Chemical Testing in the 21st Century: A Primer

We face major challenges in evaluating the health and environmental effects of chemicals. Inadequate policies have created major gaps in the available data on the safety of the ever-growing number of chemicals in commerce. Our ways of testing chemicals and assessing their risks haven't kept up with advances in science in recent decades, which point to wholly new factors that need to be considered, such as the effects of low-dose exposures, the importance of the timing of exposure, and the extent of variability in the human population. And current testing methods rarely tell us how chemicals act and can miss effects not easily detected in traditional studies conducted on whole animals.

New approaches are being developed to meet these challenges that do not rely solely on costly and time-consuming animal-based testing methods. These approaches allow testing of more chemicals at lower cost and can also help characterize the underlying mechanisms by which chemicals interact with our biology. But they face their own sets of limitations and challenges that must be met if they are to contribute to the future of toxicity testing. Because EDF believes that active engagement of organizations and researchers dedicated to improving public health is critical to the development and use of the emerging methods, we have developed this primer to provide a general introduction to the new testing approaches.

Table of Contents

Section 1: Why do we need new approaches for evaluating chemical risks?

Fetal development is a particularly susceptible "window of exposure" for harmful chemicals.

The Environmental Protection Agency (EPA) faces daunting challenges when it comes to evaluating the health and environmental effects of chemicals. The problem is three-fold. First, there is a tremendous gap between the growing number of chemicals in commerce and the available data on their health and environmental effects. This is partly due to the shortcomings of the Toxic Substances Control Act (TSCA) [for further explanation, see EDF Senior Scientist Richard Denison's discussion of key elements for TSCA reform]. Second, advances in scientific research in recent decades have identified new factors to consider in evaluating chemical risks, such as: the effects of low-dose exposures and differential effects based on the timing of exposure; epigenetic and genetic variability in the population; co-exposures to multiple chemicals as well as other factors such as nutrition and stress; and an expanding diversity of potential adverse endpoints. Third, current methods tell us very little about *how* chemicals act and can miss subtle but nonetheless that aren't easily detected by looking for obvious effects in whole animals.

These challenges cannot be met by relying *solely* on traditional animal-based testing methods—as has been done for decades. Animal testing is time-consuming and costly—both in terms of monetary costs and laboratory animal lives. We cannot fill the data gap and answer the critical questions necessary to evaluate chemical safety in any realistic time frame by traditional testing on animals alone.

In an effort to confront these challenges, EPA (in partnership with other agencies) is investing in the development of new approaches and methods for testing and assessing chemicals. This investment reflects the recognition that traditional, animal-based testing methods cannot by themselves efficiently overcome the policy and scientific challenges facing chemical risk assessment today. The future of toxicity testing must include greater use of higher-throughput, testing methods (using both non-animal tests and tests in novel animal models such as zebrafish), as well as computer modeling. Such methods hold promise not only to allow testing of more chemicals at lower cost, but also to elucidate the underlying mechanisms by which chemicals interact with our biology and, ultimately, pave the way for more accurate and efficient prediction of chemical risks.

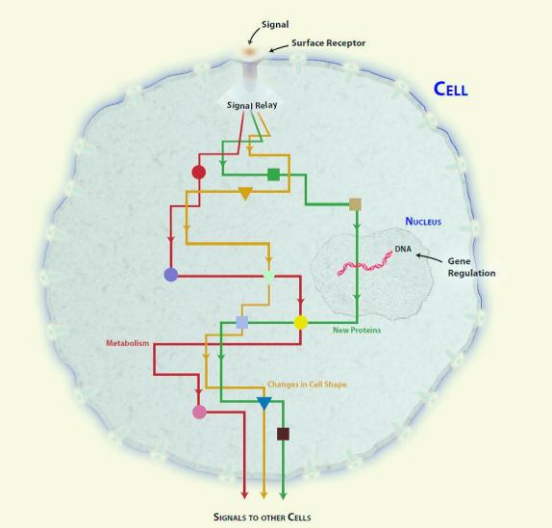
But modernizing the approach to testing chemicals is itself a daunting task. If this future vision for toxicity testing is to fulfill the promise of providing more *and* better information about more chemicals to strengthen our understanding of their potential effects on our health, then we need the active engagement of organizations and researchers dedicated to improving public health in the development and use of the emerging methods. EDF developed this primer to serve as a first step towards fostering this engagement by providing a general introduction to the new testing approaches and the programs underway at the EPA.

To learn the basics behind one of the most developed of these new testing approaches, proceed to **Section 2: Introduction to High-Throughput *in Vitro* Testing**

Section 2: Introduction to High-Throughput *in Vitro* Testing

EPA is developing and refining a wide variety of new methods to better assess and predict chemical hazard and exposure. So-called high-throughput (HT) *in vitro* testing is among the most developed of these methods. HT *in vitro* tests are conducted using cells or cell components rather than intact laboratory animals. Hence the use of the term *in vitro* (Latin for "in glass"), as opposed to *in vivo* (Latin for "within the living") used to describe tests conducted in animals. Unlike traditional, animal-based testing, HT methods examine chemicals' effects at the molecular and cellular levels in order to understand and predict adverse outcomes in the whole organism (e.g., a human).

In essence, *in vitro* tests assess whether and to what extent chemicals perturb normal cellular functions or "biological pathways." One example of a biological pathway is the sequence of molecular steps in a cell that leads to metabolic breakdown of a substance. Another example is the series of steps that lead to the expression ("turning on") of a particular gene in our DNA. Biological pathways are essential to life. They are responsible for proper execution of all our bodily functions from digesting food and regulating our reproductive cycles to ensuring normal brain development. But when something causes the normal activity of a biological pathway to go awry, we are in danger of being on the receiving end of a negative health outcome, for example, diabetes or cancer. For more background on biological pathways, visit the NIH Human Genome Research page.



External signals can initiate an internal chain of events in cells, and such biological pathways may trigger the production of metabolites, the regulation of genes, or other cellular changes. Photo by National Human Genome Research Institute

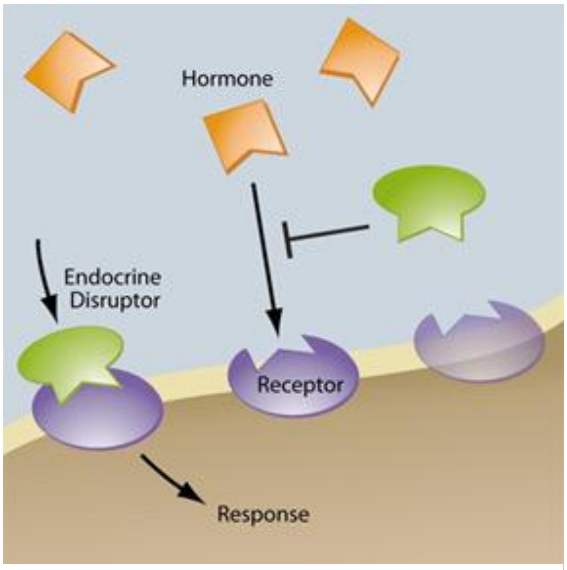
There are two basic types of *in vitro* tests, or "assays": cellular and acellular.

In a *cellular* assay, a culture (population) of cells is exposed to a chemical. During or following this exposure, disturbance (e.g., activation or suppression) of a particular biological pathway—or pathways—of interest is monitored. For example, a particular assay conducted using cultured human cells may seek to detect whether a chemical binds to the cells' estrogen receptors. If such an interaction is detected, the chemical is flagged as a potential endocrine disruptor, that is, a chemical with potential to interfere with the normal function of our endocrine system, on which we rely for normal reproductive development among many other things.

Acellular assays look for similar interactions between chemicals and biological pathways, but using cell components rather than intact cells. They examine activity at an even smaller level of biological organization by using molecules extracted from cells, such as enzymes or DNA. For example, some acellular assays examine whether a chemical interacts with cytochrome P450, a critically important enzyme involved in a number of biological activities including the

breakdown or metabolism of foreign compounds that enter our bodies, such as drugs and other chemicals. In these assays, purified cytochrome P450 is mixed with a chemical and the extent to which the normal activity of cytochrome P450 is inhibited is measured. Significant inhibition may indicate that the chemical can interfere with a key mechanism by which our bodies normally detoxify foreign substances.

In vitro assays detect *early indicators* ("initiating events") of what may ultimately lead to an adverse health effect. They provide information about the "mechanism of action" by which a chemical may alter our biology (*e.g.*, by binding to the estrogen receptor). These changes can't be easily detected or measured using traditional toxicity testing, which is more focused on determining whether a particular dose of a chemical results in an observable change in the health or normal functioning of the *whole animal* (e.g., loss of fertility, appearance of a tumor).



High-throughput (HT) *in vitro* assays can be used to identify chemicals that perturb normal biological activities. For example, assays can identify potential endocrine disruptors by detecting whether a chemical (green) interferes with the normal binding of a hormone (orange) to its receptor (purple). Photo by NIH

In sum, the aim of using *in vitro* assays is to predict adverse health outcomes by identifying those initiating or preceding events that negatively affect—or perturb—biological pathways in the cell in ways that lead to disease or debilitating conditions (e.g., asthma).

In vitro assays are not entirely new. For example, the classic Ames test, developed in the 1970s, uses bacteria to determine whether a chemical causes DNA mutations, which we know to be *one*—but certainly not the only—good predictor of whether a chemical can cause cancer in humans.

Since the development of the Ames test, perturbations in hundreds of important biological pathways have been implicated in the development of disease or a health condition.. And scientists have sought to design *in vitro* tests that examine whether and to what extent chemicals perturb each of these pathways. As a result, hundreds of such assays have been developed.

In addition to the large number of assays now available, what is also novel about *in vitro* testing today is that assays can be conducted in a "high-throughput" manner, meaning they can be run:

- quickly and inexpensively in very small volumes of solution;
- simultaneously on thousands of chemicals, or mixtures of chemicals, and;
- at multiple doses.

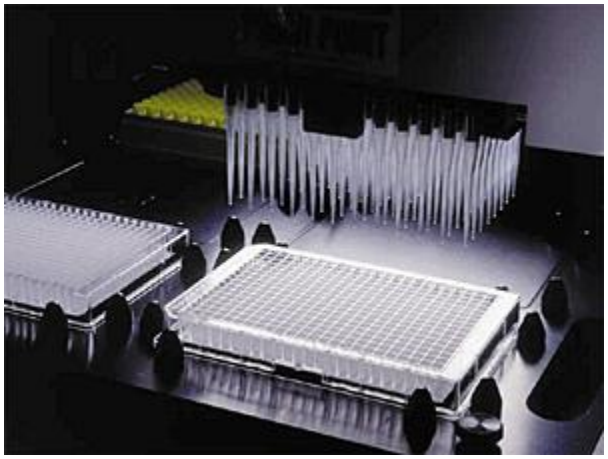
EPA's Toxicity Forecaster (ToxCast) program is focused specifically on advancing this type of high-throughput, *in vitro* testing approach. To learn more, proceed to **Section 3: EPA's Toxicity Forecaster (ToxCast) Program.**

Section 3: EPA's Toxicity Forecaster (ToxCast) Program

EPA's efforts to advance new approaches for better understanding and predicting chemical risk are coordinated through an umbrella research initiative called CompTox, housed at the EPA's National Center for Computational Toxicology (NCCT) . CompTox houses several individual research programs focused on exposure estimation and hazard prediction -- including the development of computational models that attempt to mimic the functions of whole organs and tissues.

Arguable the most developed of the CompTox programs is the Toxicity Forecaster or ToxCast program. ToxCast uses high-throughput (HT) *in vitro* testing to assess potential hazards of chemicals. For more information on HT *in vitro* testing, visit Section 2 of the primer.

Impressively, the HT technology being employed in the ToxCast program allows for thousands of chemicals to be quickly tested—at multiple doses—for effects on hundreds of biological pathways. ToxCast assays are not conducted at EPA, but rather at EPA-contracted biotech companies (for a list of these contractors, visit this EPA CompTox webpage). The HT assays are each run in individual wells on multi-well plates requiring very small volumes of material—literally drops of liquid (see image below). The whole operation is automated and uses robots to carry out the tedious, repetitive work (see robots at work here). These features—many assays run simultaneously on many chemicals at many doses, all automated—are what warrant the use of the term "high-throughput."



HT *in vitro* testing systems utilize multi-well plates, such as the one pictured above, to quickly assess the toxicity of thousands of chemicals. Photo by Nature Biotechnology.

The data generated from ToxCast are being used by EPA to build predictive models of chemical toxicity related to specific adverse health outcomes, including impairments of early development, male and female reproductive function, and vascular development (Read this EPA factsheet to learn more about how ToxCast data are being used to build predictive models).

The ToxCast program divides its work into two phases. EPA refers to Phase I, now completed, as the "Proof of Concept" phase. In this phase, EPA selected nearly 300 chemicals, mostly

pesticides that had been extensively tested using traditional methods. It ran each of them through approximately 600 HT *in vitro* assays.

Because the Phase I chemicals are data-rich, EPA is comparing the results of data generated from the ToxCast assays with data from traditional, *in vivo* tests. This comparison effort is a major aspect of EPA's work to "validate" the ToxCast assays (i.e., determine how well the ToxCast assays predict outcomes seen *in vivo*).

According to a 2010 presentation by CompTox researchers, models developed using ToxCast Phase I data to screen for chronic, developmental and reproductive toxicity endpoints had few false positives (i.e., chemicals erroneously flagged as hazardous that are not), but had many false negatives (i.e., chemicals flagged erroneously as benign that are not). A 2011 CompTox publication that evaluated the accuracy of a rat reproductive toxicity model built using ToxCast assays indicated that the model was able to distinguish chemicals that were and were not reproductive toxicants with an accuracy of 80%. The predictive power of ToxCast toxicity models will vary from model to model, of course, depending on factors such as the quantity and quality of the underlying data. A key need will be to continuously assess and improve the predictability of such models.

Phase II of ToxCast was launched in 2010 and is ongoing. In this phase, to further explore the potential utility of ToxCast, EPA is evaluating an additional 767 chemicals drawn "from a broad range of sources including industrial and consumer products, food additives, touted "green" products, nanomaterials and drugs that never made it to the market."

EPA is making data generated in both Phase I and II of the ToxCast program publicly available through this online database.

While the goals of the EPA efforts are laudable—filling massive data gaps and improving our ability to accurately predict adverse effects of chemicals, more quickly and at lower cost—there are important limitations and uncertainties that need to be recognized and addressed. Other sections of this primer provide an overview of both the benefits (see here) and the challenges (see here) to advancing this work.

For additional information from EPA on the ToxCast program visit:

- ToxCast Webpage
- Factsheet: Overview of ToxCast Program
- Factsheet: Using ToxCast to Predict Chemicals Potential for Developmental, Reproductive and Vascular Development Toxicity

For a description of related research initiatives, proceed to **Section 4: EPA's CompTox Programs**.

Section 4: EPA's CompTox Programs

In 2005, the Environmental Protection Agency (EPA) established the National Center for Computational Toxicology (NCCT) within the agency's main research arm, the Office of Research and Development (organizational chart available here). NCCT coordinates the agency's computational toxicology research program, nicknamed CompTox. This page briefly reviews the various research efforts within the CompTox program. To watch an informational video about EPA's CompTox programs, click here.

CompTox encompasses several discrete research programs, each focused on the development and application of innovative approaches and tools to improve chemical hazard, exposure and risk assessment. ToxCast is one such program. You can read more about the ToxCast program in Section 3 of this primer.

EPA created CompTox in large part to address the significant lack of health and environmental data on the thousands of chemicals for which it is responsible to ensure are safe for human health and the environment. EPA states that the purpose of the CompTox program is to "conduct innovative research that integrates advances in molecular biology, chemistry, and computer science to more effectively and efficiently rank chemicals based on risk."

The individual research programs within CompTox are generating massive amounts of data on thousands of chemicals. Sophisticated computer systems and programs are required to integrate and analyze these large data sets. This is why the term "computational toxicology" is used to describe these newer types of chemical assessment tools and approaches.

For more information on CompTox visit:

- CompTox Webpage
- Factsheet: CompTox Research Program

The following subsections describe the individual research programs within CompTox.

Toxicity Forecaster (ToxCast)

The ToxCast program is focused on the development and use of high-throughput (HT) *in vitro* testing to identify and characterize chemical toxicity. HT testing and the ToxCast program are described in detail in Section 2 and Section 3 of this primer.

For additional information from EPA on the ToxCast program visit:

- EPA ToxCast Webpage
- Factsheet: Overview of ToxCast Program
- Factsheet: Using ToxCast to Predict Chemicals Potential for Developmental, Reproductive and Vascular Development Toxicity

Tox21 Program

The Tox21 program is an inter-agency program of which EPA is apart and to which ToxCast data are contributed. Tox21, established in 2008, is a collaborative effort that leverages federal resources and expertise from EPA, National Institutes of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), and the Food and Drug Administration (FDA) to conduct high-throughput *in vitro* testing for screening and prioritizing thousands of chemicals for potential toxicity.

The technology and intent of the Tox21 program are very similar to that of ToxCast. The primary differences are:

1. There are far fewer assays being used currently in the Tox21 program. Tox21 contains approximately 50 of the 600 assays used in ToxCast.
2. Tox21 is testing many more chemicals than ToxCast. The Tox21 program includes 10,000 chemicals, whereas ToxCast includes about 2,000 chemicals.
3. Tox21 assays are conducted in house at the NIH National Chemical Genomics Center, whereas ToxCast assays are conducted by contracted biotech companies outside of federal agencies or research labs.

For more information on the Tox21 program, please visit:

- EPA Tox21 webpage
- NIH Tox21 webpage

Exposure Forecaster (ExpoCast)

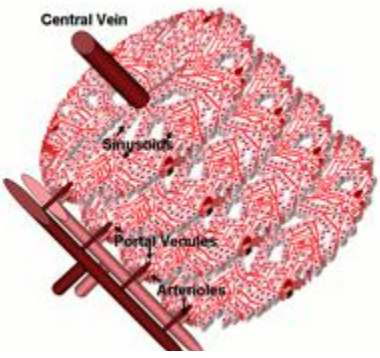
The ExpoCast program is the exposure counterpart to ToxCast. ExpoCast is focused on developing and validating computer models to estimate human and environmental exposures to thousands of chemicals. To build the models, ExpoCast is using exposure information from the Centers for Disease Control's NHANES biomonitoring program and EPA's Toxics Release Inventory and Chemical Data Reporting system (formerly Inventory Update Rule), as well as available environmental and biological fate and transport data on chemicals. Notably, information on chemical exposures from consumer products is currently lacking in ExpoCast, though efforts are being made to address this gap.

For more information on ExpoCast visit:

- ExpoCast Webpage

Virtual Tissues (v-Tissues)

The virtual tissues program seeks to develop computer simulations of complex biological structures and cell networks to predict how chemicals can affect and perturb these systems. The term *in silico* is used to describe these types of computer-based testing approaches. EPA is focused currently on developing



The virtual liver (v-Liver) project, part of EPA's CompTox virtual tissues program, is one of several computer-generated models being developed to predict and evaluate chemical toxicity. Photo by EPA

computer models of the liver and developing embryo, called v-liver and v-embryo, respectively. The v-embryo program is further divided into four focus areas: eye development, limb development, vascular system development, and embryonic stem cell development.

Developing computer simulations of these complex biological systems involves using advanced computer programming to overlay and integrate a wealth of scientific knowledge on biological pathways and processes underlying liver function and embryonic development. The predictive capacity of *in silico* testing systems like v-liver and v-embryo hinges on the extent and quality of data on which they are built. EPA is using both existing information from the scientific literature and new data generated from research programs like ToxCast to build these models.

For additional information from EPA on Virtual Tissues visit:

- Fact Sheet: Virtual Tissues
- v-Embryo Webpage
- v-Embryo Research Projects
- v-Embryo Experimental Approach
- v-Liver Webpage

CompTox Databases

EPA is creating publicly accessible online databases for both new chemical information emerging from the CompTox program and existing information compiled from other sources. These databases include:

- Aggregated Computational Toxicology Online Resource (ACToR) – ACToR is a repository of all publicly available chemical toxicity data. It also houses information related to chemical risk to human health and the environment. This database contains information on over 500,000 chemicals from more than 650 public sources.

- Distributed Structure-Searchable Toxicity Database Network (DSSTox) – DSSTox is an online public database of standardized chemical structures that are linked to "high-interest" toxicity data sets.

- Toxicity Reference Database (ToxRefDB) – ToxRefDB contains data on hundreds of chemicals, mostly pesticides, compiled from thousands of traditional *in vivo* animal toxicity studies. To date, the studies in ToxRefDB are primarily regulatory "guideline" studies that follow standardized protocols.

- Toxicity Forecaster Database (ToxCastDB) – The ToxCast database is being built to house data on thousands of chemicals being run through the hundreds of HT screening assays in ToxCast. Only a subset of the ToxCast data is currently available.

- Exposure-Based Chemical Prioritization Database (ExpoCastDB) – The ExpoCastDB contains measurements of the levels of chemicals in environmental and biological media

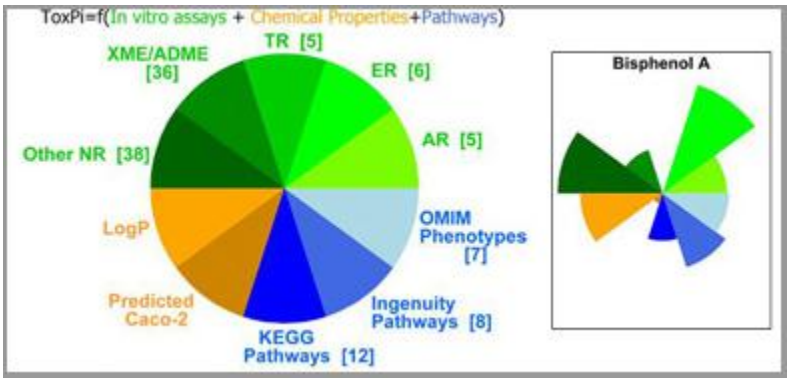
collected from homes and childcare centers. Data currently available include, to varying degrees, the amounts of these chemicals found in or on food, drinking water, air, dust, indoor surfaces and urine. The public version of ExpoCastDB currently includes limited exposure data for 99 chemicals, mostly pesticides.

For more information on these databases see:

- Factsheet: Chemical Toxicity Database
- Paper: "Aggregating Data for Computational Toxicology Applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System," February 2012.

Toxicological Prioritization Index (ToxPi)

ToxPi is a computer software tool that incorporates different sets of toxicity data on a chemical and produces a visual output of those data—a so-called "Pi." ToxPi seeks to convey the "comprehensive toxicity" of a chemical in a manner that can be quickly ascertained through visual representation and be used for prioritizing chemicals. The ToxPi tool is flexible, allowing the user to choose which datasets to include in the construction of a Pi.



ToxPi incorporates multiple data sets (left figure) to produce a visual representation of comprehensive toxicity for a specific chemical (right figure). EPA envisions using "Pis" to help prioritize chemicals.

For more information on ToxPi visit:

- ToxPi Webpage
- Factsheet: ToxPi

To learn about the possible benefits of these new technologies, proceed to **Section 5: Potential of High-Throughput *In Vitro* Approaches.**

Section 5: Potential of High-Throughput *In Vitro* Approaches

For decades we have endured a poor chemicals policy that has created a situation where too many chemicals in the marketplace have too few data to judge their safety. The lack of adequate government authority to require testing, along with its costs and concerns about use of laboratory animals, has meant that chemical safety assessment has not even come close to keeping up with the ever-expanding number of chemicals and their myriad uses in products and materials. As a result, we are unnecessarily exposed to harmful or untested chemicals and have to spend significant resources ameliorating problems that could and should have been prevented.

Scientific research has made significant strides in elucidating the many ways chemicals can affect human health and environment. We now know that some chemicals present at very low doses in our bodies can have adverse health impacts, that the timing of exposure is critical in defining an ultimate health effect, and that diversity in the population—such as genetics, age, and gender— makes us differentially susceptible and vulnerable to chemical exposures.

Inadequate chemicals policy alongside new insights gleaned from scientific research is driving EPA to develop new approaches to fill data gaps on chemicals that reflect a 21st-century understanding of the biological activity of chemicals. High-throughput *in vitro* testing (HT testing), such as that being developed in EPA's ToxCast program, is one of the technological solutions being most intensively pursued. HT testing is discussed in greater detail in Section 2 of the primer, and ToxCast is discussed in Section 3 of the primer.

HT testing hold great promise, but like all testing approaches, has limitations as well as strengths, and if used inappropriately could actually set us back, not forward. We discuss here some of the key potential benefits associated with HT testing; Section 6 of this primer discusses some of the key limitations and challenges of HT testing.

Key Benefits

- Speed

  Conventional toxicity testing methods generally involve dosing laboratory animals with a chemical of interest and after some period of time—hours to days to months to years—looking to see whether an adverse outcome, for example, a tumor, has developed. Here the scientist is observing the ultimate downstream consequence of what is presumed to be a chemical's interference with the proper function of one or more biological pathways. In contrast, HT methods primarily focus on the pathways themselves and aim to "catch" an early indicator of hazard: the perturbation of the pathway, rather than the ultimate consequence of that perturbation. This approach requires much less time—a matter of seconds to minutes. Not only does the effect happen sooner after exposure, but it can often be observed in something less than the whole animal, e.g., in a culture of cells or even a solution of cell components.

Robots quickly assess thousands of chemicals in multiple assays in an automated manner. Photo by EPA

Because these HT tests can be done in a very small volume of solution, it also means that many chemicals can be put through a battery of HT assays simultaneously and at relatively low cost. Moreover, the process can be automated, indeed even carried out by robots. Hence, thousands of chemicals can be analyzed in hundreds of assays all in a period of time far shorter than would be required to detect most adverse outcomes in intact laboratory animals. Given the massive backlog of chemicals with little or no safety data, the speed of HT tools could be very valuable, at least in screening and prioritizing chemicals by level of potential concern.

- Human relevance

  Because of the ethical problems associated with human testing, as well as the simple fact that we live so long, traditional toxicological methods use laboratory animals to assess the toxicity a chemical and predict its effects on people. According to the seminal National Academy of Sciences report "Toxicity Testing in the 21st Century: A Vision and a Strategy," use of such animal "models" is possible because in general human biology is similar to that of test animals. Nonetheless, while animal studies have served as important and useful tools, uncertainty is introduced by the need to extrapolate from animal data estimates of risk to humans. Moreover, there are some chemicals that elicit toxic effects in one species and not in another. For example, thalidomide is toxic to human fetuses, but rats are resistant to its effects.

  In addition to testing chemicals on cells originally derived from laboratory animals, EPA's ToxCast HT assays can be conducted on human cells, which are grown in culture. This may improve the ability of such tests to predict whether a chemical will be toxic to humans. Ideally, this could help lower the likelihood of a cross-species "false negative," that is, missing an effect because it happens not to occur in the lab animal chosen for a given test but would occur in a human – or, conversely, a "false positive," that is, seeing an effect in the animal model that for some reason would not occur in people.

- Multiple cell types and life stages

  HT testing methods offer the potential to look for effects of chemicals on different cell types (*e.g.,* liver cells, kidney cells, etc). This may discern different kinds of toxicity, including those resulting from a chemical's ability to disrupt a process that only takes place in certain cell types or organs. Some HT assays even use combinations of cell types taken directly from human tissues, in an effort to mimic responses of, and interactions between, cells types that are involved in the body's reaction to a particular disease or disorder (*e.g.,* asthma).

  HT testing methods also hold potential to look for effects of chemicals at different life stages. A particularly exciting *potential* application of HT tests is in evaluating chemical effects on early life stages, including fetal development. For example, the Texas-Indiana Virtual STAR

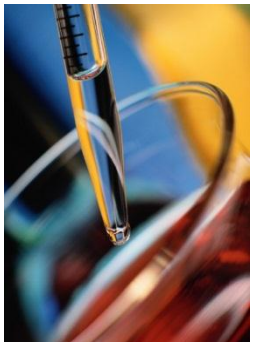
Center is using mouse embryonic stem cells to determine how chemicals may affect key biological pathways during early fetal development. This kind of research is still at an early stage, but holds the promise of providing a means to rapidly screen a large number of chemicals for developmental toxicity.

- Exposure Relevance

  Chemical testing in laboratory animals is typically done at high doses to ensure that, if an adverse effect occurs, it can be detected in a relatively small number of animals in a relatively short period of time. These concentrations are often orders of magnitude higher than what a person would actually experience. Methods are then used to extrapolate any observed effects from such high-dose exposure to lower concentrations more representative of "real-world" exposure. Extrapolation from high- to low-dose effects raises often contentious questions about the appropriate dose-response relationship and whether low-dose effects differ from those seen at high doses. An advantage of HT methods is that a wide range of doses, including very low doses, can be tested. Such capabilities may also assist in resolving disputes around chemicals' ability to cause different effects at low doses than they cause at high doses.

- Assessing Mixtures

  Human biomonitoring data confirm common sense in that they reveal that we are exposed to a complex mixture of chemicals. Such exposures extend from early fetal development through adulthood. High-throughput assays offer a means to test exposures to multiple chemicals and at multiple doses. Attempting to do so in traditional animal tests would prove prohibitive in terms of time, cost and use of laboratory animals.



New HT testing methods may improve our ability to examine the effects of mixtures of chemicals we experience in the real-world.

- Crisis Situations

  In crisis situations where there is limited time to evaluate a chemical or mixture before its use, batteries of quick high-throughput tests might help to make a more informed decision. However, even in these situations care should be taken to clearly communicate any limitations and uncertainties associated with these decisions and the data informing them. For example, EPA used some of its ToxCast assays to examine potential endocrine-disrupting effects of dispersant chemicals used to clean up the BP oil spill. Given the crisis state of the situation (and putting aside the fact that the testing came *after* millions of gallons of the dispersants had already been used, begging the question of why more thorough testing hadn't been conducted well before then), this information was helpful. However, EDF expressed concern (see here and here) regarding the poor communication of the results of such assays that was perceived as having effectively exonerated these chemicals from having *any* endocrine-disrupting activity – let alone other effects – despite the significant limitations of the available assays. The important lesson here is that the new technologies shouldn't be given explicit credit beyond their actual capabilities in any situation, crisis or otherwise.

- Green Chemistry

HT technology could help in the design of safer chemicals by quickly screening out potentially toxic chemicals during chemical development. Photo by EPA

High-throughput technologies hold promise for informing safer chemical design and selection. These assays could flag potential toxicity concerns for new chemicals during early research, design, and development phases. Many of the HT technologies used in ToxCast and related programs in fact originate from the pharmaceutical industry, where they have been used for many years in drug discovery to screen out drug candidates that appear ineffective or show indications of hazard, and to push drugs that show more potential toward further development and evaluation. Efforts are already underway to integrate HT tools into green chemical design. A workshop held in September of 2011 brought expert scientists together to discuss how HT and other computational technologies can be used for safer chemical design.

To learn about some of the barriers to full implementation of this technology, proceed to **Section 6: Challenges and Limitations of High-Throughput *In Vitro* Testing**

Section 6: Challenges and Limitations of High-Throughput *In Vitro* Testing

While high-throughput *in vitro* testing (HT testing) offers many potential benefits (detailed in Section 5), the critical question is whether these tests will improve our ability to accurately identify and predict hazardous effects of chemicals and the risks they present to the human population. Today there are a number of key limitations and challenges to using HT tests to assess chemical risks.

If HT assays are to be used more extensively in chemical testing, the following question must be asked: Can tests conducted *in vitro* accurately reflect the effects that a chemical would have in the more complex and complete environment of a whole animal, including a human? In other words, can they accurately predict adverse outcomes in whole animals, including people? This question has a number of dimensions.

Key Challenges

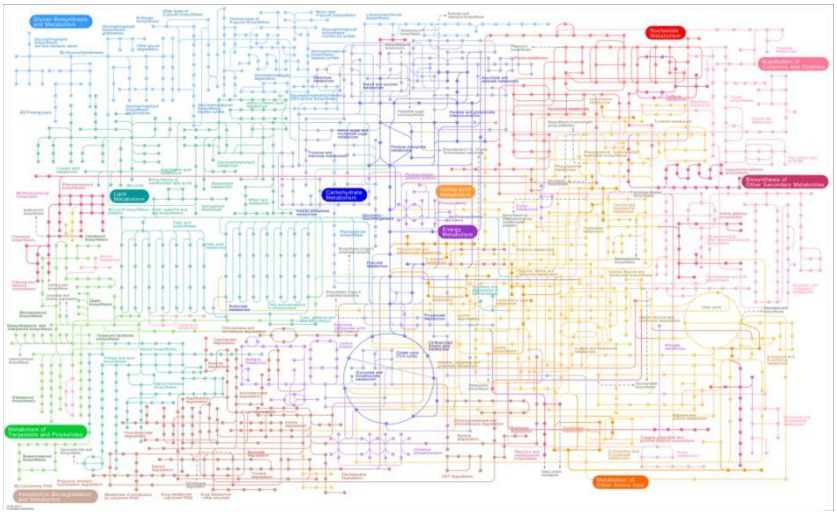
- *In vivo* versus *in vitro*

  Traditional toxicity testing aims to determine whether a particular dose of a chemical results in an observable change in the health or normal functioning of the *whole animal*. In contrast, HT tests examine whether and by what mechanism a chemical induces changes at *cellular and molecular levels*. Such changes may be precursor events leading to an actual disease outcome, in some cases picking up effects that can't easily be detected or measured in traditional whole animal tests.
  Both EPA and the National Research Council acknowledge that HT methods do not capture all relevant processes—at least not yet—that occur within the more complex system of a whole tissue or whole organism. To quote an EPA study, "The most widely held criticism of this *in vitro*-to-*in vivo* prediction approach is that genes or cells are not organisms and that the emergent properties of tissues and organisms are key determinants of whether a particular chemical will be toxic."

- Coverage of the full biological response landscape

  Determining whether a chemical perturbs a biological pathway requires that all key events in the pathway—and any auxiliary molecular activity associated with that pathway, such as epigenetic processes (see epigenetics below)—are included in the battery of HT assays being used. In other words, it's impossible to detect an adverse effect if it's not being tested for. As Dr. Robert Kavlock, Deputy Assistant Administrator for Science at the EPA has stated, "And then another lack that we have is we're looking at 467 [HT] assays right now. We may need to have 2,000 or 3,000 assays before we cover enough human biology to be comfortable that when we say something doesn't have an effect, that we've covered all the bases correctly." (The Researchers Perspective Podcast, 2010. Read the full podcast transcript [PDF].)

Likewise, during the NexGen Public Conference in February 2011, Dr. Linda Birnbaum, Director of the National Institutes of Environmental Health Sciences (NIEHS), pointed to significant gaps in our understanding of biological pathways. She described gene targets relevant to disease pathways involved in diabetes that are not currently included in the ToxCast HT battery of assays. These gene targets were identified by experts during an NIEHS workshop on chemicals and their relationship to obesity and diabetes. It will be critical for ToxCast-like efforts to continuously mine and integrate the latest science into their HT assay batteries.



EPA's use of HT *in vitro* testing must ensure adequate coverage of the biological response landscape; composed of numerous, complex, and interconnected biological pathways involved in the progression of adverse health outcomes. Photo by Kyoto Encyclopedia of Genes and Genomes

- Accounting for chemical metabolism

When chemicals are studied in whole animals, the effects observed are dependent in part on how the body metabolizes the substance. One critically important factor challenging the predictive ability of *in vitro* testing is whether and to what extent such methods capture the mechanisms animals use to metabolize chemical substances. The toxicity—or lack of toxicity—of a chemical is not always derived from the chemical itself, but rather from the rate at which it is broken down and the nature of the breakdown products (called metabolites). A classic example is the polycyclic aromatic hydrocarbon (PAH), benzo[a]pyrene: It is the metabolites of the chemical that are mutagenic and carcinogenic. Metabolism can also work in reverse, of course, rendering a toxic chemical less or non-toxic.

Many of the HT assays utilized in ToxCast and other HT systems lack explicit metabolizing capabilities. EPA is exploring ways to better account for whole-animal capabilities such as metabolism in HT testing, but until there is greater confidence that these complexities are accounted for, this factor will continue to limit the extent to which *in vitro* HT test data can be considered fully predictive of *in vivo* effects.

- Ability to account for diversity in the population

HT *in vitro* tests using genetically-identical cell lines may not accurately predict the effects of chemicals on the diverse human population.

Another challenge is the ability to account for real-world diversity among the human population that influences susceptibilities to toxic chemical exposures. Individual differences in our genomes, epigenomes, life stage, gender, pre-existing health conditions and other characteristics are integral to determining the ultimate health effect of a chemical exposure. This is a challenge, of course, for traditional animal toxicity tests as well as for newer HT methods.

Traditional animal toxicity tests typically use inbred, genetically identical (isogenic) animal strains to generate results that must then be extrapolated to predict a chemical's effect in the much more diverse human population. Similarly, newer HT methods typically use homogenous populations of cells or components drawn from such cells. While there are often good reasons to start with genetically homogenous populations of animals or cells, their use limits the ability to make accurate predictions about effects in very diverse human populations.
This challenge has not escaped federal researchers, who are testing thousands of compounds on different human cell lines to better account for differential susceptibility to effects. Indeed, recently published scientific research reveals that genetically diverse cell lines responded differently to certain compounds in a HT testing system, suggesting that using a diversity of cell lines is one approach to incorporating genetic diversity in the population.

- Accounting for multiple exposures and different timing of exposure

We know that in the real world we are exposed to a complex mixture of chemicals, not one chemical at a time. And we are learning that the timing of such exposures—early in fetal development, early childhood, or in adulthood—influences the health outcome. Capturing this complexity of exposure presents a fundamental challenge to the use of HT testing (as well as to traditional toxicity testing).

- Accounting for different patterns of exposure

The ultimate impact of a chemical exposure on our health may be quite different depending on the duration, frequency, and level of exposure. For example, the effects of acute, high dose exposures can be quite different than those that result from continuous, low dose exposures. Issues relating to the frequency and duration of exposure have been acknowledged by agency experts in a peer-reviewed publication: "A related challenge is the understanding of what short timescale (hours to days) *in vitro* assays can tell us about long-timescale (months to years) processes that lead to *in vivo* toxicity end points such as cancer." The frequency and duration of real-world chemical exposures will need

to be either directly addressed in HT assays or otherwise somehow integrated into the interpretation of HT testing data.

- Determining a significant and adverse level of perturbation

  At some point, an informed decision will need to be made as to what level of chemically-induced perturbation observed in an HT assay is considered sufficiently indicative or predictive of an adverse effect in a human. In other words, even if an assay *performs* perfectly (i.e., yields no false positives or false negatives – see bullet below), determining how to interpret and translate HT data into a measure of actual toxicity in humans is a challenge—further complicated by the issues of individual and population variability discussed earlier.

  Just as in our efforts to deal with data from existing testing methods, there will need to be decision rules that govern how to extrapolate HT data to humans so as to measure the *intensity* of effect at a given dose, not just whether or not there is an effect. Translating and interpreting such data to inform decisions about toxicity and risk to humans will also require transparent and clear delineation of where value judgments or assumptions enter into decision-making.

- Insufficient accounting for epigenetic effects

  Epigenetics is a burgeoning field of science that studies how gene expression and function can be altered by means other than a change in the sequence of DNA, i.e., a mutation. Epigenetic programming of our genes is critical for normal human development and function. For example, epigenetics is the reason why the single fertilized egg we all began as differentiates into the more than 200 different types of cells that make up our adult bodies.
  Evidence is increasing that certain chemicals can interfere with normal epigenetic patterns. For example, epigenetic changes induced by tributyltin have been shown to influence the programming of stem cells to become fat cells instead of bone cells. The current ToxCast battery of assays is quite limited in explicitly measuring epigenetic effects of chemicals (see this NIEHS presentation slide 13 and this description of one of the few such assays currently available currently available).

- False Negatives/False Positives

  Fundamental to the success of HT assays is their ability to correctly identify chemicals that are – or are not – of concern. EPA's approach to validating HT tests largely involves testing chemicals with already well-defined hazard characteristics based on traditional animal testing. By using well-studied chemicals, the agency plans to determine the extent to which HT assays accurately detect those hazards identified in the whole-animal studies.
  Determining how accurately a HT assay identifies a chemical's hazards includes assessing the "false positive" and "false negative" rates of the test. If the false negative rates are high in the HT assays used to screen chemicals for further assessment, a

potentially hazardous chemical could be erroneously determined to be of low concern and side aside. As Dr. Bob Kavlock, explained, "You want to have as few false negatives as possible in the system, because if you put something low in a priority queue, we may never get to it, and so you really want to have confidence that when you say something is negative, it really does have a low potential." (The Researchers Perspective Podcast, 2010. Read the full podcast transcript [PDF].)

During the 2011 NIEHS Workshop on the "Role of Environmental Chemicals in the Development of Obesity and Diabetes," experts examining organotins and phthalates noted that ToxCast high-throughput assays did not successfully identify chemicals known to interfere with PPAR—a protein important for proper lipid and fatty acid metabolism—in assays designed to flag this interference.

New chemical testing approaches offer great potential to address some of the long-standing limitations of chemical risk assessment (detailed in Section 1 of this primer). But as discussed above, there are several major challenges—many also applicable to traditional toxicity testing—that need to be met in the development and use of HT testing and other newer approaches. Moreover, new challenges will continue to arise as a consequence of the ever-evolving nature of science. While there may not be immediate solutions to the challenges we face, it is profoundly important that current limitations are acknowledged, characterized and communicated to decision-makers and stakeholders so that new data or new assessments can be appropriately interpreted and appropriately used.

To learn about how new chemical testing technology is being incorporated into risk assessment, proceed to **Section 7: Advancing the Next Generation (NextGen) of Risk Assessment**

For a commentary on the need for public engagement in the development and use of these new methods, proceed to **Section 8: The Need for a Public Interest Perspective**

Section 7: Advancing the Next Generation (NexGen) of Risk Assessment

NexGen is an EPA-led effort to integrate data derived from new testing approaches, including CompTox, into risk assessment. The NexGen program is housed within the National Center for Environmental Assessment which itself is part of EPA's Office of Research and Development.

Presented in a 2012 publication in *Environmental Health Perspectives*, the NexGen program is seeking to address the following broad set of questions:

- "How can these new data and methods substantively improve our understanding of risk?"
- "Can scientifically sound assessments be made faster, cheaper, and/or more accurate using these new methods, and better address a variety of environmental management challenges (risk context)?"
- "How can these new types of information best be incorporated into risk assessments and utilized to inform risk managers and the public?"
- "What new policies and procedures are needed to produce consistent, reasonable, and robust assessments?"

To develop a framework for incorporating new types of data into risk assessment, NexGen is conducting a series of case studies, or "prototypes" that each pair specific chemicals with an associated disease. The first three disease-chemical prototypes are: 1) cancer and benzene, benzo(a)pyrene and other polycyclic aromatic hydrocarbons; 2) endocrine disruption and BPA, phthalates, and perchlorate; and 3) lung injury and ozone and chlorine.

Like Tox21, NexGen is an inter-institutional program and includes the Environmental Protection Agency's Computational Toxicology Program, National Institutes of Environmental Health Sciences, National Toxicology Program, Centers for Disease Control Agency for Toxic Substances and Disease Registry, National Human Genome Research Institute, and the State of California's Environmental Protection Agency.

For more information on NexGen visit:

- NexGen webpage
- Publication in Environmental Health Perspectives: "Advancing the Next Generation of Risk Assessment" August 2012.

To learn about the need for greater public engagement in this area, proceed to **Section 8: The Need for a Public Interest Perspective**

Section 8: The Need for a Public Interest Perspective

EDF supports EPA's investment in developing new approaches to identify and characterize the potential health concerns posed by chemicals. There are practical policy-related and scientific reasons for developing new tools that:

The expertise of academic and NGO communities is needed to help guide the development and implementation of EPA's new toxicity testing programs.

- reflect a more sophisticated and current scientific understanding of our biology;
- are able to be efficiently applied at low cost to large numbers and mixtures of chemicals;
- significantly improve our ability to reliably predict adverse effects chemicals may exert at various doses and stages of life and across a diverse human population;
- help us understand not only whether, but also *how*, chemical exposures can exert adverse effects;
- improve not only our ability to predict adversity but to design out of new chemicals problematic characteristics of chemicals that lead to adverse effects; and finally
- are credible, understandable—at least at a basic level—and accepted by the full range of decision-makers and stakeholders.

The call for a new vision and future of toxicity testing was outlined in the National Academy of Sciences' (NAS) 2007 publication, "Toxicity Testing in the 21st Century: A Vision and a Strategy." The problem, as described in the report, is that "The current system, which relies primarily on a complex set of whole-animal-based toxicity-testing strategies for hazard identification and dose-response assessment, has difficulty in addressing the wide variety of challenges that toxicity testing must meet today." The report lays out a vision for developing and incrementally integrating new methodologies into the practice of risk assessment. ToxCast is a first big step—albeit incomplete and imperfect—towards realizing that vision.

While the new testing approaches hold great promise for dealing with long-standing deficiencies in risk identification and assessment, they also present challenges that must be confronted if they are truly to lead to improvements. It is critical that these challenges, and the efforts taken to address them, are clearly communicated to the public. This is especially important as these methods begin to more formally enter into EPA decision-making activities, whether for purposes of prioritizing chemicals, informing risk assessments, or making regulatory decisions.

If federal agencies are serious about advancing these testing methods to a point where they can form the core of a new testing and assessment paradigm, there will need to be broad confidence that they can serve as a sufficient basis for making regulatory and other policy decisions. A corollary implication is that data derived using the new methods must be able to meet statutory and regulatory standards governing how the safety of a chemical is to be determined. This is the ultimate challenge: to move the new methods from the research and development phase to serve as a part of the basis for risk management and other regulatory determinations.

To meet this challenge, regulatory bodies will ultimately need to attain sufficient buy-in or acceptance from relevant stakeholders in the industry, academic, NGO and governmental sectors. And to achieve that buy-in, at a minimum each of the challenges laid out in this primer will need to be addressed. All of this will not happen overnight, of course, and will likely take many years and be constantly evolving. But it is imperative that this ultimate challenge is kept in sight, guiding the development of newer testing strategies as they move forward.

Academic scientists and public interest groups have a critical role to play in ensuring that EPA incorporates the best science into its new testing program and that these efforts serve the public interest.

Academic scientists have unique expertise and knowledge regarding experimental design, validation of data, model development, and of course, the mechanisms by which chemicals interact with, and ultimately, affect our biology. While EPA does have expertise in these areas, the ambition and promise of this endeavor demands the input of a diverse community of scientists to ensure that the best thinking and information is brought to bear in the development and use of these methods. Ultimately, these new approaches are only as strong as the science and critical-thinking upon which they are built.

Similarly, ensuring that EPA communicates and uses newer methods in a transparent and appropriate manner that meets public health needs requires the input of the public interest community. Effective engagement of this community should not be assumed to demand a high level of technical expertise, though it clearly requires some familiarity with the nature and aims of newer approaches and programs. Some of the most difficult challenges surrounding the use of these methods relate to value judgments. For example, determining acceptable levels of confidence and uncertainty for particular uses of new types of data is inherently value-laden and needs to reflect the expectations, values, and tolerances of the public. The public also has an important role to play in helping EPA to present and communicate information related to these methods in a manner that is transparent and accessible to the lay person.

A sustained dialogue between EPA and outside stakeholders including academia and the public interest community is needed. EDF created this primer as a resource for NGOs, academics and others interested in engaging to build a better future for toxicity testing and risk assessment.

To help our communities enter into and sustain a dialogue with EPA, EDF has also coordinated a series of webinars that outline the basics of the agency's testing programs and the promises and challenges ahead. Video recordings of these sessions can be accessed through the Chemical Testing in the 21st Century: Webinar Series webpage.