observable changes in the organism (Blaauboer and Andersen, 2007)</td>
</tr>
<tr>
<td>pathway of toxicity (PoT)</td>
<td>A cellular response pathway that, when sufficiently perturbed, is expected to result in an adverse health effect (NRC, 2007)</td>
</tr>
<tr>
<td>adverse outcome pathway (AOP)</td>
<td>A pathway of events, starting with a molecular initiating event in which a chemical interacts with a biological target, leading to a sequential series of higher order effects to produce an adverse outcome with direct relevance to a given risk assessment context (Ankley et al., 2010)</td>
</tr>
<tr>
<td>dosimetry</td>
<td>An estimation of the external or internal dose in organisms or tissues resulting from the exposure to a chemical (exposure being a function of dose and time)</td>
</tr>
<tr>
<td>reverse dosimetry</td>
<td>The process of calculating the dose to which an organism would be exposed to produce a concentration in tissues that is equivalent to a concentration measured in an <em>in vitro</em> system</td>
</tr>
<tr>
<td>nominal concentration</td>
<td>The amount of a compound added to the culture medium of an <em>in vitro</em> test system divided by the volume thereof</td>
</tr>
<tr>
<td>point of departure (PoD)</td>
<td>The concentration or dose of a compound that is taken from a concentration- or dose-effect relationship in a test system and is used as a starting point for extrapolations in a risk assessment</td>
</tr>
<tr>
<td>quantitative <em>in vitro-in vivo</em> extrapolation (QIVIVE)</td>
<td>The process of estimating the environmental exposures to a chemical that could produce target tissue exposures in humans equivalent to those associated with effects in an <em>in vitro</em> toxicity test. This calculation is done based on an <em>in vitro</em> concentration-effect relationship and physiologically-based kinetic modelling (Yoon et al., 2012).</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative structure activity relationship</td>
</tr>
<tr>
<td>Area-under-the curve (AUC)</td>
<td>The integral of the curve in a concentration-time or dose-time diagram</td>
</tr>
<tr>
<td>DEREK (Lhasa Ltd)</td>
<td>A knowledge-based system that identifies structural alerts for a wide variety of toxicities and target organs</td>
</tr>
<tr>
<td>TOPKAT (Accelrys)</td>
<td>A QSAR-based toxicity prediction system that contains models for carcinogenicity, developmental toxicity, skin sensitization and various effect levels (e.g., the chronic LOAEL)</td>
</tr>
<tr>
<td>METEOR (Lhasa Ltd)</td>
<td>A knowledge-based system for metabolite prediction, linked to DEREK</td>
</tr>
<tr>
<td>HazardExpert (CompuDrug Ltd)</td>
<td>A knowledge-based system that identifies structural alerts for a wide variety of toxicities and target organs</td>
</tr>
<tr>
<td>Multicase (MultiCase Inc)</td>
<td>A system designed to identify automatically all the molecular fragments that may exist within a set of diverse chemicals tested under a common protocol for any kind of endpoint</td>
</tr>
</tbody>
</table>
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 measurable variable?
- Is it quantifiable?
- Does it represent a chemical-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 (finger-print)?
- 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 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 is characterized by a multitude of parameters that are initially within their homeostatic range. After a chemical insult (indicated by an arrow hitting the horizontal time axis) many of these parameters (e.g., metabolites, transcripts or cell organelle functions) will change in a time-dependent manner. For the selection of relevant BoT, these parameters may be grouped according to their relation to cell fate and hazard of the chemical. The first group reflects the immediate mechanism of action of the chemical (e.g., binding to an enzyme). As chemicals may have multiple targets, the predictivity of changes of one of these early parameters is often low, but it can be useful as BoT, especially for pharmaceuticals. The second group reflects the downstream mode of action (MoA) of the chemical and often is very suitable as BoT. Some parameters change without having a predictive value for the fate of the cell (epiphenomena) or they are cellular counter-regulations of the initial insult. They are not suitable as BoT. In late phases, there is a strong change of parameters, e.g., related to cell death. These appear useful at first sight, but they are often unspecific, and often only reflect a general breakdown of homeostasis. A complicated group of changes is related to altered cellular differentiation. They reflect a new form of homeostasis and are difficult to interpret. They can be useful in the field of developmental toxicity, but their use as BoT requires great care and validation. The gaps in some box outlines symbolize that such changes phase in and phase out at different time points that cannot be sharply 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 explanation see text, section 6.
studied by employing in vitro co-cultures of the relevant cells (Henneveer et al., 2005; Hallier-Vanuxem 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; Kinnser-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 topicaly 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 (Heneweir et al., 2005; Cantón 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 (Crumpl 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).

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 EC50 values may be a good choice as this parameter is the mathematically most robust datapoint to determine. Furthermore, the EC60 or EC90 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.

Fig. 4: Modeling and determination of points of departure (PoD)
The concentration-dependent change of a BoT, which is related to the hazard of the respective chemical tested, is shown. Such curves, obtained from experimental data and modeling, are used to determine the appropriate PoD for quantitative in vitro→in vivo extrapolations (QIVIVE). The PoD is named after its role as starting point for QIVIVE modeling and calculations, and it is expressed as a concentration. It is NOT pre-defined which part of the curve is most appropriate for determination of the PoD. This information depends on the underlying biology and toxicological relevance. For instance IC50, IC90 or lowest observed effect level (LOEL) may be used, if this is toxicologically/biologically justified. Two examples are given: PoD1 is defined by the BMD10, i.e., the lowest concentration that results in a 10% change of the baseline; PoD2 is defined as the IC60 of the curve. Concentrations smaller than the PoD would be considered non-significant with respect to hazard-prediction. Different statistical and modeling approaches are available to define such PoD.
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, 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 PBPK 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 biomolecules (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 toxicological 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


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.


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


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