Chemical Safety for Sustainability (CSS): Human \textit{in vivo} biomonitoring data for complementing results from \textit{in vitro} toxicology

A Commentary

Joachim D. Pleil\textsuperscript{1*}, Marc A. Williams\textsuperscript{2}, and Jon R. Sobus\textsuperscript{1}

\textsuperscript{1}Human Exposure and Atmospheric Sciences Division, National Exposure Research Laboratory

\textsuperscript{2}Environmental Public Health Division, National Health and Environmental Effects Laboratory, US Environmental Protection Agency

Research Triangle Park, NC 27711, USA

Abstract

The U.S. Environmental Protection Agency (EPA) has instituted the Chemical Safety for Sustainability (CSS) research program for assessing the health and environmental impact of manufactured chemicals. This is a broad program wherein one of the tasks is to develop high throughput screening (HTS) methods and follow-up confirmation for toxicity at realistic environmental exposure levels. The main tools under this task are \textit{in vitro} toxicity testing, \textit{in silico} molecular modeling, and \textit{in vivo} (systemic) measurements documentation. The \textit{in vivo} research component is intended to support and corroborate \textit{in vitro} chemical toxicity prioritization with observations of systemic perturbations and statistical parameters derived from intact (living) organisms. Based on EPA’s Biomonitoring Framework for human health research, such observations are intended to link environmental exposures to a cascade of biomarker chemicals to help identify and clarify adverse outcome pathways within the context of systems biology. This commentary discusses the issues regarding interpretation of \textit{in vitro} changes from HTS as an adverse result, an adaptive (non-adverse) response, or a random/irrelevant occurrence. A second goal is to inform \textit{in vitro} strategies as to relevant dosing (potency) levels at the cellular level that reflect realistic systemic exposures. Although we recognize the high value of \textit{in vivo} animal toxicity testing, herein we focus on observational (minimally-invasive) human biomonitoring methods and propose complementary \textit{in vivo} testing that could help guide the design of high-throughput analyses and the ultimate interpretation of their outcomes.

*corresponding author: pleil.joachim@epa.gov

\textit{key words:} exposure science; toxicological risk; computational toxicology; biomonitoring; systems biology, high-throughput screening
Introduction

Every year, about 1,000 new chemicals are introduced into U.S. commerce to join the more than 84,000 chemicals already listed in the Toxic Substances Control Act inventory (15USC2601). Of these, only a few hundred or so have undergone rigorous evaluation for public health safety. This commentary describes the evolution in modern toxicological research moving towards a major reliance on in vitro, high throughput screening (HTS), and proposes that there continues to be an important role for parallel research with in vivo biomarker research especially for discovering new endogenous compounds from a systemic/metabolic response that may not develop at the cellular or molecular level. This is particularly valuable for discerning protective repair functions at the organism level that could mitigate toxicity found at the molecular or cellular level.

We suggest that observational studies, representing the “intact organism” or “systems biology” approaches of environmental exposure science will be necessary to provide guidance for in vitro safety assessment of manufactured chemicals. Although we focus here on the niche of minimally invasive human biomonitoring methodologies, we also view the more traditional animal dosing experiments as valuable tools.

Chemical Safety for Sustainability (CSS)

In the U.S. Environmental Protection Agency’s (EPA) Framework for Chemical Safety for Sustainability (CSS) working document (EPA, 2011), it is acknowledged that:

“Although chemicals are essential to modern life, we lack innovative, systematic, effective, and efficient approaches and tools to inform decisions that reduce the environmental and societal impact of chemicals while increasing economic value.”

And it is further recognized that:

“...new transformative approaches are needed to improve the information used in (chemical safety) assessments.”

The U.S. EPA CSS program is very broad and includes numerous modeling strategies, molecular and cellular level bench research, invasive and observational (behavior, learning, memory, etc.) animal experiments, as well as observational human studies. Here, we are addressing only a subset of research under the CSS portfolio comprised of human studies that are directly or indirectly linked to environmental exposures.

Toxicity testing in the 21st century

The original initiative for modernizing toxicological and environmental research came in 2005 when EPA, the National Institute of Environmental Health Sciences (NIEHS), and the National Toxicology Program (NTP) requested an independent assessment and a long-range strategic plan.
of environmental toxicology testing from the National Research Council (NRC) (http://www.epa.gov/pesticides/science/nrc-toxtesting.html). The result was the seminal work entitled “Toxicity Testing in the 21st Century: A Vision and a Strategy” (NRC, 2007). Government Agencies, environmental organizations, research institutes, and the chemical industry have embraced this vision and have instituted a series of initiatives for toxicological high throughput testing including “Tox21” and “ToxCast” (http://epa.gov/ncct/Tox21/, http://www.epa.gov/ncct/toxcast/). The proposed new approaches are centered on in vitro (cell-line) testing which offers three specific advantages:

1. cost effectiveness, in that thousands of chemicals can be tested simultaneously,
2. reduced ethical considerations, in that no intact animals are used,
3. more relevant dosing strategies, in that more suitable (lower) levels are used for assessing effects at the cellular and molecular level.

The U.S. Government Agencies, EPA, NIEHS, NTP, National Institutes of Health (NIH), and other organizations are now embarking on a mission to augment the current in vivo (primarily animal, some human) approaches that are very labor and resource intensive with high-throughput methodology invoking in vitro toxicity testing and in silico molecular modeling (Andersen et al., 2010; Collins et al., 2008; Hubal, 2009; Raunio, 2011; Schmidt, 2009). The overall philosophy is to evolve the observational study of exposure to disease outcome (population based exposure assessment and disease epidemiology) to more predictive models based on target specific (cell line) studies and mechanism based (toxicity pathway) computer based studies. More specifically, there is a shift in to prioritize chemicals for further testing and to use in vitro assays to understand underlying cellular function (Judson et al., 2010). Beyond the explicit U.S. Government efforts, there are complementary programs for the development of toxicity testing directed or sponsored by organizations including the Hamner Institute for Health Sciences (Research Triangle Park, NC), the American Chemical Council (Washington, DC), the Evidence Based Toxicology Consortium at Johns Hopkins University (Baltimore MD), and European Commission Research Directorate (Brussels, Belgium).

As mentioned above, EPA and NIEHS have implemented collaborations entitled “Tox21” and “ToxCast” that combine computational chemistry, high throughput screening (HTS), and toxicogenomic assessment to prioritize chemicals for further evaluation and to discover modes of action (MOA) at the cellular level (Ankley et al. 2010, Boekelheide et al., 2011, Andersen et al. 2010) The ultimate goal is to build a rich database of chemical properties and in vitro test results that will continually improve the capabilities to predict chemical toxicity based on alternative methods (Dix et al., 2007; Rusyn et al., 2012). We propose, however, that these advantages should be tempered with an understanding that in vitro cell lines are not identical to intact organisms and that human systems biology and external exposures should be considered in making ultimate decisions about new chemicals entering the environment (Edwards and Preston, 2008; Pleil and Sheldon, 2011). Also, HTS results could be linked to “key events” or biological change that may initiate adverse response without fully understanding how they fit into a known adverse outcomes progression. Figure 1 is a representation of the “human relevance” to “throughput” continuum of toxicological testing. (The figure is adapted from various public presentations courtesy of Dr. Robert Kavlock of US EPA and is based on Tox21 concepts.
articulated by a consortium of Government organizations contributing to the National Center of Translational Sciences of the NIH).

Although philosophically it may be an improvement from traditional animal testing to assess mode of action with human cells and much lower environmental doses, measurements of the environment and of systemic in vivo responses continue to be crucial to properly design and interpret in vitro tests. Furthermore, developing, categorizing, and maintaining data from empirical measurements of “unremarkably exposed” humans will allow us to deduce when subpopulations (or individuals) are out of the norm, and thus may trigger further evaluation using in vitro methodologies (Edwards and Preston, 2008; Pleil, 2012). Finally, we reiterate that the human in vivo information will allow us to include systemic repair functions into risk assessments that are otherwise lost if we rely on the cellular or molecular level damage assessments alone.

Current limitations of HTS

The emphasis on high-throughput and computational research, is not without scientific risk. Although in vitro challenges to cell lines with thousands of different chemicals are an elegant and highly efficient approach, the results are not completely conclusive. To design and evaluate the proper experiments to determine if any sets of in vitro responses are truly realistic and probative, there are (at least) four basic questions that need to be answered:

1. How does in vitro potency relate to in vivo potency?
2. How does an in vitro response relate to an adverse outcome in vivo?
3. How can in vitro testing provide within- and between-subject variance components for the human population?
4. How can in vitro testing provide critical life-stage information for adverse effects?

These are all important considerations for designating a specific chemical as “safe” or acceptable within reasonable bounds. They address the complexity of the human systems biology that may not be reflected in cell lines or computer models, especially with regard to repair function, differential metabolism (toxicity activation or detoxification), individual host factors, and environmental exposure parameters. There may be other more complex questions that could be explored in the future such as the synergistic effects of complex chemical mixtures and the cumulative risks from repeated exposures, but these are beyond the scope of this article.

Potency consideration: The first issue regarding observations of in vitro potency and the prediction of in vivo potency revolves around dosing at the cellular level; that is, how can the in vitro dose be administered to mimic a systemic dose encountered by an intact organism (human) in the environment? Aylward and Hays (2011) discuss this in detail and conclude that human biomonitoring data should be collected and that exposure factors that affect absorption, distribution, metabolism, and elimination (ADME) in the organism should be considered to establish the potency relationships. Additionally, we need to consider the impact of chemicals on protective immune responses and increased susceptibility to allergic inflammatory and
perhaps infectious disease (Williams, 2011; Williams et al., 2007; Williams et al., 2008). We also should include emerging concepts on epigenetic factors that modify DNA externally and affect expression of proteins, cellular transcription, and cancer outcomes (Feil and Fraga, 2011; Ross and Davis, 2011; Sandoval and Esteller, 2012) as well as endocrine disruption, inflammation, and autoimmune responses (Hartnett and Egan, 2012; Pollard et al., 2010; Schug et al., 2011). All of these considerations will affect the dosing strategies at the molecular and cellular levels.

In recognition of this need to collect and interpret human exposure data, EPA has implemented a new program named “ExpoCast” as the exposure science counterpart to ToxCast (http://www.epa.gov/ncct/expocast/) (Cohen Hubal et al., 2010). ExpoCast is intended to provide the overarching framework to characterize “…biologically relevant exposures (to) link human exposure data for chemical prioritization and toxicity testing.” (Kavlock and Dix, 2010).

The tasks of linking environmental exposures to internal dose, and eventually to bio-indicators and effects have been discussed in two EPA biomonitoring framework articles (Sobus et al., 2011; Tan et al., 2012). A number of case study articles have also appeared in the literature that serve to illustrate empirically how internal biomarker levels of environmental chemicals are related to external exposures. Over the years, researchers at EPA and other institutions have studied ADME and classical pharmacokinetics (PK), exposure reconstruction, and physiologically based pharmacokinetic (PBPK) models of specific chemicals and mixtures (Buschmann, 2006; Egeghy et al., 2011; Furtaw, 2001; Goldsmith et al., 2010; Kim et al., 2007; Lipscomb et al., 2012; Lorber and Egeghy, 2011; Sahmel et al., 2010; Tan et al., 2012).

Our own focused studies at the EPA National Exposure Research Laboratory (NERL) and the National Health and Environmental Effects Laboratory (NHEERL) have included observational and pharmaco-kinetic work with methyl tertiary butyl ether, (Buckley et al., 2001; Kim et al., 2007; Lindstrom and Pleil, 1996a; Pleil et al., 2007; Prah et al., 2004), documentation of polycyclic aromatic hydrocarbons (PAHs) exposures in blood and plasma (Pleil et al., 2010), military jet fuel exposures in human blood and exhaled breath (Liu and Pleil, 1999, 2001; Pleil et al., 2000), volatile trihalomethane and chlorinated compounds in breath from inhalation and dermal exposures (Lindstrom and Pleil, 1996b; Lindstrom et al., 1997; Pleil and Lindstrom, 1997), and trichloroethylene exposures in blood and breath (Pleil et al., 1998). We have also been heavily involved in larger-scale human biomonitoring studies of specific chemical classes with particular focus on pesticides (Egeghy et al., 2005; Morgan et al., 2007; Morgan et al., 2005; Morgan et al., 2008; Naheer et al., 2010; Thomas et al., 2010).

The research articles listed above are certainly not exhaustive, but illustrate how different approaches might be implemented and how the resulting data could be interpreted. Such focused and targeted exposure case studies, regardless of scale, have demonstrated how specific environmental pollutants impact the public and have helped establish benchmarks for environmental epidemiology. EPA will develop estimates for potency values based on such human observational studies and publish a “Biomarkers Knowledge Base” as part of the CSS research program output strategy. For future work, specific targeted biomarker data will continue to be an important component for assessing broader groups of compounds with similar modes of action and chemical structures.
Adverse outcome: The premise of HTS in vitro testing is to challenge different human cell lines with a range of concentrations of manufactured chemicals and observe chemicals/responses that are related to specific key events that can be related to toxicity pathways at the cellular and subcellular level, and adverse outcome pathways (AOP) (Ankley et al. 2010). The AOP concept has been developed as a descriptive tool linking a specific harmful effect to a direct initiating event caused at the molecular level with the ultimate goal of informing risk assessment and identifying susceptible populations (Perkins et al., 2011; Stephens et al., 2012; Watanabe et al., 2011).

Although the literature tends to focus on ecotoxicological pathways, some examples of AOP in human cell lines are activation of aryl hydrocarbon receptor, activation of estrogen receptor, and tumor suppressor protein (p53) mutation that relate to immunosuppression, reproductive outcomes, and cancer, respectively (Ankley et al., 2010; Bhattacharya et al., 2011). There are undoubtedly hundreds of other toxicity pathways that could be seen in vivo as well, however, the question remains, at what point in the progression of all possible responses in the Petri dish or well-plate array can a change be considered adverse at the organism level?

This issue has engendered a branch of molecular methods testing wherein researchers are constructing a “Taxonomy of Adverse Effects” that can be used to classify and organize changes at the cellular and subcellular levels (Boekelheide and Campion, 2010). In the long run, the taxonomy approach will likely prove extremely useful in diagnosing in vitro changes and categorize chemicals on some form of sliding scale of in vivo adversity. In the near future, however, observational human biomarker measurements and limited invasive (traditional) animal testing must support this role. We have considered this issue from the perspective of top-down or discovery style human biomarker (exposome) analysis (Hubbard et al., 2009; Pleil et al., 2010; Sawyer et al., 2008) and from the perspective of “forward mapping” of observable parameters to putative adverse endpoints (Pleil and Sheldon, 2011). Biomarker discovery is based on a stratified un-targeted approach wherein a broad spectrum of chemicals are measured in a particular medium (e.g. blood, breath, urine) and interpreted within the context of an observed range of measured environmental exposures. We understand that there are many possibilities for purely random associations, and that a statistical correlation between biomarkers and intervention group may not necessarily be adverse. Furthermore, a response could be very consistent and repeatable, yet still be irrelevant with respect to health. As such, we need to consistently (and statistically) remove the weaker associations as is done in gene-wide association studies (GWAS) and is now being applied to epigenome-wide and environment-wide association studies (EWAS) as well (Murcray et al., 2011; Rakyan et al., 2011; Thomas, 2010). The objective of discovery is to eventually focus down to a small group of probative compounds within the exposome that represent the biologically relevant bioindicators and can be unambiguously related to an adverse pathway (Birnbaum, 2010; Lioy and Rappaport, 2011; Patel et al., 2010). The ultimate goal is to link such chemically defined AOPs and their observations in vitro to causality of in vivo diseases. In fact, such “top-down” in vivo approaches are considered to be the path towards eventually discovering the causes of about 80% of all human disease: those attributable to environmental causes (Rappaport, 2012; Rappaport and Smith, 2010).
Variance components: One of the important factors in estimating risk is to understand the variance components within- and between-human subjects; only then, can one begin to understand how to mitigate adverse effects from chemical factors (Lin et al., 2005; Peretz et al., 2002; Sobus et al., 2010). This is a domain that cannot be easily addressed with in vitro and in silico methods and requires multiple repeat measures of many different individuals. The reasoning is that temporal changes and individual host factors contribute to variance, as do broader changes in the environmental conditions dictated by ecological (meso-scale) conditions (Pleil, 2009; Pleil et al., 2012). Although in vitro approaches can be used to make static dose-response observations at the molecular and cellular level, there are no simply accessible systems biology factors that can create organism level variances in the Petri dish (Blaauboer, 2008; Edwards and Preston, 2008). This creates difficulty for assigning probative responses beyond the confines of the individual cell. Certainly in vitro systems can be established for different types and genetically distinct cell lines, but the effort of mimicking human genetic diversity would be overwhelming. Knowing the “normal” range of a cellular or organism responses is critical when assigning the level of adversity or whether the response is “unremarkable” and falls into the adaptive category (Pleil, 2012).

There are three basic approaches to discovering biomarker variance components in systemic responses to environmental triggers: calculation based studies of environmental measurements correlated with human meta-data (height, weight, age, gender, etc.), observational longitudinal (repeated measurements over time) studies, and observational cross-sectional/panel (snapshot in time) studies. All have been regularly implemented at the EPA Office of Research and Development Laboratories (Pleil et al., 2000; Samet et al., 2007; Williams et al., 2009). In our recent experience, we have found that comparisons between observational studies that incorporate detailed human meta-data and measurement data and those that are observational studies of the snapshot type that the range in variance of biomarker measurements is 10-times less in the former (Pleil, 2009; Sobus et al., 2010). Furthermore, in many biomarker studies of environmental level exposures, it has been found that the one of the most important host-factors for explaining between-subject variance is gender (Brown et al., 1998; Hubbard et al., 2009; Lin et al., 2006; Pollard, 2012). Certainly there are additional human host-factors that affect biomarker response including obesity, age, health state, microbiome (gut and pulmonary bacteria), nutrition, stress-level, just to name a few, that are commonly reported in the literature. These host-factor parameters contribute to orders of magnitude variability and cannot be addressed with in vitro experiments. As such, the parallel in vivo study of the interaction of environmental levels of chemicals with human cohorts is important to distinguish a real response from an adaptive or random effect.

Life-stage: A main tenet of public health safety is the protection of vulnerable sub-populations from environmental toxicants (Landrigan and Goldman, 2011; Woodruff et al., 2011). Two such groups are children (both pre- and post-natal) and the aged population; children are considered especially susceptible as they are experiencing the highest growth/developmental rates and have not fully established their immune systems whereas the seniors are subject to decreasing efficiency of repair function and generally declining health state (Dietert, 2009; Ginsberg et al., 2004; Jafri, 2011; Sacks et al., 2011; Tulve et al., 2008; Williams et al., 2000). Although life stage is recognized in the HTS community, the proposed resolution relies on statistical estimates of population variability coupled with in silico models to achieve sufficiently conservative...
exposure limits based on calculations of a biological pathway altering dose (BPAD) (Judson et al., 2011).

Cross-sectional environmental exposure studies, both completed and ongoing, have shown repeatedly that “...children are not just small people...” and thus cannot be treated the same within epidemiological and toxicological structures (NRC, 1993). Studies of the senior population have found that predisposition to adverse external effects is amplified by health state (asthma, mobility, nutrition, cardio-pulmonary function), as well as changes in metabolic clearance of exogenous compounds (Alexeeff et al., 2008; Wang et al., 2010). In fairness, there may also be protective/adaptive parameters in the senior population; for example, point mutations are less likely to proliferate into tumors due to slowed cellular reproductive rates and less efficient metabolism, and germ-line mutations are unlikely to be passed forward (Campisi, 2003; Mays Hoopes, 2010; Sartorius and Nieschlag, 2010). As such, observational in vivo studies of sub-populations are important to address the differential toxicity experienced due to age-based parameters.

**Concluding remarks**

The CSS program is designed, in part, to address the reality that chemical toxicity evaluations will require faster and more efficient methods to keep up with the explosion of new manufactured compounds and formulations in industrial chemicals, consumer products, fuels, pharmaceuticals, and other high-volume uses of novel chemicals. These new chemicals may affect human health and welfare through environmental distribution or through contact from intended use. The implementation of broad scale in vitro toxicity testing for addressing the safety of new manufactured chemicals is an important step forward for the protection of human health in an environment constantly increasing in chemical complexity. The caution that we bring, however, is that in vivo systems biology and associated observational biomonitoring at the intact organism level, remain as crucial components for proper design and interpretation of in vitro and in silico results. Realistic biomarker data are necessary to determine in vitro dose parameters, to link an in vitro response to a true response in the intact organism, to assess the statistics of real-life variance in the context of in vitro monocultures, and finally to determine how real-world systemic vulnerabilities and protective repair functions can be related to the microcosm of observations in cellular and molecular level adverse outcome pathways. The temptation for “cheaper/faster” toxicological screening of chemicals should be tempered with the reality of the complexity of the gene x environment interaction in human systems biology.

**Acknowledgements**

The authors thank Stephen Edwards and Ross Highsmith from US EPA for many insightful discussions regarding biomarkers and their relationship to toxicity testing. We also thank Cecilia Tan, Elaine Cohen-Hubal, Linda Sheldon, Robert Kavlock, Nicolle Tulve, Myriam Medina-Vera, Elin Ulrich, Michael Madden, and Andrew Ghio of US EPA, Stephen Rappaport of University of California, Berkeley, Terence Risby of Johns Hopkins University, and Rebecca Fry and Leena Nylander-French of University of North Carolina, Chapel Hill for their invaluable
contributions to the overall concepts of integrating biomonitoring into environmental exposure science.

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. This research has been subjected to (EPA) Agency review and approved for publication.

Figure 1. Representation of the continuum of toxicological assays ranging from low (LTS) to medium (MTS), to high (HTS), to ultra high throughput (uHTS) throughput screening. The trade-offs between direct relevance to human systems and efficiency of processing samples are shown in the lower pennants; at the left are *in vivo* experiments with humans and animals, progressing through cellular level gene expression, in vitro chemical assays in well-plate arrays, and finally to uHTS at the molecular (protein, enzyme, peptide) level. This figure is adapted from presentations courtesy of Dr. Robert Kavlock, US EPA; it is based on concepts developed by the National Center of Translational Sciences, National Institutes of Health: (http://www.ncats.nih.gov/research/reengineering/tox21/tox21.html).
References

15USC2601, T., Title 15: Commerce and Trade.
Collins, F.S., Gray, G.M., Bucher, J.R., 2008. Toxicology. Transforming environmental health
protection. Science 319, 906-907.

Landrigan, P.J., Goldman, L.R., 2011. Children's vulnerability to toxic chemicals: a challenge and opportunity to strengthen health and environmental policy. Health affairs 30, 842-850.


Lorber, M., Egeghy, P.P., 2011. Simple intake and pharmacokinetic modeling to characterize exposure of Americans to perfluorooctanoic acid, PFOA. Environmental science & technology 45, 8006-8014.


Williams, R., Rea, A., Vette, A., Croghan, C., Whitaker, D., Stevens, C., McDow, S., Fortmann,


Woodruff, T.J., Burke, T.A., Zeise, L., 2011. The need for better public health decisions on chemicals released into our environment. Health affairs 30, 957-967.