



Symposium Report

Correlating *In Vitro* Data to *In Vivo* Findings for Risk Assessment

Ambuja S. Bale¹, Elaina Kenyon², Thomas J. Flynn³, John C. Lipscomb⁴,
Donna L. Mendrick⁵, Thomas Hartung⁶, and Geoffrey W. Patton⁷

¹National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA; ²National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA; ³Neurotoxicology and *In vitro* Toxicology Branch, Division of Toxicology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, MD, USA; ⁴National Center for Environmental Assessment, U.S. Environmental Protection Agency, Cincinnati, OH, USA; ⁵Division of Systems Biology, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA; ⁶Johns Hopkins University, Bloomberg School of Public Health, CAAT, Baltimore, USA and CAAT-Europe, University of Konstanz, Germany; ⁷Division of Food Contact Notifications, Office of Food Additive Safety, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, MD, USA

Summary

A special session at the Toxicology and Risk Assessment Conference in Cincinnati, OH, USA in May, 2012 presented approaches to expand current uses of in vitro toxicity data for risk assessment. Evaluation of xenobiotics through use of in vitro study methods is increasing exponentially and these methodologies offer a relatively fast and considerably cheaper way to determine toxicities in comparison to traditional animal-based approaches. One of the challenges with in vitro data is to effectively use this information for risk assessment purposes. Currently, in vitro studies are used as supportive information for hazard characterization and to identify mechanisms associated with toxicity. Being able to effectively correlate in vitro effects with in vivo observations represents a major challenge for risk assessors. The presentations in this special session provided innovative approaches toward effectively using in vitro data for the human health risk assessment process.

Keywords: risk assessment, in vitro-to-in vivo correlation, biomarkers, toxome

1 Introduction (Ambuja Bale and Geoffrey Patton)

This 2012 Toxicology and Risk Assessment Conference (TRAC) session focused on correlating *in vitro* data with *in vivo* findings, a challenge formulated in the National Research Council report *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC, 2007). The major theme of this report was to move away from whole animal testing toward the use of alternative *in vitro* methods, in agreement with the 3R concept proposed by Russell and Burch (1959) to replace, reduce, and refine animal testing. This strategy represents a paradigm shift away from observing adverse effects of high doses of chemicals in experimental ani-

mals and towards looking at whether chemicals activate key toxicity pathways in the species of interest. Recommendations made in the report that were also themes of the TRAC session were to harness the potential of *in vitro* high-throughput assays representing key toxicity pathways and to perform *in vitro* targeted testing studies to complement toxicity pathway analysis. This information could be effectively used beyond hazard identification or mechanism of action analysis to replace the need for data from animal experiments in the human health risk assessment process. The presentations given in the TRAC session gave an update on how *in vitro* data may be correlated with *in vivo* findings to allow implementation of the NRC recommendations.

Received October 1, 2013; accepted in revised form November 8, 2013; Epub November 14, 2013; <http://dx.doi.org/10.14573/altex.1310011>

Disclaimer: The views expressed in this article are the authors' personal opinions and do not reflect those of the U.S. EPA or U.S. FDA.



2 New Approaches to Toxicology (Donna Mendrick)

2.1 Use of Biomarkers to Identify Adverse Effects

The number of adverse drug reactions (ADRs) and deaths caused by drugs is growing faster than the number of prescriptions (Moore et al., 2007). Clearly, toxicity testing in experimental animals does not predict toxicity in humans with full accuracy.

One approach towards increasing the accuracy of predicting toxicity is to identify and validate biomarkers of toxicity. A biomarker is a clinical sign that is objectively measured and evaluated (Biomarkers Definitions Working Group, 2001). It can be, for example, an individual protein, gene, or metabolite whose level changes over the course of a disease. A biomarker of toxicity generally is a biomarker of adversity.

Biomarkers can be used to study the course of disease and the benefit of treatment in patients; they may cross-correlate and are not necessarily unique for a certain disease. In humans, they are primarily measured in body fluids, as it is undesirable to take tissue samples from human patients unless necessary and in many clinical situations it is not possible.

A biomarker of toxicity is most useful for predictive toxicity testing if it is consistently found across several species, including humans, in association with an adverse health outcome and can be tracked in body fluids. Such a biomarker allows translation of effects found in animals to humans.

Qualification of a biomarker entails its correlation with other accepted signs of adverse health effects. For example, a set of urinary protein biomarkers in rats was qualified by CDER (Center for Drug Evaluation and Research) and EMA (European Medicines Agency) against the histological evaluation of tissue samples and against changes in serum creatinine and blood urea nitrogen (BUN). The study found a good correlation between the biomarkers and the histological changes but less so between the biomarkers and the creatinine and BUN values from the serum samples. Therefore, qualifying these biomarkers in patients is challenging in the absence of kidney biopsies, as a good correlation with the serum markers cannot be expected based on the data from the animal experiments. So, how can the biomarker be qualified in patients in the absence of histology data and does the close correlation between the urinary biomarkers and the histology data in the rats suggest that the serum biomarkers measured in patients are under-performing?

Regulatory qualification of a biomarker is a very rigorous process and is contextually driven. This qualification process is generally for a particular type of assay and use. The benefit of following this process is that it leads to acceptance and use of the specific biomarker for regulatory purposes. If a biomarker is not needed for regulatory use, such a qualification process is not necessary, though it does certainly help to establish it as a reliable and relevant research tool in the scientific community. For example, new liver biomarkers of injury are being compared to ALT (alanine aminotransferase) values in serum. However, the performance characteristics of ALT have never been officially qualified and it may therefore not be an ideal gold standard, although it has been used as a marker of liver injury for decades.

Technologies such as genomics, epigenomics, metabolomics, transcriptomics, and proteomics are being used to detect common patterns or molecular signals through examination of tissue, blood, and urine samples that can be qualified as biomarkers. For example, male Sprague-Dawley rats were administered acetaminophen in 0.5% methylcellulose by oral gavage and then sacrificed at multiple time points. Tissue and body fluids from the rats were analyzed and it was found that biomarkers of toxicity representative of the tissue damage could be followed in the body fluids. Therefore, it would be sufficient to take fluid samples at different times from one animal instead of sacrificing one at each time point, thus reducing the number of animals necessary for such a study. Samples were analyzed by metabolomics 6 h, 24 h, 3 days, and 7 days after administration of acetaminophen to include recovery from damage, as in the clinic, a liver transplant may be avoided if there are signs of recovery from liver injury. Glycolic acid was identified as a biomarker of acetaminophen-induced liver injury that normalized quickly with recovery of liver injury. Further research is needed in this area, but this example demonstrates how qualified biomarkers could be employed effectively in a clinical setting.

Metabolomics in body fluids, such as urine, saliva, or blood, can also be used to determine the level and extent of exposure to chemicals. For example, urinary acetaminophen-NAC levels drop considerably 24 h after administration of acetaminophen. Quantification of the levels over time can help to predict when the dose was administered and help with treatment strategies for intoxication.

In terms of epigenetic mechanisms, 10 urinary microRNAs (small, non-coding RNAs) have been identified as biomarkers of hepatotoxicity. MicroRNAs are far more stable in extracellular fluids than messenger RNA. These urinary microRNAs were increased in rats treated with the hepatotoxicants acetaminophen and carbon tetrachloride but were not altered by penicillin or vehicle (Yang et al., 2012). Since acetaminophen and carbon tetrachloride cause different types of injury in the liver, an area of further research might be to compare the pattern of biomarkers induced by the two chemicals.

The biomarker field has expanded and has been useful for establishing exposure to and toxicity of compounds. However, the omic technologies used to identify biomarkers study many parameters in very few biological samples and potential biomarkers arising from such studies must be rigorously qualified. In the MAQC-II project (Shi et al., 2010) 36 expert teams evaluated 6 microarray datasets. The conclusions from this large-scale evaluation were that many different statistical approaches could be used to find biomarkers but correct training and test approaches as well as documentation are required for quality assurance.

A further challenge in transferring a biomarker from experimental animals to the clinic is that relatively homogenous populations are used for animal experiments, which do not represent the variability of the patient population.

2.2 Use of Zebrafish and Stem Cells to Identify Developmental Effects

In the US, about 3% of babies are born with a birth defect. 66% of these defects have unknown causes, 28% are assumed to be



genetic, and about 6% are assumed to be due to diabetes and drugs. The *in vivo* animal study designed to identify developmental toxicity effects of compounds has been in place for 30 years and a higher throughput and cheaper approach, particularly for screening compounds, is needed.

Stem cells and zebrafish are being used increasingly to examine developmental changes caused by compounds, in particular regarding neural development. It is investigated whether compounds added to murine embryonic stem cells that are induced to differentiate into more mature neurons disrupt this process. The expression of pluripotent markers, e.g., Nanog and Sox2, decreases as cells differentiate, while the expression of neuronal markers, such as Nestin and Sox1, increases (Lee et al., 2012). Changes in this process and changes in the morphological differentiation of the stem cells caused by chemicals can be followed.

Zebrafish embryos have very specific developmental stages and their translucence allows evaluation of the organs for signs of developmental toxicity as is being done at the National Center for Toxicological Research. One study found a correlation between F-actin integrity and development (Kanungo and Paule, 2011). High throughput technologies have been developed for this model: Manually dechorionated embryos were placed into individual wells of a 384-well plate at 28 h post-fertilization following ethanol exposure. High throughput imaging documented the effects of ethanol on the axon length (Kanungo et al., 2011). Zebrafish offer exciting opportunities to increase the screening capabilities for new chemical entities although there are several drawbacks including the requirement for the test compounds to be water soluble (Lee et al., 2012).

In summary, new approaches are needed that increase the throughput of toxicity studies. Increasing costs and a concern for using animals are pushing the field to develop higher throughput and less invasive approaches. There is a high barrier to replacing animal studies since the scientific field is accustomed to them. New translational biomarkers are needed in *in vitro* assays, animal studies, and in humans so that exposure and toxicity can be better predicted and to enable the acceptance of alternative models.

3 Mapping the Human Toxome (Thomas Hartung)

Toxicology was stirred up in 2007 by the NRC report *Toxicity Testing in the 21st Century: A Vision and a Strategy* (NRC, 2007), stating that scientists must embrace new technologies from the biotech and bioinformatics revolution. It also puts forward the concept of pathways of toxicity. This concept has been embraced well as a scientific endeavor. An example is a *Science* article, first-authored by Francis Collins of the National Institutes of Health (NIH), in which he said, “we propose now a shift from primarily *in vivo* animal studies to *in vitro* assays, *in vivo* assays with lower organisms and computational modeling for toxicity assessments” (Collins, 2008). Similarly, Peggy Hamburg of the Food and Drug Administration (FDA) says “with an advanced field of regulatory science, new tools including func-

tional genomics, proteomics, metabolomics, high-throughput screening and systems biology, we can replace current toxicology assays with tests that incorporate the mechanistic underpinnings of disease and of underlying toxic side effects” (Hamburg, 2011). In 2009, the National Academy of Science (NAS) Toxicity Testing in the 21st Century (Tox-21c) paradigm was accepted by the U.S. Environmental Protection Agency (US EPA) as its strategy for evaluating chemicals. Therefore, there is an impetus in these new approaches.

The primary concept is that there possibly is a limited number of pathways of toxicity. This theory is supported by the fact that there are certain critical infrastructures in a cell. Only when these infrastructures are hit does this lead to damage, otherwise cells can compensate. If it is possible to identify crucial pathways of toxicity, define them on a molecular basis, and use rather simple testing systems to establish a battery of pathways of toxicity-based tests, fewer tests and animals could be used to assess the toxicity of compounds. So, instead of taking the 450 different cell types of humans into culture and looking at all the possible co-cultures for interactions, it is better to go back to the common denominator – toxicity pathways. For example, oxidative stress is commonly observed as a marker of toxicity and perhaps there really is a rather limited number of pathways of concern. After identifying key pathways, one could start to annotate pathways to different cell types, with the understanding that there will be some overlap. Then we might understand why, for example, the hepatocyte is vulnerable and the cardiomyocyte is not; or how certain groups of toxicants have a certain effect by perturbing a certain pathway of toxicity; or that certain manifestations of cellular phenotype require certain types of disruption. Generating a comprehensive list of toxicity pathways – or a largely comprehensive list – could be useful in understanding why, for example, a rat shows a certain adverse effect and a rabbit does not. This comprehensive list of pathways of toxicity is also known as the Human Toxome (<http://humantoxome.com>).

Perhaps one could conclude, for the first time, that a substance is “clean” or “non-toxic” by finding no effects on the conserved toxicity pathways. If we know that a certain pathway of toxicity needs to be engaged to trigger adverse effects and a substance does not perturb the pathway, we may conclude that the respective hazard can be excluded. Negative results from animal tests cannot definitively lead to this conclusion, since there may be species differences. Similarly, with *in vitro* tests effects may depend on culture conditions and cell types and it is difficult to test all possible permutations and combinations to conclude that a substance is non-toxic.

The National Academies, when talking about pathways of toxicity, consider several different pathways: endogenous hormones, DNA damage, different nuclear receptors, hypo-osmolarity, Nrf-2 to oxidative stress, heat shock proteins. It is very challenging to determine how these toxicity pathways can all be annotated in a common way and put together. The Center for Alternatives to Animal Testing (CAAT) is aiming to map and identify novel pathways and develop a method of annotating these pathways by hazard, chemical/toxin class, cell type, and species (Hartung and McBride, 2011). While these approaches seem to be complementary to approaches by EPA and The



Hamner Institutes, they are actually being developed in a strong collaboration, which means our NIH consortium involves both US EPA's ToxCast and the Hamner Institutes, the other major players implementing the Tox-21c vision.

The Pathways of Toxicity Mapping Center (PoToMaC) at Johns Hopkins University is a vision for an institutionalized effort towards producing a workflow for metabolomics and transcriptomics, soon also microRNAs, as tools for identifying pathways of toxicity. The concept is to start with the existing validated or pre-validated models, which have many advantages compared to other tests. First of all, they have very robust protocols, which have been shown in validation exercises to be good cell models and have received some regulatory acceptance. Reference substances are defined for them and, importantly, thresholds of adversity are defined. Estrogenic endocrine disruption is being examined as a pilot case since toxicity is known to be pathway-based. In the present studies with endocrine disruptors, two assays, i.e., MCF-7 and T47-D, are being used. Obviously, broad testing for endocrine disruptors needs to be done in the near future, so there may be additional benefits. Metabolomics and transcriptomics are primarily being used as well as emerging bioinformatic integration of the two.

At this moment, the concept of a pathway of toxicity (PoT) is not completely developed. The Organisation for Economic Co-operation and Development (OECD) talks about adverse outcome pathways and defines toxicity pathways as a small part of these, i.e., from the chemical biological interaction to cellular hazard manifestation. First, approaches for mapping, annotating, and validating need to be addressed. Finding feasible solutions for the time and spatial resolution is necessary. Some additional needs are identification of distinct pathways and variants of pathways. Governance for the databases to be created and new technologies, other than those few we are now using, are needed to really make up a complete picture of a pathway. It is about moving away from empirical and towards mechanistic approaches. Some predictive models, animal as well as *in vitro* models, are available as well as reference toxicants. In addition, omics phenotyping and other signatures of toxicity could be used to help develop a proof for the existence of pathways. There would not be reproducible signatures if there were not pathways to produce them. So this result is one of the strongest arguments in favor of a limited number of pathways of toxicity. And we might already make some use of Integrated Testing Strategies (Hartung et al., 2013), but the interesting point is, we now can feed this into the identification of pathways of toxicity, because we have at the same time a mechanistic understanding from the past and we also have the molecular biology and biochemistry, which are helping us. This should be later integrated into a systems biology approach of systems toxicology.

When a system is in a certain homeostasis and then it is disrupted, it will counteract this disruption by producing a new, different level of homeostasis under stress, and only if this is exceeded, excessive stimulation will result in some type of exhaustion and subsequently hazard manifestation. This concept can and should be extended to what one would observe at the cellular level. We have a situation where cells in the body or in culture are in a certain homeostasis and then the chemical-

biological interaction hits acutely. If the cell survives this hit, the pathways of defense and a lot of epiphenomena will lead to a new stabilization. But if the hits continue, either over time or if excessive doses are applied, they may lead to chronic manifestations. Currently, typical measurements occur at very late stages where a new homeostasis has been established, because a signature of toxicity is observed at this point. Having information at later stages and not earlier stages is one of the key problems, since the initial insult (or hit) and the new equilibrium (from the hit) are two different things.

Toxicometabolomic strategies have been used to start identifying and elucidating pathways of toxicity (Bouhifd et al., 2013). An advantage is that a lot of other information is available, especially with regards to known mechanisms and identified modes of action. The pathway analysis can build on many decades of toxicology and mechanistic toxicology. However, there are several challenges, including bringing together the different areas of biology, mass spectrometry, and especially bioinformatic pathway identification. Additionally, complex detection methods, the sample handling, the throughput, the reproducibility, and toxicology information are needed. The end goal is to produce a Human Toxome database into which, like a "Wiki" approach, scientists can feed in further pathways identified in a similar way. This all is not only about producing technological results. It is about coalition and concept forming. It is about building a framework in which these things can happen.

Identification of pathways of toxicity is a very important issue that is of concern worldwide. For example, Europe is concerned about a ban on systemic toxicity testing for cosmetics ingredients that came into force in 2013. In a report invited by the European Commission (Adler et al., 2011), the concern was with regards to replacement of animal studies with alternative methods and that the scientific field has not been able to reach that threshold. An independent review of this report was conducted where it was concluded that systemic toxicity tests in cell cultures and computers are not available (Hartung et al., 2011). This spurred thirty-five expert scientists to develop a consensus roadmap for the development of alternative methods (Basketter et al., 2012). These recommendations highlight the need for identifying pathways of toxicity and moving forward toward alternative toxicity testing.

4 *In Vitro* to *In Vivo* Extrapolation for Application in Physiologically Based Pharmacokinetic (PBPK) Models **(John Lipscomb and Elaina Kenyon)**

Toxic responses are a function of the concentration of the toxicologically active chemical moiety in the target tissue. Distribution of the toxicant to the tissue is broadly regarded as toxicokinetics (TK) and the events surrounding/driving the response of the tissue can be broadly described as toxicodynamics (TD). For TK, *in vitro* systems offer the advantage of studying individual steps or processes (metabolism, transport) in isolation; the challenge is to integrate observations into a system representing the intact organism so that they can be

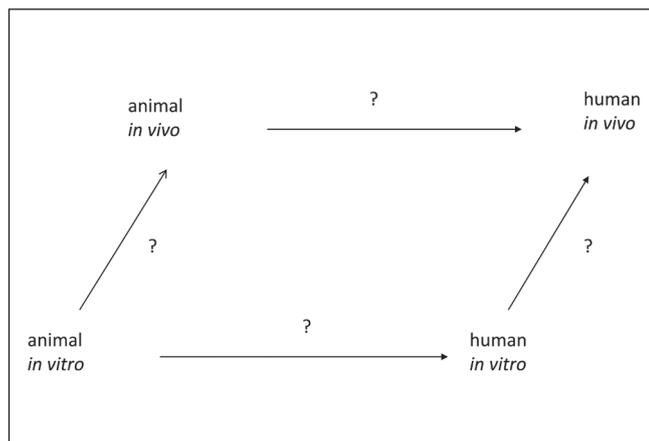


Fig 1: The parallelogram approach involves comparison of *in vitro* observations in a test species with *in vivo* findings from the same test species

The results of this comparison are applied to *in vitro* observations with human samples to predict the human response *in vivo*.

appropriately interpreted. For TD, *in vitro* systems offer the advantage of studying the response under circumstances under which the TK influence can be controlled. One of the primary challenges for TD is that the observed responses must be linked to the adverse health outcome of interest, which may require an in depth understanding of the mode or mechanism of action. Regardless of the application, *in vitro* systems offer the advantage of studying the human species in the absence of constraints surrounding intentional human exposures (Lipscomb and Poet, 2008). Often, a parallelogram approach is employed in which *in vitro* observations in a test species are compared with *in vivo* findings from the same test spe-

cies (Fig. 1). The results of this comparison are applied to *in vitro* observations with human samples to predict the human response *in vivo* (Sobels, 1989). This section reviews some *in vitro* systems used for TK and demonstrates the utility of the resulting data for application in PBPK models.

Toxicokinetics is typically defined as absorption, distribution, metabolism, and elimination. Transport across cell membranes and transporter mechanisms are important in each process, but this section will focus on the study of metabolism because IVIVE (*in vivo* to *in vitro* extrapolation) for metabolic rate parameters has been more broadly applied in the field of PBPK modeling. Several systems are available for the *in vitro* investigation of metabolism, and these cover a broad range of biological complexity. These systems include the isolated perfused liver, tissue slices, hepatocytes, subcellular fractions (microsomes and cytosol) and recombinantly expressed enzymes (see Tab. 1). Their characteristics, advantages, and limitations have been well summarized in Barter et al. (2007) and Lipscomb and Poet (2008).

The isolated perfused liver offers the most complex and physiologically realistic system – it maintains architecture at the level of the cell and tissue, maintains the balance of cytoplasmic and membrane-bound enzymes, cellular transport mechanisms, as well as physiologic levels of cofactors. Blood or medium flow can be well regulated and blood or medium can be adjusted to mimic the balance of constituents that determine extrahepatic protein binding.

Tissue slices have been widely used and are relatively easy to prepare. They also are available commercially, fresh or frozen, from experimental animals and humans. Slices maintain cellular architecture and the balance of cytoplasmic and membrane-bound enzymes, and may maintain physiologic balance of cofactors, but they do not maintain tissue blood flow. Slices can be used to study TK or TD events, but toxicant exposure may be problematic due to the need for toxicants to diffuse through the

Tab. 1: *In vitro* systems used for the study of metabolism in liver^a

<i>In vitro</i> system	Characteristics and enzymes present	Scaling factor ^b
Isolated perfused liver	Intact cellular and organ structure with complete enzyme and physiological milieu	Liver weight
Tissue slices	Microsomal and cytosolic enzymes plus cellular structure including membrane and biliary transporters	HomPPGL (mg homogenate protein per g liver)
Hepatocytes	Microsomal and cytosolic enzymes plus cell membrane transporters	Hepatocellularity per g liver (HPGL)
Microsomes	Cytochrome P-450s, flavin-containing monooxygenases, esterases, glutathione S-transferases (GSTs), glucuronyl transferases	Microsomal protein per g liver (MPPGL)
Cytosol	Sulfotransferases, N-acetyltransferases, amino acid conjugators, GSTs, glucuronyl transferases	Cytosolic protein per g liver (CPPGL)
Expressed or purified enzymes	Enzyme of interest with constitutively expressed ancillary proteins or native lipid environment	MMPGL (or CPPGL) x enzyme isoforms (pmol/mg)

^a Lipscomb and Poet (2008)

^b Barter et al. (2007); liver weight is a required scaling factor in all cases.



outer layers of cells to get to the inner layers of the cells. Therefore, most of the metabolism may occur in the outer layers only, complicating not only internal exposure estimation, but also the representation of metabolic rate, which is routinely based on protein or enzyme recovery in the entire slice.

Isolated hepatocytes are often prepared in the laboratory from test species and are available from human organ donors or from surgically resected tissues. Recent advances have resulted in methods to cryopreserve isolated hepatocytes, which offer the advantage of a degree of consistency from experiment to experiment. While this preparation does not preserve architecture at the tissue level, the physiologic ratio of cytoplasmic to membrane-bound enzymes is maintained and physiologic balances of cofactors can be attained.

Perhaps the most widely used *in vitro* system is that of sub-cellular fractions, i.e., cytosol and microsomes. Microsomal protein (MSP) comprises microsomal lipid and membrane bound proteins (enzymes) and preparation requires isolation from other cellular components by homogenization and centrifugation. This separation from other cellular components allows individual steps to be studied without the influence of competing factors – the major advantage and disadvantage combined. While these systems are relatively cost-effective, the disadvantages include a lack of physiological balance between cytosolic and membrane bound enzymes, lack of maintenance of physiologic levels of cofactors, an absence of transport mechanisms and an often overlooked complication of the membrane arrangement (geospatial co-localization) of enzymes and co-enzymes (e.g., cytochrome P-450 (CYP) enzymes and their oxidoreductases).

Recombinantly expressed systems offer the advantage of studying actual human enzymes without the complications and precautions required by exposure to actual human tissues. The optimal application of recombinantly expressed enzymes may be in the identification of metabolites of a given compound, or in the confirmation of the activity of the studied enzyme toward a given substrate. Comparisons of animal to human metabolism are complicated due to the infrequent recombinant expression of corollary enzymes from test species. The quantitative usefulness of metabolic rate data from recombinant enzymes is complicated by differences in the lipid constituency of the biological membrane in which the enzyme is bound (complicating examinations of substrate solubility/exposure to the active site of the enzyme) versus the constituency of the human lipid membrane; as well as alterations of any associated co-enzymes. With respect to cytochrome P-450 enzymes, human oxidoreductases may not be expressed and the geospatial co-localization of P-450 enzymes and oxidoreductases can be altered (Lipscomb and Poet, 2008).

The need for IVIVE for metabolic rate parameters has become increasingly important because the application of PBPK models has become relatively widespread in health risk assessment. PBPK models can estimate target tissue concentrations of toxicants following exposure via the oral, dermal, and inhalation routes and are useful in risk assessment because they can extrapolate these tissue concentrations across dose, route, time, and species. These models comprise a series of simultaneously based, simultaneously solved ordinary differential equations.

The equations are populated with parameters that describe the anatomy (e.g., compartment size, spatial relationships), physiology (e.g., blood flow, specialized function) and biochemistry (e.g., uptake, partitioning, metabolism) of the experimental animal or human and its toxicologically and physiologically significant organs and tissues. Metabolism parameters (e.g., V_{max} , K_m , and Cl) can be among the most influential parameters determining the target tissue concentrations of a toxicant (Clewell et al., 2008). V_{max} and K_m are interdependent constants describing saturable metabolism. V_{max} is the theoretical maximal initial rate of metabolism, and K_m is the concentration that drives reaction rates to one-half the rate at V_{max} (Figure S1 in supplementary data at <http://www.altex-edition.org>). V_{max} is determined by holding the amount of enzyme constant and increasing the amount of substrate until no further increase in metabolic rate is observed. V_{max} is directly proportionate to the amount of enzyme present, but K_m is independent of enzyme concentration. K_m can be influenced by the solubility of the substrate in solution and by other factors, including binding to macromolecules in addition to the enzyme. *In vitro* systems provide a means to control extrinsic factors that can influence the determination of these parameters (Lipscomb and Poet, 2008).

Aside from a full toxicokinetic treatment of metabolism as offered through PBPK modeling, another measure that can be used is intrinsic clearance (Cl_{int}). Cl_{int} is estimated by dividing V_{max} by K_m (Fig. S1). When units of measurement are mg/h/mg MSP and mg/l, the expression reduces to liters cleared per hour per mg MSP, and is scaled to the whole liver by accounting for the yield of MSP per g and per liver (Houston, 1994; Carlile et al., 1997). Cl_{int} is often measured for drugs or compounds in clinical development, and experiments are often developed that also include means to control for or to estimate the impact of binding of substrate to non-enzymatic proteins. Cl_{int} estimates can be useful in estimating doses to be used in preclinical drug assessments and have proven quite valuable for that application (Houston, 1994). However, measurements of Cl_{int} must be rather cautiously interpreted beyond this circumstance because if the underlying experiments fail to account for binding, solubility, or reactivity, the results can be unreliable.

While both K_m and V_{max} can be extrapolated to the *in vivo* setting, maximal metabolic rates may not be observed *in vivo* for several reasons. One of the most influential factors is the delivery of the substrate to the liver via hepatic blood flow. Metabolic rate may be restricted by the rate at which the substrate is delivered to the liver, a function of both the solubility of the chemical in blood and the rate of hepatic blood flow (Q_L). Under this condition, called flow-limited metabolism, the metabolic capacity of the liver is greater than the rate of substrate delivery to the liver in blood. This is often seen with halogenated solvents, and has been well characterized by physiologically based pharmacokinetic modeling (Kedderis, 1997). Table 2 illustrates the issues with blood flow detailed above and explains why measures of intrinsic clearance, in contrast to hepatic clearance (Cl_H), may not always be the best measures of *in vivo* clearance. Cl_H has another term in it relative to $V_{max} / K_m - Q_L$. Cl_H is actually the model whose structure is incorporated into PBPK models.

Tab. 2: The impact of hepatic blood flow (Q_L) on hepatic clearance (Cl_H) of xenobiotics

Q_L (L/h)	V_{max} (mg/h)	K_m (mg/l)	Cl_{int} (l/h)	Cl_H (l/h)
0.005	5	10	0.5	0.00495
0.05	5	10	0.5	0.04546
0.5	5	10	0.5	0.25
5	5	10	0.5	0.4545
50	5	10	0.5	0.4951
500	5	10	0.5	0.4995

$Cl_H = (Q_L \times V/k) / (Q_L + V/k)$. When hepatic blood flow is much lower than V/k , flow (delivery to the liver) determines the overall clearance rate.

In the context of metabolism, *in vitro* systems can be used qualitatively as well as quantitatively. Qualitatively, because MSP contains multiple enzymes and families of enzymes, it can be used in conjunction with specific chemical and immunologic inhibitors to identify enzymes associated with specific metabolic processes. Quantitatively, metabolic rate constants can be identified through kinetic analysis and then extrapolated back to the intact organ. This requires that the amount of MSP per unit mass of tissue be determined. One favored method is to select an enzyme or protein expressed only in MSP and focus on determining the amount of that enzyme in intact tissue and in resulting MSP. Using CYP2E1 as the example, an ELISA has been developed and applied (Snawder and Lipscomb, 2000). Quantitatively, the equation is:

$$(\text{nmol CYP2E1/g tissue}) / (\text{nmol CYP2E1/mg MSP}) = \text{mg MSP/g tissue}$$

Once V_{max} values have been determined in MSP, they are readily converted from units of mg/min/mg MSP to mg/min/g liver (MMPGL) by simple multiplication. Several values for the amount of MSP per g liver are available that have been generated using various proteins and enzymes (Table S1 in supplementary data at <http://www.altex-edition.org>).

The reliability of PBPK model predictions is directly related to the accuracy of the chemical-specific parameters used as model inputs. A fundamental question in regards to IVIVE for metabolic rate constants is how well and to what extent parameters derived from *in vitro* data predict *in vivo* metabolism. In an effort to systematically begin examining this issue for environmental chemicals, a preliminary analysis using data from the published literature was conducted. *In vivo* estimates of $V_{max} C$ (mg/h/kg) from *in vivo* vapor uptake studies to scaled-up $V_{max} C$ values derived from *in vitro* studies of microsomal metabolism were compared. Criteria for chemical selection were designed to minimize potential sources of variation. Use of a common set of techniques for *in vivo* data (vapor uptake) and *in vitro* data (microsomal metabolism) that measures disappearance of parent chemical from the atmosphere of either a closed recirculating chamber containing a rat or a vial containing microsomes

Tab. 3: Preliminary comparison of *in vivo* and scaled *in vitro* metabolism parameters^a

Chemical	Parameter	Units	<i>in vivo</i>	<i>in vitro</i>	Ratio
1,1-dichloropropene	$V_{max}C$	mg/h/kg	5.58	7.22	0.77
1,2-dichloroethane	$V_{max}C$	mg/h/kg	3.25	5.50	0.59
2,2-dichloropropane	V_{max}/K_m	l/h/kg	1.55	0.80	1.94
bromodichloromethane	$V_{max}C$	mg/h/kg	12.8	2.65	4.83
chloroform	$V_{max}C$	mg/h/kg	6.8	3.77	1.80
toluene	$V_{max}C$	mg/h/kg	7.5	2.93	2.56
benzene	$V_{max}C$	mg/h/kg	3.3	2.05	1.61

^aKenyon et al. (2012). Ratio calculated as *in vivo/in vitro*.

was one criterion. Variation due to inherent species, strain, gender, and age differences was minimized by using data derived only from mature male F344 rats. Results of this comparison are shown in Table 3 (Kenyon et al., 2012).

Inspection of this table reveals that most of the estimates are within a factor of two to three of each other as shown in the column labeled "Ratio". Given the various assumptions inherent in this approach, this degree of agreement is fairly good. Some of these assumptions will be highlighted and described subsequently, but for now the chemical to focus on is bromodichloromethane (BDCM), a drinking water disinfection byproduct that has been shown to cause cancers of the colon and kidney in rats. It is noteworthy that the *in vivo* and *in vitro* scaled estimates for $V_{max} C$ differ by a factor of 4.8 for BDCM, with the *in vitro* scaled to *in vivo* estimate being much lower. One possible explanation for this difference is the assumption that *in vitro* metabolism in liver quantitatively accounts for most *in vivo* metabolism. Comparatively lower estimates of $V_{max} C$ on the basis of hepatic microsomal data might be explained by quantitatively significant extrahepatic metabolism. Although extrahepatic metabolism is frequently thought of as being of mechanistic significance for organ-specific toxicity, it may in some cases be quantitatively significant enough to impact whole body metabolic clearance which is what is actually measured by the *in vivo* vapor uptake technique.

Obvious sources of uncertainty in this type of extrapolation are the scaling factors used in the calculation. In this case that would be mg of microsomal protein per g liver (MMPGL) and liver weight (LW). Typically, rates of microsomal metabolism are reported normalized to mg of microsomal protein. In order to scale these rates to the *in vivo* case, it is necessary to multiply the rate of metabolism by MMPGL and LW to scale it to the whole body. In practice, these factors are rarely measured and reported for the experimental conditions under which the *in vitro* metabolism data were generated. Thus it is generally necessary to use default scaling factors (Table S1 in supplementary data at <http://www.altex-edition.org>) which may vary due to both intrinsic host factors and experimental conditions.



Another noteworthy assumption embedded in IVIVE for metabolic rate constants is that other physiological processes such as binding to tissue constituents and blood flow to the liver do not significantly alter the concentration of chemical available for metabolism as discussed previously. The impact of these physiological processes can sometimes be evaluated in companion *in vitro* assays (e.g., measuring binding to specific proteins) or their impact assessed via sensitivity analysis combined with the use of PBPK models. The key factor to remember is that IVIVE must be conducted and considered in the context of the experimental system used to generate the data and any chemical-specific attributes (e.g., solubility, reactivity) that may influence the interpretation of data obtained from these systems. Some of these issues and impacts are further explored in the remainder of this section.

When evaluating metabolic rate constants in a single organ, units of expression for metabolic rate may be sufficiently described as rate per unit subcellular fraction (e.g., nmol/min/mg MSP). However, this level of expression is problematic when attempting to extrapolate measured *in vitro* metabolic rates to other organs, tissues, or species. There are two reasons for this: 1) different tissues contain different concentrations of MSP, and 2) MSP from different tissues contains different concentrations of individual enzymes. Table S1 illustrates rat-human differences in measured MSP per g liver; Table S2 demonstrates some differences in MSP content for specific CYP enzymes. When it can be determined that metabolism proceeds via a single enzyme, this complication can be overcome by determining the amount of the given enzyme per unit mass of tissue, and avoiding the estimation of MSP content altogether, but it requires one additional step – metabolic rate determined initially as rate per mg MSP must be converted to units of rate per pmol enzyme. This procedure and its application for IVIVE are illustrated in Figure 2 for trichloroethylene (Lipscomb et al., 2003). This procedure has also recently been applied in the IVIVE of the CYP2E1-mediated hepatic and renal metabolism of chloroform (Sasso et al., 2013).

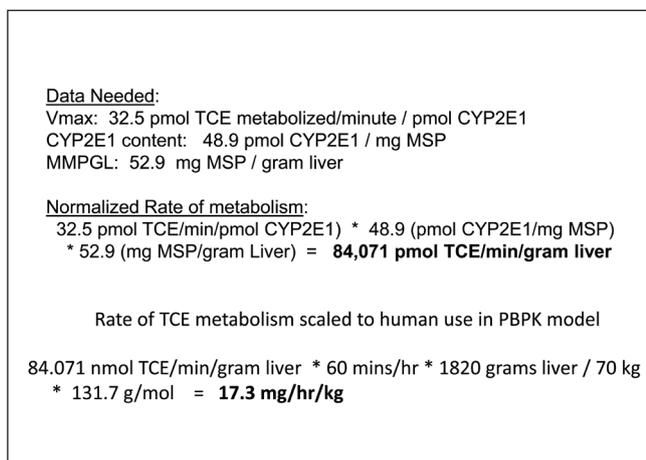


Fig 2: Illustration of IVIVE scaling calculations for trichloroethylene based on microsomal metabolism measurements

Another important issue to consider is parameter variability and its impact on IVIVE and dose metric estimation in the context of PBPK model application. An example of a study on variability and the influence of enzyme content is an analysis of the variability of CYP2E1 and resulting influence on the oxidation of trichloroethylene (TCE) (Lipscomb et al., 2003). Because the oxidation of TCE results in the formation of toxicologically active metabolites and because CYP2E1 content was known to vary among humans, an analysis was conducted to capture human variability of hepatic CYP content and include this variability in a PBPK model (Fig. 3). The amount of CYP2E1 in human hepatic MSP samples was determined via ELISA. The results indicated a range of approximately 12-fold (11-134 pmol CYP2E1/mg MSP) from a sample of 60 individuals. For 20 individuals, the amount of MSP per g liver (25.2-103.5 mg MSP/g liver) was determined. A statistical combination of these data sets according to the statistical method of moments indicated a variability of hepatic CYP2E1 (1231.7 - 5004 pmol CYP2E1/g liver) of approximately 4-fold. A biochemical analysis was used to quantify the specific activity of CYP2E1 toward TCE (GSM \pm GSD: 32.5 \pm 1.538 pmol TCE metabolized/min/pmol CYP2E1), and was also included. A PBPK model was reparameterized to include the bounds of variability and exercised to simulate exposure at regulated environmental contaminant levels. The result is that a 6-fold difference in metabolic capacity (V_{max} ; pmol TCE oxidized/min/g liver) resulted in a 2% change in TCE oxidation *in vivo*. The underlying factor in mediating the effect was hepatic blood flow – for TCE, the hepatic metabolic capacity far outstrips the rate of delivery to the liver. This case is illustrative and demonstrates the value of sensitivity analysis in identifying the most influential causes of toxicokinetic variability. However, the data describing the variability of human hepatic CYP2E1 content can be used in other models for CYP2E1 substrates, where the specific activity of CYP2E1 for the given substrate is known.

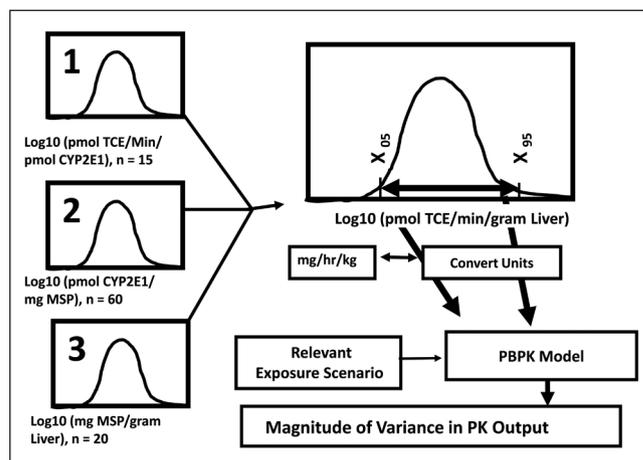


Fig 3: Framework for extrapolation and incorporation of *in vitro* derived metabolic rates in PBPK modeling
 The model was exercised to simulate environmentally and occupationally relevant exposures.

5 Building “Reality” Into *In Vitro* Hepatotoxicity Assays (Tom Flynn)

One of the major barriers to the acceptance of *in vitro* assays is the perception that the assays do not reflect reality. After all, how could a two-dimensional growth of cells in a Petri dish be informative about three-dimensional animals that have multiple organs with multiple interactions between them? However, there are some ways that scientists can build enough “reality” into *in vitro* assays to make them possibly more acceptable to the risk assessment community.

The first question is how does one address known *in vivo* realities in an *in vitro* model? For example with *in vitro* hepatotoxicity assays, one could start off by developing a list of desirable properties of what the *in vitro* model for hepatotoxicity should look like, especially that it should display as many human liver-specific functions as possible. The model should focus on post-mitotic cells. People typically and mistakenly use hepatic cell lines while they are still actively proliferating, and that would not model reality since the adult liver is almost exclusively post-mitotic. Also, with respect to cell lines, a low-glucose medium should be used instead of the commonly used high-glucose medium, since it is well known that high glucose medium can mask mitochondrial toxicity. In choosing doses of a test compound, one should use a “reasonable” maximum concentration in the test system based on expected exposures in humans. And finally, if multiple concentrations of test agent are used, take heed of the concentration response, since it is less informative to only report the EC₅₀ and not the effects at other concentrations. In building reality into *in vitro* assays, one has to first consider cell lines that accurately simulate mechanisms of toxicity and second, develop a good prediction model. In this discussion, the focus is on applying these concepts to *in vitro* hepatotoxicity assays.

One of the cell lines used to study hepatotoxicity *in vitro* is HepG2/C3A, which is a sub-clone of the commonly used HepG2 human hepatocellular carcinoma that was developed by Kelly (1994). One little-appreciated property of HepG2 cells is that they will undergo differentiation *in vitro* if grown to confluence. The doubling time, which is approximately 24 h while they are in log growth phase, increases to over 200 h after confluence. While this is not totally post-mitotic, it is good enough for modeling a post-mitotic tissue. Other desirable properties

of HepG2/C3A cells are that they synthesize and excrete physiological levels of albumin but little or no alpha-fetoprotein, they can be grown in glucose-free medium, they have low but measurable activities of all the major cytochromes P450, and they are commercially available.

For building reality into the hepatotoxicity assay, it was important that the HepG2/C3A line is capable of recapitulating as many of the known mechanisms of liver toxicity as possible. As listed in Table 4, this system has been used to model many of the known mechanisms of liver toxicity. At the plating density used, $6 \times 10^4/\text{cm}^2$, the cells reach confluence in 4 to 5 days. The cells are allowed another 3 days to differentiate, and treatment of the cells with drugs or chemicals is typically started on day 8. Figure S2 (in supplementary data at <http://www.altex-edition.org>) depicts HepG2/C3A cells at culture day 8. Unlike many transformed cells, these cells do not pile up on each other, and they retain many of the characteristics of normal liver cells.

Following the development of a well-characterized biological system (e.g., HepG2/C3A cells), it was important to develop a good prediction model by first creating a gold standard list of “known” positives and negatives that can be used to determine the operating parameters of the assay, specifically parameters such as sensitivity, specificity, and positive prediction value. To create a short and fairly comprehensive list, the criteria for selection focused on liver toxicants for which human data were available. The responses observed after treatment with the known liver toxicants were highly concordant with the known *in vivo* effects. One particularly interesting observation with the pharmaceutical human liver toxicants was that the lowest concentration producing a toxic response in the *in vitro* assays was at the upper level of the therapeutic blood concentration range.

One of the challenges with *in vitro* testing is analyzing and interpreting the data in an effective manner. The statistician evaluated a number of standard mathematical models and analyses for interpreting the data that were generated with the HepG2 cell line assays. Ultimately, the goal was to review the mathematical models and determine the best predictive model for the *in vitro* assay. Also, since multiple endpoint assays were being used, it was important to demonstrate that there was very little or no cross-correlation between assays, as highlighted in Table S3 (supplementary data at <http://www.altex-edition.org>).

Tab. 4: Endpoint assays and the mechanisms of hepatotoxicity modeled

Endpoint assay	Hepatotoxicity mechanism modeled
Total double-stranded DNA (H33258 fluorescence)	Cell death (necrotic)
Rhodamine 123 uptake and retention	Mitochondrial membrane depolarization, P-glycoprotein induction/ inhibition
Nile red uptake	Steatosis (“fatty liver”), phospholipidosis
Dihydrodichlorofluorescein oxidation	Oxidative stress
Ethoxyresorufin-O-deethylase (predominantly CYP1A)	Modulation of CYP450 activities
Benzoyloxyresorufin-O-debenzylase (predominantly CYP2B and CYP3A)	Modulation of CYP450 activities



After developing the prediction model, it was tested using another group of model compounds as listed in Table S4 (supplementary data at <http://www.altex-edition.org>). As is also shown in Table S4, the uniqueness of the predictive model is reflected by how the response is evaluated at every concentration tested. The predictive model was able to distinguish and differentiate between similar compounds like androstenedione and testosterone, and between some of the flavonoids listed in Table S4. A key point here is that, out of all the compounds that were evaluated in this system, either in the model development set or in the validation set, 100% cell lethality was only observed with propranolol at the two highest test concentrations. Thus, this model really focuses mostly on overall metabolic changes that reflect pathways of toxicity because multiple endpoints are being examined which capture multiple physiological pathways within these cells.

Another factor that was built into this liver toxicity prediction model takes into account the reality that not all dose responses are monotonic. In fact, it has been shown that as many as 20% to 30% of the endpoints measured *in vitro* have a non-monotonic or U-shaped dose response. The predictive model detected and interpreted these responses (see Tab. S4).

In the *in vitro* HepG2/C3A cell prediction model that was developed, one of the objectives was to find the lowest liver active concentration and to determine whether or not it could be interpreted as a lowest observed effect level (LOEL), a value that is of interest to risk assessors and would be meaningful for risk assessment of chemicals and drugs. Another advantage in this approach was that the gold standard list was based exclusively on human data so that the validation was against only one species, which is also the same species of origin of the liver cells. The actual parameter returned by the prediction model is the probability of a binary classifier being liver active or not for each observation, thereby allowing for an estimation of the percentage of observations indicating liver activity at each concentration rather than a simple overall dichotomous classification.

Known sex differences in liver toxicity also can be modeled in an *in vitro* hepatotoxicity assay. One of the most obvious differences between men and women is the difference in blood levels of steroid hormones. To build more “reality” into the model, mixtures of hormones that were consistent with those in both human males and females were added to the culture system. Another parameter that might be related to sex differences in liver toxicity is inflammation because the immune systems of men and women are very different. For example, the incidence of autoimmune disease is much higher in women, and women recover from sepsis much faster than men. In modeling the immune response for the liver, the liver Kupffer cells, a class of macrophage, were considered. Kupffer cells respond to any kind of inflammatory challenge by releasing pro-inflammatory cytokines, primarily IL-1 β , IL-6, and TNF- α . To add an *in vivo* reality to the *in vitro* model, a mixture of these cytokines was included that is consistent with an inflammatory state in the human liver.

Acute phase proteins released by liver in response to inflammation, such as albumin and interleukin-1 receptor antagonist (IL-1ra), were measured in the *in vitro* assay because these pro-

teins are expressed in HepG2 cells. However, HepG2 cells do not express C-reactive protein, which was the primary reason this protein level was not studied in these assays. But IL-1ra has been shown to be equally as responsive and informative as C-reactive protein. With the albumin response, a significant interaction between the hormones and the cytokines was observed as indicated in Table 5. The next step in this series of studies is to understand the impact of the hormonal composition and the presence or absence of cytokines on the response to known liver toxicants.

In terms of building further reality into an *in vitro* hepatotoxicity assay, one needs to consider the marked prevalence of non-alcoholic fatty liver disease. There are several models that could be incorporated to make the *in vitro* system more representative of some *in vivo* factors. In the case of *in vitro* hepatocytes, adding an excess amount of fatty acids to the culture medium for 48 h will result in the cells accumulating neutral lipid similar to a fatty liver, as depicted in Figure S2 (supplementary data at <http://www.altex-edition.org>). Prior to treatment with oleic acid, there is a uniform distribution of neutral lipids (yellow/green fluorescence) and polar lipids (red fluorescence) with a Nile red stain. After oleic acid treatment, there is a significant shift to primarily neutral lipid droplets. In Figure 4, a dose-dependent increase in total neutral lipid stored in the cell is observed with increased

Tab. 5: Effects of sex hormones and cytokines on acute phase proteins in HepG2 human hepatoma cells

Hormones	Cytokines	Acute phase protein % control \pm SE	
		Albumin	IL-1ra
None	–	100 \pm 1	100 \pm 3
None	+	89 \pm 1	553 \pm 41
Female	–	100 \pm 2	112 \pm 6
Female	+	88 \pm 2	625 \pm 47
Male	–	129 \pm 6	133 \pm 4
Male	+	85 \pm 2	723 \pm 26
	Effect	Albumin	IL-1ra
	Hormones	0.0794	0.0903
	Cytokines	0.0021	0.0396
	Interaction	0.0093	0.1813

HepG2/C3A cells were grown for 8 days in either male hormone, female hormone, or sex hormone-free medium, then exposed to a mixture of pro-inflammatory cytokines (IL-1 β , IL-6, TNF- α) for 72 h. The amounts of human albumin and human IL-1ra in the medium were measured by ELISA. Data are expressed as percent of control (no hormones, no cytokines) mean \pm SE. Statistically significant ($p < 0.05$) effects are shown in bold. Data are from Flynn and Ferguson (2010).

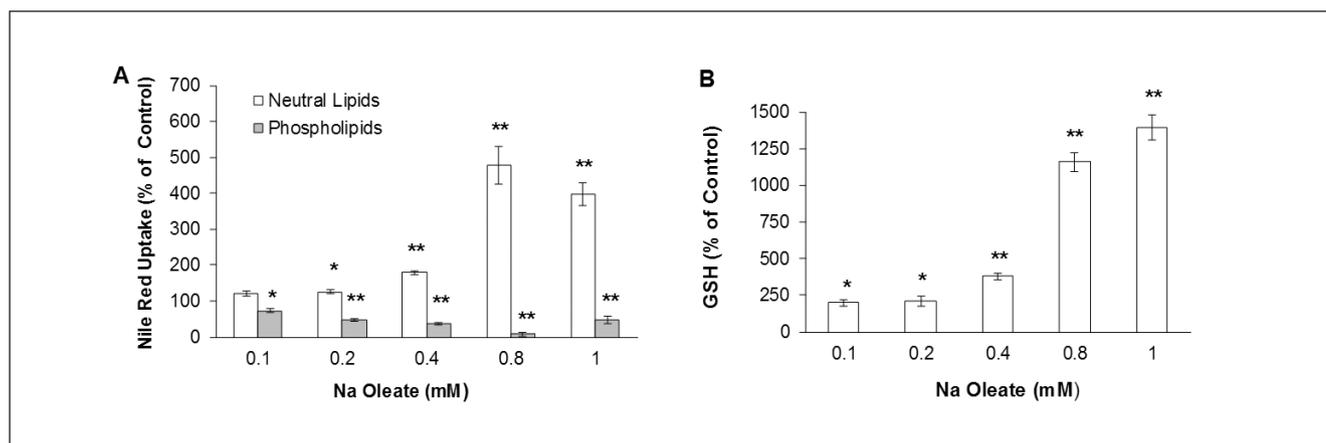


Fig 4: Effects of 48 h treatment with oleic acid on HepG2/C3A cells

A) Effects of oleic acid on neutral and polar lipid content. Neutral and polar lipids were quantified in a spectrofluorometer in cells stained with Nile red. B) Effects of oleic acid on reduced glutathione (GSH) content. Data are from Garcia et al. (2011).

concentration of oleic acid in the medium. Conversely, the total phospholipid is reduced. Therefore, in developing a prediction model, this information must be incorporated as it appears that the accumulating neutral fat crowds out the other intracellular membranes, such as the endoplasmic reticulum. The amount of total reduced glutathione also increased in the cells in response to the oleic acid treatment in a dose dependent manner. Future studies will determine the effect of the excess intracellular fat on the response to toxicants.

In conclusion, the primary goal has been to build reality into *in vitro* hepatotoxicity assays with an eye towards using the assays for actual risk assessment rather than merely as screening assays to guide future animal studies. The approaches outlined here for hepatotoxicity assays will hopefully provide some guidelines or a pathway that will allow scientists to see how these assays may someday become useful as replacements for animal-based liver toxicity testing.

6 Conclusion

Recent advances in the development of *in vitro* assays make the results more useful for risk assessment. This session highlighted approaches in which toxicologists are more effectively incorporating *in vitro* and alternative data to understand the toxicity of chemicals and drugs. Models have been and are being developed to address specific sensitive subpopulations. Omics technology has been used to help identify common effects of toxicants, such that the effects of unknown compounds can be compared to those of known toxicants in order to assist with risk assessment strategies. *In vitro* information is being incorporated into PBPK models that are routinely used in risk assessment of drugs and other chemicals. These advances in effectively incorporating *in vitro* data into risk assessment and linking the findings to observed effects *in vivo* is the result of a collaborative effort amongst federal agencies, non-profit science organizations, aca-

dem institutions, and industry. Future work in this area toward addressing the recommendations made by the NRC in *Toxicity Testing in the 21st Century* (NRC, 2007) will require continued collaboration and cooperation amongst these key groups.

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.
- Barter, Z. E., Bayliss, M. K., Beaune, P. H., et al. (2007). Scaling factors for the extrapolation of *in vivo* metabolic drug clearance from *in vitro* data: reaching a consensus on values of human microsomal protein and hepatocellularity per gram of liver. *Curr Drug Metab* 8, 33-45.
- Barter, Z. E., Chowdry, J. E., Harlow, J. R., et al. (2008). [Covariation of human microsomal protein per gram of liver with age: absence of influence of operator and sample storage may justify interlaboratory data pooling.](#) *Drug Metab Dispos* 36, 2405-2409.
- 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.
- Biomarkers Definitions Working Group (2001). Biomarkers and surrogate end points: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69, 89-95.
- Bouhifd, M., Hartung, T., Hogberg, H. T., et al. (2013). Review: Toxicometabolomics. *J Appl Toxicol* 33, 1365-1383.
- Carlile, D. J., Zomorodi, K., and Houston, J. B. (1997). Scaling factors to relate drug metabolic clearance in hepatic microsomes, isolated hepatocytes and the intact liver – studies with induced livers involving diazepam. *Drug Metab Dispos* 25, 903-911.
- Clewell, H. J., Reddy, M. B., Lave, T., and Andersen, M. E. (2008). Physiologically based pharmacokinetic modeling. In S. C. Gad (ed.), *Preclinical Development Handbook: ADME*



- and *Biopharmaceutical Properties* (1167-1227). John Wiley & Sons, Inc.
- Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science* 319, 906-907.
- Flynn, T. J. and Ferguson, M. S. (2008). Multiendpoint mechanistic profiling of hepatotoxicants in HepG2/C3A human hepatoma cells and novel statistical approaches for development of a prediction model for acute hepatotoxicity. *Toxicol In Vitro* 22, 1618-1631.
- Flynn, T. J. and Ferguson, M. S. (2010). An in vitro system for studying potential biological mechanisms of human sex differences in susceptibility to acute liver injury. *Toxicol Lett* 198, 232-236.
- Garcia, M. C., Amankwa-Sakyi, M., and Flynn, T. J. (2011). Cellular glutathione in fatty liver in vitro models. *Toxicol In Vitro* 25, 1501-1506.
- Hamburg, M. A. (2011). Advancing regulatory science. *Science* 331, 987.
- Hartung, T., Luechtefeld, T., Maertens, A., and Kleensang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX* 30, 3-18.
- Hartung, T. and McBride, M. (2011). Food for thought ... on mapping the human toxome. *ALTEX* 28, 83-93.
- Hartung, T., Blaauboer, B., and Bosgra, S. (2011). An expert consortium review of the EC-commissioned report "Alternative (non-animal) methods for cosmetics testing: Current status and future prospects – 2010". *ALTEX* 28, 183-209.
- Houston, J. B. (1994). Utility of in vitro drug metabolism data in predicting in vivo metabolic clearance. *Biochem Pharmacol* 47, 1469-1479.
- Kanungo, J., Lantz, S., and Paule, M. G. (2011). In vivo imaging and quantitative analysis of changes in axon length using transgenic zebrafish embryos. *Neurotoxicol Teratol* 33, 618-623.
- Kanungo, J. and Paule, M. G. (2011). Disruption of blastomeric F-actin: a potential early biomarker of developmental toxicity in zebrafish. *Mol Cell Biochem* 353, 283-290.
- Kedderis G. L. (1997). [Extrapolation of in vitro enzyme induction data to humans in vivo](#). *Chem Biol Interact* 107, 109-121.
- Kelly, J. H. (1994). Permanent human hepatocyte cell line and its use in a liver assist device (LAD). U.S. Patent No. 5290684.
- Kenyon, E. M., Pegram, R. A., Eklund, C. E., and Lipscomb, J. C. (2012). Comparison of in vivo derived and scaled in vitro metabolic rate constants for some volatile organic compounds (VOCs). *The Toxicologist, Supplement to Tox Sci* 126(S-1), 1892. SOT Annual Meeting, San Francisco, CA, USA.
- Lee, H. Y., Inselman, A. L., Kanungo, J., and Hansen, D. K. (2012). [Alternative models in developmental toxicology](#). *Syst Biol Reprod Med* 58, 10-22.
- Lipscomb, J. C., Fisher, J. W., Confer, P. D., and Byczkowski, J. Z. (1998). In vitro to in vivo extrapolation for trichloroethylene metabolism in humans. *Toxicol Appl Pharmacol* 152, 376-387.
- Lipscomb, J. C., Teuschler, L. K., Swartout, J. C., et al. (2003). Variance of microsomal protein and cytochrome P450 2E1 and 3A forms in adult human liver. *Toxicol Mechanisms Meth* 13, 45-51.
- Lipscomb, J. C. and Poet, T. S. (2008). In vitro measurements of metabolism for application in pharmacokinetic modeling. *Pharmacol Therap* 118, 82-103.
- Moore, T. J., Cohen, M. R., and Furberg, C. D. (2007). Serious Adverse Drug Events Reported to the Food and Drug Administration, 1998-2005. *Arch Intern Med* 167, 1752-1759.
- NRC (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: National Academy Press.
- Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London, UK: Methuen. Reprinted by Universities Federation for Animal Welfare, Potters Bar, UK, 1992.
- Sasso, A. F., Schlosser, P. M., Kedderis, G. L., et al. (2013). [Application of an updated physiologically based pharmacokinetic model for chloroform to evaluate CYP2E1-mediated renal toxicity in rats and mice](#). *Toxicol Sci* 131, 360-374.
- Shi, L., Campbell, G., Jones, W. D., et al. (2010). The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* 28, 827-838.
- Snawder, J. E. and Lipscomb, J. C. (2000). Interindividual variance of cytochrome P450 forms in human hepatic microsomes: Correlation of individual forms with xenobiotic metabolism and implications in risk assessment. *Reg Toxicol Pharmacol* 32, 200-209.
- Sobels, F. H. (1989). Models and assumptions underlying genetic risk assessment. *Mutat Res* 212, 77-89.
- Yang, X., Greenhaw, J., Shi, Q., et al. (2012). [Identification of urinary microRNA profiles in rats that may diagnose hepatotoxicity](#). *Toxicol Sci* 125, 335-344.

Correspondence to

Ambuja S. Bale, PhD
 U.S. Environmental Protection Agency – Office of
 Research and Development
 National Center for Environmental Assessment
 1200 Pennsylvania Ave., NW
 Washington, DC 20460, USA
 Phone: +1 703 347 8643
 Fax: +1 703 347 8693
 e-mail: ambujabale@gmail.com