Green Chemistry

Dynamic Article Links

Cite this: DOI: 10.1039/c2gc35055f

www.rsc.org/greenchem

PAPER

Designing endocrine disruption out of the next generation of chemicals†

T. T. Schug,*[a] R. Abagyan,[b] B. Blumberg,[c] T. J. Collins,[d] D. Crews,[e] P. L. DeFur,[f] S. M. Dickerson,[g] T. M. Edwards,[h] A. C. Gore,[i] L. J. Guillette,[j] T. Hayes,[k] J. J. Heindel,[a] A. Moores,[l] H. B. Patisaul,[m] T. L. Tal,[n] K. A. Thayer,[o] L. N. Vandenberg,[p] J. C. Warner,[q] C. S. Watson,[r] F. S. vom Saal,[s] R. T. Zoeller,[t] K. P. O'Brien*[g] and J. P. Myers*[u]

Received 12th January 2012, Accepted 4th September 2012
DOI: 10.1039/c2gc35055f

A central goal of green chemistry is to avoid hazard in the design of new chemicals. This objective is best achieved when information about a chemical's potential hazardous effects is obtained as early in the design process as feasible. Endocrine disruption is a type of hazard that to date has been inadequately addressed by both industrial and regulatory science. To aid chemists in avoiding this hazard, we propose an endocrine disruption testing protocol for use by chemists in the design of new chemicals. The Tiered Protocol for Endocrine Disruption (TiPED) has been created under the oversight of a scientific advisory committee composed of leading representatives from both green chemistry and the environmental health sciences. TiPED is conceived as a tool for new chemical design, thus it starts with a chemist theoretically at "the drawing board." It consists of five testing tiers ranging from broad *in silico* evaluation up through specific cell- and whole organism-based assays. To be effective at detecting endocrine disruption, a testing protocol must be able to measure potential hormone-like or hormone-inhibiting effects of chemicals, as well as the many possible interactions and signaling sequellae such chemicals may have with cell-based receptors. Accordingly, we have designed this protocol to broadly interrogate the endocrine system. The proposed protocol will not detect all possible mechanisms of endocrine disruption, because scientific understanding of these phenomena is advancing rapidly. To ensure that the protocol remains current, we have established a plan for incorporating new assays into the protocol as the science advances. In this paper we present the principles that should guide the science of testing new chemicals for endocrine disruption, as well as principles by which to evaluate individual assays for applicability, and laboratories for reliability. In a 'proof-of-principle' test, we ran 6 endocrine disrupting chemicals (EDCs) that act *via* different endocrinological mechanisms through the protocol using published literature. Each was identified as endocrine active by one or more tiers. We believe that this voluntary testing protocol will be a dynamic tool to facilitate efficient and early identification of potentially problematic chemicals, while ultimately reducing the risks to public health.

[a]*Division of Extramural Research and Training, Cellular, Organ and Systems Pathobiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. E-mail: schugt2@niehs.nih.gov, kpobrien@advancinggreenchemistry.org, jpmyers@ehsciences.org*
[b]*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA*
[c]*Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, USA*
[d]*Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA, USA*
[e]*Section of Integrative Biology, University of Texas at Austin, Austin, TX, USA*
[f]*Center for Environmental Studies, Virginia Commonwealth University, Richmond, VA, USA*
[g]*Advancing Green Chemistry , Charlottesville, VA, USA*
[h]*School of Biological Sciences, Louisiana Tech University, Ruston, LA, USA*
[i]*Institute for Neuroscience, University of Texas at Austin, Austin, TX, USA*
[j]*Department of Obstetrics and Gynaecology, Medical University of South Carolina, Charleston, SC, USA*
[k]*Laboratory for Integrative Studies in Amphibian Biology, Molecular Toxicology, Group in Endocrinology, Energy and Resources Group, Museum of Vertebrate Zoology, and Department of Integrative Biology, University of California, Berkeley, CA, USA*
[l]*Centre for Green Chemistry and Catalysis, Department of Chemistry, McGill University, Montréal, QC, Canada*
[m]*Department of Biology, North Carolina State University, Raleigh, NC, USA*
[n]*Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR. Current Address: Integrated Systems Toxicology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA*
[o]*Office of Health Assessment and Translation, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA*
[p]*Center for Regenerative and Developmental Biology and Department of Biology, Tufts University, Medford, MA, USA*
[q]*Warner Babcock Institute for Green Chemistry, Wilmington, MA, USA*
[r]*Center for Addiction Research, University of Texas Medical Branch, Galveston, TX, USA*
[s]*Division of Biological Sciences and Department ,University of Missouri-Columbia, Columbia, MO, USA*
[t]*Biology Department, University of Massachusetts-Amherst, Amherst, MA, USA*
[u]*Environmental Health Sciences, Charlottesville, VA, USA*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c2gc35055f

Introduction

As noted by Anastas and Warner,[1] most efforts at reducing risk to human health from chemicals have focused on reducing the probability and magnitude of exposures. That approach works, until it fails. Failure, unfortunately, is virtually inevitable, because of accidents and practices not part of the 'intended use of a product.' There are a multitude of examples of unintended exposures including accidents like the accidental release of methyl isocyanate gas at Bhopal, BP's Deepwater Horizon oil spill, the recycling of electronic waste by children in China and India, and household dust in California containing flame retardants.

Green chemistry takes a different approach. One of its fundamental goals is to synthesize chemicals that are not hazardous for human health and the environment. To achieve this goal efficiently, chemists must be able to assess potential hazards of the chemicals that they develop.

We use the word 'hazard' deliberately: 'hazard' is embedded in green chemistry as one of the two determining elements of risk. It is commonly accepted that risk is a function of inherent hazard and exposure. Green chemistry deals with risk by seeking to eliminate inherent hazard rather than by controlling exposure.[1] Ideally this assessment would take place as early in the design process as feasible so that decisions can be made whether to pursue further development. If a hazard is identified, the chemist can opt either to cease development of that chemical or to manipulate the molecular structure to design against hazard.

In an ideal world, it would be possible to predict with confidence the potential toxicity of new molecules based on their structure and physical characteristics. Well-known weaknesses in these approaches, however (note for example the 'Structure Activity Relationship Paradox' discussed below), render this approach not just inadequate, but potentially misleading. In this endeavor, such potential for false positives and false negatives is unacceptable. Actual biological experiments are therefore necessary.

Because chemists typically are not trained in toxicology or other relevant fields, developing the means to achieve this goal requires collaboration between environmental health scientists and green chemists. This collaboration, systematically applied and constantly adjusted to reflect new scientific discoveries, would help lead to a new generation of inherently safer chemicals.

In this paper we explore how chemists can apply principles and tests from the environmental health sciences to identify potential endocrine disruptors. Specifically, we propose a five-tiered testing protocol, TiPED. We begin with computational approaches as the fastest and the least expensive assays. Subsequent tiers involve increasingly specialized tests to determine the potential for endocrine disrupting characteristics of a chemical under development. Some of the assays are based on known mechanisms of action; some are designed to catch disruptions for which the mechanisms or receptors are as yet unknown. We present the overall structure of the protocol with assay examples that could be used in each tier.

We noted above that actual biological experiments are necessary for predicting toxicity. This is especially the case with endocrine disruption because of the complex signaling events that control endocrine activity within and between cells, tissues and organs. We discuss this issue in greater detail as we discuss the strengths and weaknesses of our different tiers.

We present the tiers in a logical sequence for a chemist designing a new chemical: from the simplest approach (and least expensive) through the more complex (and often more expensive). We recognize, however, that different users will have different needs. A user can start anywhere in the system, not necessarily with Tier 1. An academic research chemist drawing a molecule *de novo* will have different issues and questions than an industrial chemist with a molecule already in hand; the former would be more likely go through the protocol in a linear progression. The latter might use assays in a later tier to get a quick read on the likelihood of potential problems. Some users may want a straight "harm/no likely harm" answer, abandoning failed molecules rather than developing them into products. Others, after getting a "harm" result, might pursue a series of increasingly specific assays to identity mechanisms of biological action so that they might redesign the product. To reiterate, though presented here in a linear fashion for the sake of new chemical design, other users can enter the system where it best meets their needs.

To support the tiered assay system, we also identify a suite of principles that should be used to guide implementation (discussed after summary of TiPED). These principles focus both on general concerns about toxicity testing as well as unique characteristics of endocrine disrupting compounds (EDC) that makes their detection particularly challenging.

At this stage, TiPED is a scientific framework in progress. This paper presents the overall strategy, its scientific rationale and the principles that govern its design and implementation. The formal protocol itself will be presented on the TiPED website (www.TiPEDinfo.com). The website will undergo formal peer review and invite constant input from EDC specialists and chemists who use it.

Scientific understanding of endocrine disruption is advancing rapidly. New mechanisms of endocrine disruption, new targets for EDC action and new ways to measure the effects of EDCs are being reported regularly. Any effective testing protocol must evolve as new scientific discoveries are reported. The guiding principles behind this testing protocol, however, remain constant.

We choose to focus on endocrine disruption for three reasons. First, the body of evidence that has emerged from the past 20 years of research on this class of mechanisms has grown, indicating it is a serious public health issue. Second, it is clear that the current paradigm focused on exposure, instead of hazard, has failed to protect public health from endocrine disruption. Measurements by the U.S. Centers for Disease Control and Prevention document widespread exposure to multiple EDCs at levels that current scientific research suggests may not be safe. Third, despite a 1996 Congressional mandate to develop toxicity assays for EDCs, the United States Environmental Protection Agency (U.S. E.P.A.) has made little progress in implementing the use of EDC assays in the regulatory process. With this focus in mind, we invited leading experts in endocrine disruption science to collaborate with leading green chemists to develop a testing protocol that could be used by chemists as a voluntary—not regulatory—design tool (Table S1†). This allowed us to focus on scientific issues, rather than regulatory debates.

1. What is endocrine disruption?

The endocrine system uses chemical signals—hormones—to direct development and reproduction, regulate body function and metabolism, and influence behavior and immunity.[2] In its broadest sense, endocrine disruption takes place when an agent alters hormone signaling or the response to hormone signaling, and in so doing alters some aspect of the organism under hormonal control. According to the Endocrine Society, the world's authoritative scientific association of clinical and research endocrinologists, an endocrine-disrupting chemical (EDC) is an exogenous chemical, or mixture of chemicals, that can interfere with any aspect of hormone action.[3]

Endocrine disruption can be caused by diverse mechanisms. Hormones work by binding with protein receptors in the cell membrane, the cytoplasm or the nucleus. Binding initiates gene activity or physiological processes (depending upon the receptor, its location, hormone concentration, and the developmental state of the cell/tissue/organism) that are part of and essential to normal organismal function. EDCs work by interfering with that signaling process. They are not necessarily structurally similar to hormones; many, but not all, are lipophilic.

Mechanisms of action include: the EDC binds to the receptor and adds to the normal signal; the EDC binds to the receptor and blocks the normal signal; the EDC affects hormone synthesis (increasing or decreasing the amount of natural hormone that is available for signaling); the EDC alters hormone metabolism or hormone transport and storage within bodily tissue (again, increasing or decreasing hormone amount); and/or the EDC affects the levels of mature hormone receptor *via* disruption or modulation of gene expression, folding, or transport.

A central part of the phenomenon of endocrine disruption is receptor binding, which depends upon the molecular conformation of the hormone and its receptors. Molecular structure is a good, but imperfect predictor of whether binding will occur; chemists can use information about structure both to predict potential hazard (described below) as well as to guide manipulation of a chemical's structure to avoid hazard.

A crucial aspect of hormone action is that it takes place at extremely low concentrations. For an estrogen, for example, typical physiological levels of the biologically-active form of an estrogen are extremely low, in the range of 10–900 pg ml^{-1} (high parts per quadrillion to low parts per trillion). This is possible because of the specificity of hormone binding to its receptor, and is biologically necessary because of the large number of signaling molecules present at any one time. Specificity and extreme sensitivity make it possible for an enormous number of signaling molecules to co-exist in circulation [4] without disrupting each other's signaling. The specificity also evolved, presumably, to reduce or avoid disruption by exogenous compounds with which organisms have had evolutionary experience.

Within the past century, over 80 000 new chemicals have been synthesized and used in ways that have resulted in widespread human exposures. A subset of these chemicals are toxic; a subset of these toxic chemicals are toxic due to endocrine disruption. A small number of these chemicals have been created explicitly to alter hormone signaling, *e.g.*, the estrogenic drug diethylstilbestrol and many pesticides (for target species). Other chemicals have molecular structures that unintentionally bear sufficient resemblance to hormones such that they are capable of binding, with varying degrees of affinity, to hormone receptors, or of interacting at the molecular level with other molecules involved in hormonal activity. Often EDCs are much less potent than the endogenous hormones in binding with receptors. An increasing number of examples appearing in the peer-reviewed literature, however, show that in some signaling pathways exogenous hormone-mimics can be equipotent and capable of provoking biological responses at picomolar (pM) levels or lower.[5]

Most early research on EDCs focused on the effects of disruption of sexual reproduction *via* interactions with the estrogen and androgen nuclear receptors. Evidence gathered over the past decade now shows that the mechanisms and endpoints vulnerable to endocrine disruption are much broader than originally understood. Indeed, EDCs are now known to affect metabolism, diabetes, obesity, liver function, bone function, immune function, learning and behavior *via* a panoply of receptor systems and signaling pathways. In addition, the actions of EDCs on reproduction are now known to go far beyond nuclear sex steroid hormone receptors. In principal, there is virtually no endocrine signaling system or hormone pathway immune to disruption (Fig. 1).

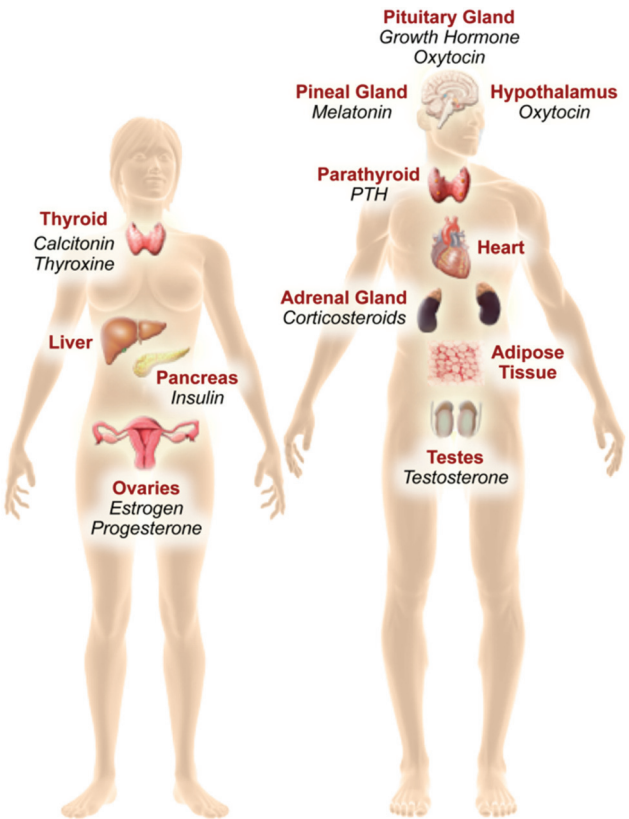

Fig. 1 The endocrine system is comprised of the hypothalamus, pituitary, adrenal gland, parathyroid, pineal gland, thyroid, pancreas, and reproductive glands. Other tissues and organs such as the liver, heart, and adipose tissue have secondary endocrine functions, and may also be targeted by EDCs. Endocrine glands secrete a hormone, which is carried throughout the body *via* the blood, and may bind to its specific receptor in target organs. For instance, estrogen is released by the ovary and binds to estrogen receptors (ERα and ERβ) distributed throughout the body and brain.

The majority of research on EDCs has examined the consequences of their interactions with nuclear hormone receptors (NRs), especially estrogen receptors alpha and beta (ERα, ERβ), the androgen receptor (AR), among others. NRs are a superfamily of transcription factors, proteins that can bind to DNA and influence the expression of nearby genes. NRs play central roles in development, physiology and disease. In humans, there are some 48 identified NRs. Many others remain "orphans," meaning that their endogenous ligands have not yet been identified. When activated, NRs undergo conformational changes that allow recruitment of co-regulatory molecules and the chromatin-modifying machinery of the cell. The ultimate action of NRs is to influence the transcriptional machinery of target genes. NRs also interact with other intracellular signaling pathways. Examining how chemicals bind to these receptors can provide important information concerning their endocrine disrupting potential. There are *in vitro* assays, some of which can be performed as part of high throughput, screening systems that can confirm chemical binding to the majority of NRs. The strengths and weaknesses of *in vitro* tools in predicting hazard will be discussed below in the section on Tier 2.

Endocrine disruption also takes place outside the cell nucleus. Many natural steroid hormones bind to cell membrane-bound receptors, which in turn partner with a variety of well-known signaling cascade proteins. Recent evidence demonstrate that EDCs may exert hormonal effects *via* these non-nuclear hormone receptors as well. Rather than acting as transcription factors, membrane hormone receptors act *via* intracellular signaling molecules to affect phosphorylation and calcium flux within a cell. Disruption of this pathway is another way by which EDCs may alter endogenous hormone actions.

Thus, EDCs can act *via* multiple pathways and receptor-based mechanisms (Fig. 2). At higher doses they may also exert receptor-independent actions *via* more traditional mechanisms of toxicity. Their effects are species, tissue- and cell-specific, and are influenced by metabolism.

2. Testing for endocrine disruption

The complex biology of endocrine disruption means that *no single assay nor single approach* can be used to identify chemicals with EDC characteristics. Instead, a combination of approaches is necessary, including computational methods as well as both *in vitro* and *in vivo* testing. Compared to current practice, a carefully composed battery of assays can dramatically reduce the likelihood that a newly developed chemical will later be found to be an EDC.

In vitro methods can test for many types of EDC activity. Actual endocrine disruption, however, involves perturbing the action of one or more hormones within a whole organism. Today's *in vitro* and computer models do not incorporate the complexity that this involves. For this reason, *in vivo* assays will also be necessary.

Two additional characteristics of the endocrine system must inform a strategy to detect potential EDCs. First, like endogenous hormones, EDCs may display non-monotonic dose-response curves.[2,6] This means that effects observed at low dose levels may be completely unpredictable, and indeed the opposite, of effects observed at high levels. Multiple mechanisms underlie the non-monotonicity of endocrine systems. Thus it is critical to assess chemicals over a wide concentration range *in vitro* and wide dose range *in vivo* to determine whether they have EDC characteristics.

Second, the effects of an exposure to EDCs vary with the life stage in which it is experienced (Fig. 3). Thus, the consequences of exposure during periods of development (fetal, childhood and adolescence, including puberty) can vary among periods and may also yield very different effects compared to exposures in adulthood. While adult exposure to EDCs can certainly be an important factor in adverse health outcomes, key times in development are likely to be more sensitive to endocrine disruption.

Adverse effects during periods of developmental transition are likely to occur at concentrations of the chemical that are far below levels that would be considered harmful in the adult.[7,8] These vulnerable life stages, including fetal, childhood, and pubertal development, are of particular concern because it is during these stages that the individual is changing physiologically and morphologically. These periods of transition are marked by massive changes in the endocrine environment as the new phenotype (or body plan) is being developed.

This heightened sensitivity during developmental transitions results from multiple factors. Most important, the organizational activities of hormones (formation of organs, brain organization, *etc*.) are not reversible, whereas the activational activities (regulation of reproduction, immune system modulation, *etc*.) that prevail in adulthood are reversible. Second, the protective mechanisms available to the adult such as DNA repair mechanisms, a competent immune system, detoxifying enzymes, liver metabolism, excretion, and the blood/brain barrier are not fully functional in the fetus or newborn. Third, the developing organism has an increased metabolic rate as compared to an adult or aged organisms and this, in some cases, may result in increased or reduced toxicity.[8]

Lastly, any strategy designed to test for EDC activity must examine organisms during different developmental stages because the suite of endogenous hormones present during development vary from one stage to another. A developing organism may be at a stage when it would not normally be exposed to a certain hormone—and thus, exogenous exposure to an EDC that acts upon that hormone's receptor or signaling system will activate a pathway that should not be active at that life stage. Therefore, prenatal exposure to environmental factors can modify normal cellular and tissue development and function through developmental programming, such that the individual may have a higher risk of reproductive pathologies and metabolic and hormonal disorders later in life.

3. Tiered Protocol for Endocrine Disruption (TiPED)

3.a. Overview. We propose a five-tiered system, TiPED, to help chemists determine potential endocrine disrupting activity of a new chemical (Fig. 4). The Tiers are organized from the simplest and least expensive screens to a whole animal lifetime assessment with the goal of identifying chemicals with endocrine disrupting potential early in the synthetic and testing process.

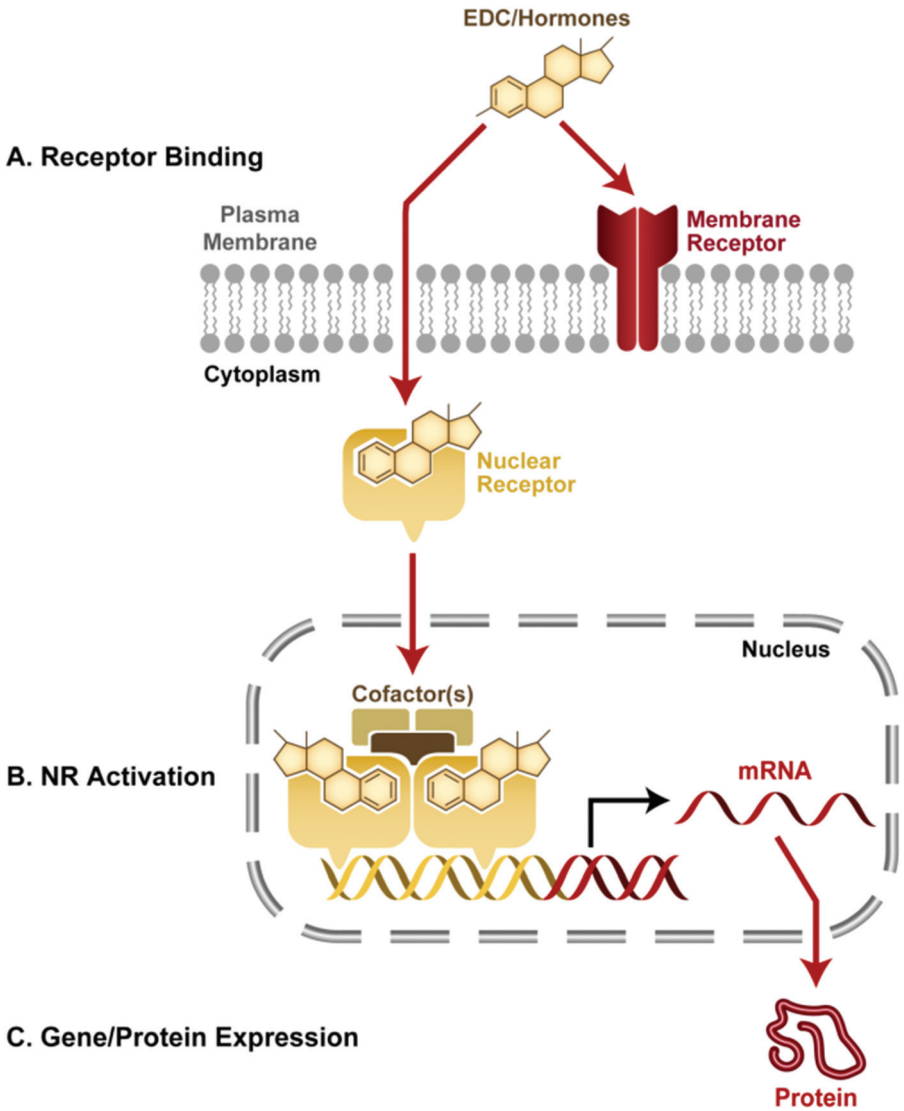

Fig. 2 This schematic depicts disruption of receptor signaling by an EDC, one of many possible ways that EDCs can interfere with endocrine system function. A. In this example, the EDC is a small lipophilic molecule, which can pass through the cell's plasma membrane and bind to a nuclear hormone receptor (NR). B. The NR is activated by EDC binding, and it translocates to the nucleus where the cell's transcriptional machinery, such as cofactors, are recruited to form a complex on the hormone response element of a hormone-responsive gene. C. The assembled complex promotes transcription of downstream DNA into RNA and eventually protein. Ultimately, gene and protein expression of hormone responsive genes may be influenced by EDC binding to nuclear hormone receptors.

While each Tier by itself will be informative, confidence about whether or not a compound alters endocrine function is increased by combining evidence from multiple tiers. Cost would suggest that inquiries begin with the first and second tiers, and if these tests prove negative then assays from higher tiers be used. This linear approach is the most logical and economic from a new chemical design perspective, but there may be reasons to start elsewhere. Ideally, multiple tests examining endpoints across taxa, encompassing different life stage effects and a range of doses, would be conducted.

3.b. TiPED Tier 1: computation-based assessments. A logical starting point for a chemist designing a chemical *de novo* would be to assess the physical and chemical properties of a molecule, such as density, boiling point, vapor pressure, refractive index, viscosity, surface tension, polarizability, partition coefficients, log *P*, *etc*.[9] Tier 1 encompasses an array of computational approaches that utilize statistical, computer and mathematical models to predict EDC properties of molecules. Early-stage identification of potential for endocrine disruption using *in silico* methods has the highly desirable advantages, compared to higher tiers, of speed of detection, lower cost, efficiency, avoidance of animal use and sustainable resource management.

Currently available computational-based assessments can be grouped into four distinct, complementary approaches:

• Chemical reactivity: these approaches are based upon the presence of a toxicophore, a specific chemical group within a larger molecule with identified toxic properties, a.k.a. toxicophores as defined by Williams (2002),[10] *e.g.* 1,3-benzodioxole group containing molecules in kava extract,[11] or azo-fragment (R–N═N–R′) in some dyes;

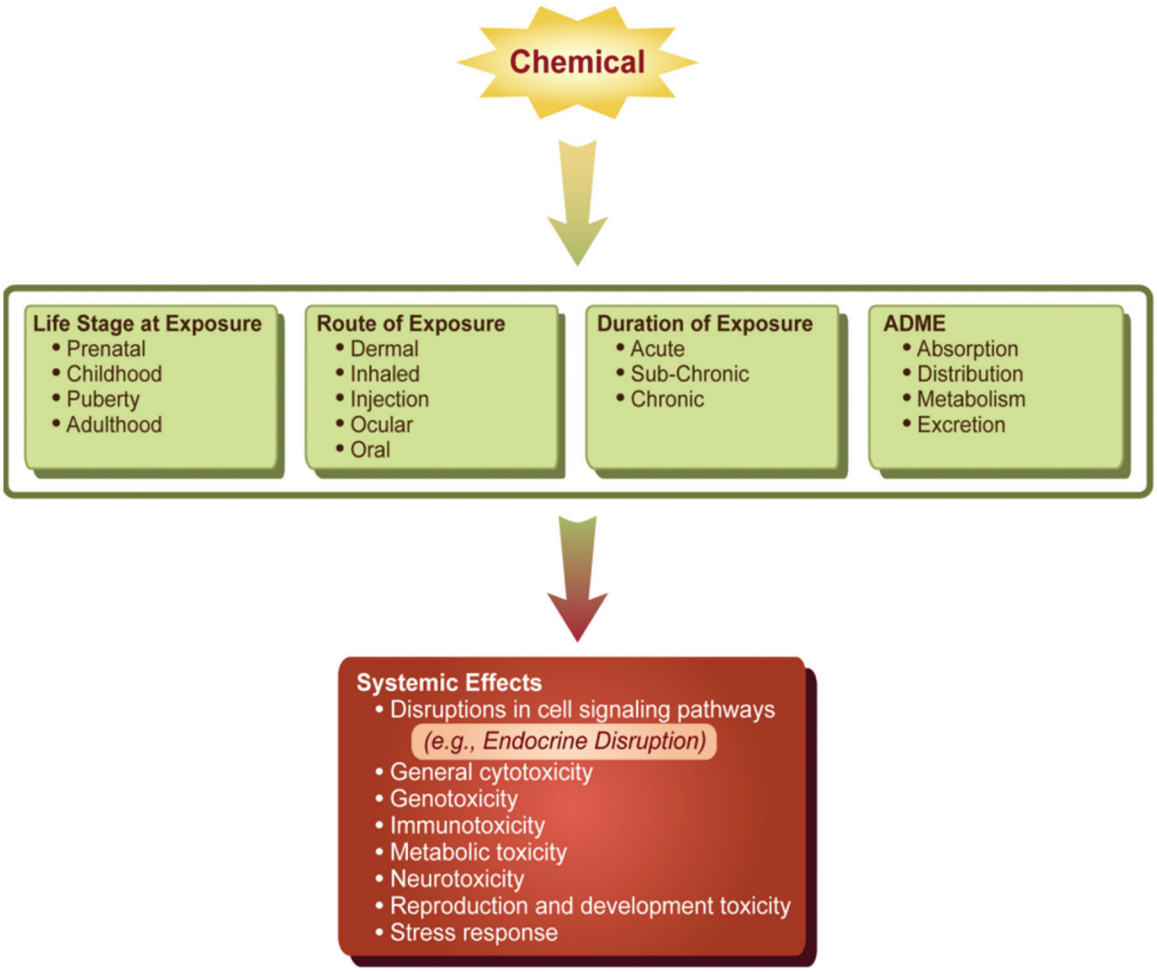

Fig. 3 Multiple factors contribute to a chemical's ultimate systemic effect on an organism, including age at exposure, route and duration of exposure, and metabolism of the chemical. There may be a period of latency following the exposure, such that effects of a chemical may not manifest until later in life.

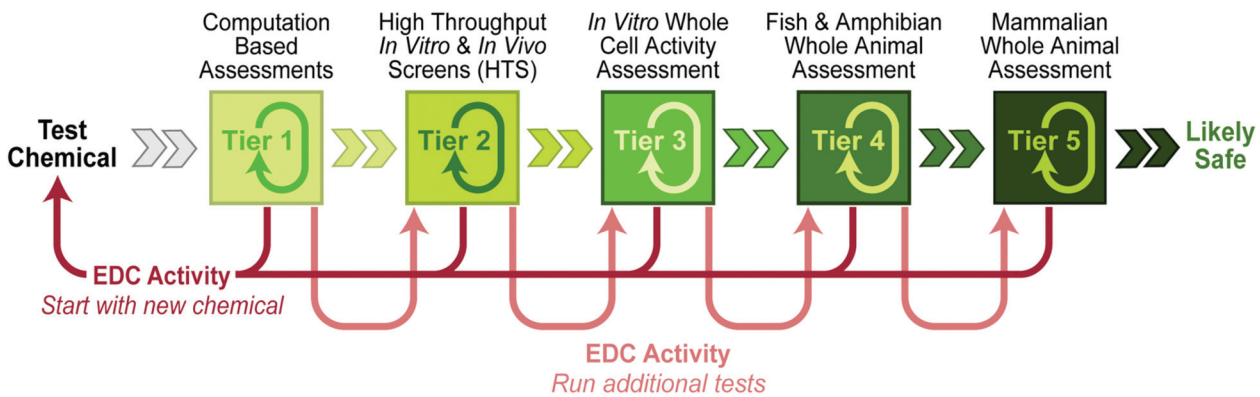

Fig. 4 Tiered tests for endocrine disruption. The progressive approach (A) to using this tiered system runs from left to right, from the simplest, fastest and cheapest on the left (Tier 1) to the most expensive on the right (Tier 5). Failure to find EDC activity in one tier then leads to testing at the next highest tier (after replication with other assays within the same tier). Chemists taking the plate approach (B) would begin at a tier that best fits their individual needs, with the choice reflecting prior knowledge (or hypotheses) about potential mechanisms of action, as well as their access to assay systems. Results from initial tests would then inform the next steps.

• Physico-chemical properties: statistical predictions of toxicity based on physico-chemical parameters, such as lipid solubility, octanol-water partition coefficient, log *P*, a hydrophobicity measure that correlates with ubiquitous interactions and certain elimination/activation pathways;

• Q/SAR: approaches based on the assumption that molecules with similar chemical structures will have similar biological activities;

• Modeling of biological activity: this approach uses a flexible 3D model of the novel molecule to predict whether it will fit within the binding pocket of a specific biomacromolecular target associated with an endocrine disruption pathway, *e.g.* a nuclear hormone receptor.

We recommend evaluating molecules with unknown characteristics through multiple computational assays because each method has distinct strengths and weaknesses. For the purposes of this paper, only Q/SAR and molecular docking will be discussed in detail (however, see Table S2:† "*Tools available for in-house computational-based assessments of EDC activity*" for a listing of other computational tools available on-line).

Quantitative/Structure Activity Analysis (Q/SAR). A series of papers in the mid-1960s laid the foundations for quantitative structure activity relationships (Q/SAR) by quantifying relationships between a chemical's biological activity and its physicochemical properties.[12] The Q/SAR approach utilizes statistical tools to generate predictive models of biological activity based on a number of descriptors unique to a chemical's molecular structure/properties (*i.e.* molecular weight, number of H-bond acceptors/donors, log *P*, solubility, *etc*.). The test chemical's structure and molecular properties are then compared to the same structures and properties of an experimental data set (a training set of well-characterized molecules where biological activities are well established). The aim is to quantify structural similarity to other chemicals with known biological activity, with the assumption that the untested molecule may possess the same biological activity by virtue of its structure. Since its introduction, Q/SAR has become a widely used tool to predict biological activity of chemicals, and a number of laboratories have applied this approach to predict the endocrine disrupting activity of environmental and pharmaceutical chemicals.[13,14]

Although it is potentially a useful statistical tool, obtaining a meaningful Q/SAR predictive model on toxicity is problematic, and depends on several factors, including the quality and availability of biological data, the statistical methods employed, and the choice of descriptors. A useful Q/SAR model would incorporate the following characteristics:

(1) Include a training set comprised of a sufficient number of molecules that cover the range of properties to be predicted by the model.

(2) The number of compounds in the training set should be far more numerous (at least 5 to 10 fold) than the number of non-correlated descriptors used to calculate the model. Furthermore, the descriptors should be biophysically relevant to the property being predicted.

(3) The model should be applicable to novel compounds and allow for mechanistic information related to the endpoint of interest.

(4) Preferably, the simplest model should be selected.

For the purposes here, a chemist should consider the following limitations of the Q/SAR approach when selecting a Tier 1 method to predict EDC potential:

• The "SAR Paradox", the fact that molecules of similar structure often have very dissimilar biological activity.[15]

• Each Q/SAR model predicts a specific endpoint, and only for chemicals with the identical mechanism.

• Q/SAR models do not perform well with chemical structures outside the training set.

• Most nuclear receptors have not been the focus of Q/SAR modeling, and there almost certainly are receptors yet to discover. Existing Q/SAR models predict only a subset of potential endocrine-activity and as such are insufficient.

• Q/SAR models do not predict whether the compound agonizes or antagonizes a receptor.

• Care must be taken to avoid deriving an over-fitted model (*e.g.* one that describes random error or noise, rather than an underlying relationship) and generating useless interpretations of structural/molecular data.

In sum, while Q/SAR models currently can be used as statistical tools for broad statements of probability they are not sufficient for predictive toxicology, especially for endocrine disruption; additional tools must be used to provide a fuller picture.

Modeling of biological activity (pocket modeling, molecular docking). The simplest way to think about a molecule and its receptor is to picture them as a lock and key, with a caveat that both of them are somewhat flexible. In a molecular docking model, the goal is to determine the correct orientation and adjustments of these two components. Specifically, molecular docking predicts the preferred orientation a molecule will adopt when bound to another molecule (*i.e.* the receptor) to form a stable complex. This information can be used to predict the binding affinity, or strength of association between the two molecules. Because the relative orientation of two molecules influences whether agonism or antagonism of the receptor results from their interaction, this method is useful for determining what type of signal a novel chemical is predicted to generate at the receptor. The limitation of this approach is that the molecular docking method requires an available crystal structure of the ligand-binding domain of interest, or at least of its close relative, as well as understanding of the domain's flexibility, and structures being altered by residence in different cellular locations, such as plasma membrane *vs.* aqueous compartments.

The main approach used by scientists that study molecular docking simulates the actual docking process, whereby the ligand moves into position within the receptor's active site following a series of rigid body transformations and internal changes to the ligand structure, such as torsion angle rotations, as well as changes in the binding pocket structure (Fig. 5).[11] Unlike simple comparisons of the complementarity of receptor and ligand shapes, simulation approaches can incorporate both ligand and receptor flexibility into the model, thus it is more reflective of what actually happens during ligand–receptor interactions. A disadvantage of this approach is that it is more time-consuming.

Molecular docking modeling tools have been developed in connection with pharmaceutical chemistry and are now being adapted to predict endocrine disruption potential. Initial studies have demonstrated the acute accuracy of the tool, *e.g.* accurately modeling the interaction of polybrominated diphenyl ethers (PBDEs) with the ER[16,17] and AR,[18] as well as preliminary studies of a panel of NRs with crystallographic structures.[19] Recent tests of PPARγ models demonstrate the very strong

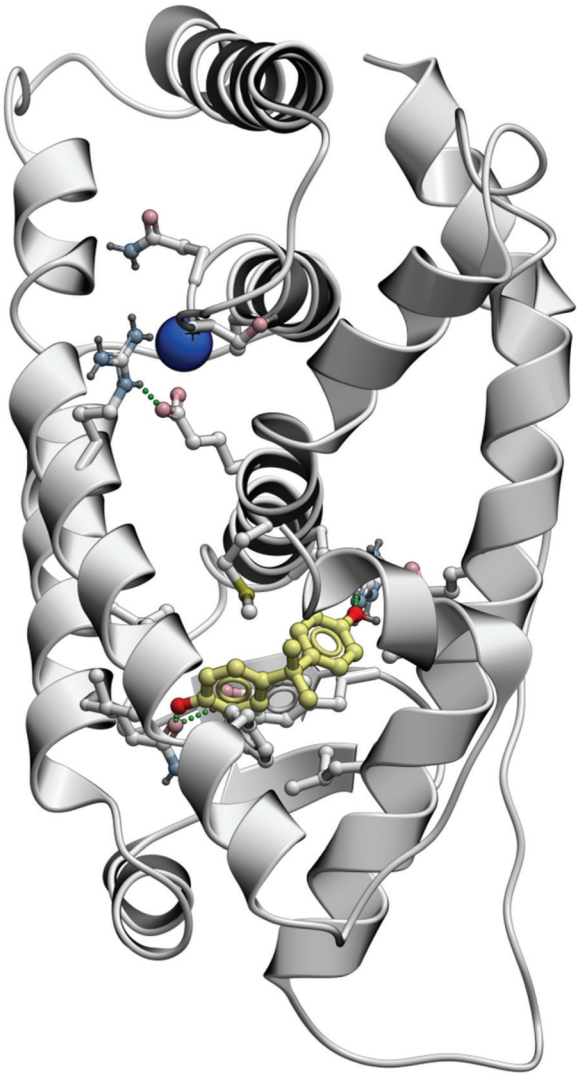

Fig. 5 This figure depicts the interaction of bisphenol A (BPA) with the estrogen receptor, at the ligand-binding domain.

(at close to 100% accuracy) discriminating ability of the docking models. As this particular tool is further developed and refined, its utility in predicting EDCs will become extremely valuable as part of Tier 1 in the TiPED toolbox.

3.c. Tier 2: high-throughput *in vitro* screens (HTS). HTS are now available using cell-based and cell-free methods. The two primary examples in the U.S. are TOXCAST at the U.S. E.P.A.;[20] and Tox21, is a joint effort U.S. E.P.A., National Institutes of Environmental Health Sciences/National Toxicology Program, National Institutes of Health and the Food and Drug Administration.[21] These screens were created to allow for rapid testing of many chemicals across many potential endpoints (see Table S3:† "*Receptors and other endpoints that can be assessed using Tier 2 high-throughput screening*"). Originally developed for use in drug discovery, they work well at detecting pharmacologically-active compounds with strong effects. Efforts underway at Tox21 have made significant progress to use these assays to identify compounds with weak activity, as well.

TiPED's use of HTS differs from that of the pharmaceutical industry in two ways. First, green chemists are likely to be interested in the potential for EDC activity among a small number of new chemicals, not hundreds or thousands that might be of interest in drug discovery. This is because the synthetic green chemist is usually not screening hundreds or thousands of existing compounds for effects, but instead is focused on a small number of newly synthesized molecules. Second, HTS were not designed initially to detect weak activities, even though those weaker signals may be biologically relevant and indicative of EDC activity. Hence care must be exercised in HTS use and interpretation.

With significant limitations discussed below, HTS offers the opportunity to test chemicals quickly to further explore Tier 1 findings for agonist or antagonist activity of identified molecular targets such as nuclear hormone receptors, cell surface receptors, cellular kinase signaling pathways, *etc.* Tier 2 therefore has two purposes and outcomes:

(i) HTS allows direct testing for the ability of the compounds to modulate biological signaling pathways important for endocrine disruption. For example, these screens test for estrogen, anti-androgen, anti-thyroid or obesogen activity.

(ii) HTS also informs the *in silico* screening in Tier 1, thereby allowing the models to be quickly and accurately refined to have better discriminative and predictive properties. This improves the suite of Tier 1 assays to minimize false positive and false negative results, allowing for continued development of Tier 1 assays.

Tier 2: increasingly, HTS assays represent rapid, sensitive and cost-effective strategies for identifying EDC activities, and as they are refined, they promise to allow large numbers of candidate chemicals to be tested for endocrine disrupting activities. An important refinement will be to identify the most predictive subset of assays required and this will be a natural consequence of early testing.

With respect to endocrine disruption, the simplest and most developed HTS assays measure the binding affinity of a chemical to NRs, provided the compound is sufficiently small (<1000 Da) and lipophilic. Examples of such assays include radioligand competition binding assays, whereby molecules compete with radio-labeled ligand for binding to the receptor, scintillation proximity binding assays, which measure the reaction of compounds with receptor-coated beads, and a fluorescence resonance energy transfer assays, in which a fluorescent signal is generated *via* the ligand-dependent interaction between the fluorescently-labeled ligand binding domain of a NR and co-activator proteins. HTS that assess relative binding affinities to many known human NRs can be done in approximately a week at a cost in the range of $10 000.

Several commercial labs can currently perform these screens, although attention to quality control is important. For example, an independent replication of PPARγ assays by a university laboratory specializing in this receptor found that of the 19 chemicals reported by ToxCast to be PPARγ activators, only 4 were bona fide activators while 3 were antagonists and the remainder inactive (Blumberg, unpublished data). Thus, currently available HTS can provide some useful but incomplete information concerning the likelihood of endocrine disruption by assessing the ability of a chemical to bind to known hormone receptor

systems. The advantage here is not only that the assays are fast and relatively inexpensive but also that one can get an indication of possible signaling pathways that might be disrupted by the chemical across many signaling pathways. A positive at this point might lead to rethinking of the chemical structure or using the information on possible signaling pathways to inform where to look in either Tier 3 or Tier 4.

The Tox21 program includes multiple endocrine signaling pathways using a titration-based format, and this quantitative HTS (qHTS) platform tests each chemical at multiple (7–15) concentrations, thus creating wide-ranging dose response profiles of compounds.[22] This system supports miniaturized cell-based assays in a 1536-well-plate format providing the throughput to test thousands of compounds at the same time in a single assay. For example, in a recent study, the Tox21 project screened ~2800 chemicals at 15 concentrations against a panel of 10 human NRs—the androgen receptor (AR), estrogen receptor α (ERα), farnesoid X receptor (FXR), glucocorticoid receptor (GR), liver X receptor β (LXRβ), peroxisome proliferator-activated receptors δ and γ (PPARδ and PPARγ), retinoid X receptor α (RXRα), thyroid hormone receptor β (TRβ), and vitamin D receptor (VDR)—in a qHTS format.[23–25] Data were used to generate concentration–response curves for every compound to identify both potential agonists and antagonists. The study reported better reliability for the agonist-mode than for the antagonist-mode assays, which was likely due to interference of cytotoxicity in the latter assays.[25] Overall, the results demonstrate the feasibility of using qHTS to quickly screen many compounds at many concentrations to test for potential endocrine activity.

Over the next few years Tox21 will be expanded to cover, to the extent possible, all known pathways and receptors involved in endocrine signaling. In addition, more overlapping endpoints and redundancy in the pathways assessed will provide internal validation. Data from Tox21 will be made publically available.

HTS methods, however, have limitations. Most importantly, *in vitro* HTS in Tier 2 typically do not assess an integrated, whole cell activity but only cell binding and pathway activation. They work in the 'known world' of endocrine signaling, leaving the chance that the molecule in question targets mechanisms that are currently unknown and therefore not targeted by HTS.

Further, HTS tools are unable to determine whether any metabolites of the chemical being tested have potential endocrine disrupting activities. This is a significant limitation of what can be learned with this approach and another reason why whole animal tests are a necessary component of testing for EDC activity. Tox21 is working to develop strategies to address this metabolism limitation (R. Tice, pers. comm.).

Another drawback to these types of HTS is that in many cases only partial receptors or fusion proteins are used, which may lead to false negatives. Tox21 includes assays for both full-length and particle receptors, at least initially for ER and AR; this may be expanded depending upon results obtained. Biological signaling pathways have cell type-specific requirements, and although cells used for HTS may provide sensitive readouts of receptor activity in a general context, the cells used in these assays are typically established cell lines that may not have the full suite of transcriptional machinery required for activity in a particular tissue of interest.

A positive finding from assays in Tier 2 that the chemical being tested demonstrates EDC activity presents the chemist with several choices. The simplest would be to abandon work with the chemical. A second option would be to perform additional and different Tier 2 assays that target the same mechanism, because of the possibility of false positives. A third choice would be to confirm the result with additional analysis that can help guide efforts to avoid the EDC effect through modifications to the molecule's structure.

No positive findings in Tier 2, in contrast, would be a strong signal that the chemical in question warrants additional examination in Tiers 3 to 5.

3.d. Tier 3: *in vitro* whole cell activity assessment. In Tier 3, we employ sensitive cell-based assays that are known to accurately reflect *in vivo* signaling and functional outputs in whole cell systems. Cell-based assay systems have the advantage of providing an integrated response stimulated by individual chemicals or mixtures of EDCs. This is a critical next step to assess whether a chemical can activate signaling pathways that lead to functional outputs such as cell division, differentiation, or cell death. These assays are generally more sophisticated than Tier 2 *in vitro* HTS and require more time, cost and experience to conduct. However, the targeted sensitivity of these assays will provide important biological validation on whether or not the compound of interest displays endocrine disrupting activity.

The primary advantage of Tier 3 testing is the ability to examine functional outputs resulting from receptor binding and pathway activations. Thus a positive result in Tier 3 is a strong indicator of EDC activity. Some of the assays we propose for Tier 3 are already used in Tier 2. This redundancy is necessary for two reasons: first, this reflects concerns about quality control in current HTS systems and the resulting frequency of false positives and negatives. Second, Tier 3 offers the opportunity to probe more deeply into specific biological mechanisms that may suggest to the chemist molecular modifications to eliminate EDC activity.

To screen for EDC activity, most whole cell assays fall into one of the following categories: cell proliferation assays, phospho-activation of regulators, enzyme or transporter assays, hormone secretion, or receptor-dependent gene or protein expression assays (see Table S4:† "*Examples of current assays, biological endpoints, and references*"). For example, cell proliferation assays are conducted in cell lines that respond to specific hormone exposure by increasing proliferation (for example rat pituitary cells or human breast cancer cell lines), and are sensitive to very low levels of EDCs. Many kinases, phosphatases, and transcription factors are regulated by rapid post-translational modifications such as phosphorylation, which can be monitored with the use of phospho-specific antibodies. Enzyme assays can be utilized to determine whether a compound alters or inhibits activity of important endocrine system enzymes (*e.g.* those involved in the synthesis, release, or metabolic degradation of steroid hormones). Hormone assays measure the concentration of a specific hormone in the cell media in which the cells are growing using either enzyme-linked immunoassay (ELISA) or radio-immunoassay (RIA). Receptor-dependent gene expression assays measure the ability of a test compound to stimulate receptor-dependent induction of reporter gene

expression in transient- or stably-transfected yeast or mammalian cells (with or without transfection of receptors). Examples of Tier 3 assays include the MCF-7 human breast cancer cell proliferation assay, BG-1 Lumi Cell ER assay, GH3 cell proliferation assay, rapid non-genomic protein activations, PPAR, yeast estrogen screen (YES), and arylhydrocarbon receptor (AhR) assays, among others (Table S4†). These options allow much more detailed testing for (anti-)estrogenic and (anti-)androgenic activity, as well as disruption of signaling through the thyroid receptor (TR), corticosteroid receptors (*e.g.*, mineralocorticoid, glucocorticoid), retinoic acid receptor (RAR), Retinoid X Receptor (RXR), Vitamin D Receptor (VDR), PPAR, and metabolic altering properties of chemicals.

The estrogen-dependent proliferation assay using MCF-7 human breast cancer cells is one of the best characterized estrogen-response assays.[26] If the assay finds cell proliferation, a second assay is run with both the test chemical and a known anti-estrogen. If the anti-estrogen suppresses the proliferation, this indicates that the proliferation is estrogen-mediated. If not, the proliferation is due to some other mechanism.

More extensive testing in this tier can be carried out in specialized cell systems designed to detect endocrine disrupting effects that target specific organ systems. For example, co-culture systems are available to test transcriptional endpoints using inflammation, neurotoxic, metabolic, pulmonary toxic, and reproductive development models.[27] Non-genomic signaling screening assays are available to detect downstream kinases and other second messengers. Tier 3 testing can also be used to assess the activity of metabolites of test chemicals by incubating the parent chemical in a liver cell preparation [S9 fraction, available commercially[28,29]] and then testing the metabolites or fractions in a cell culture system; it is possible that the parent compound is safe but a metabolite may be toxic (*i.e.*, the pesticide methoxychlor[30,31]).

A major disadvantage of Tier 3 assays is that they are generally specific for only one mechanism of action, necessitating the employment of a battery of screens to address all possible mechanisms of endocrine disruption. One approach to circumvent this limitation is to examine mitogen-activated kinase activations that are downstream integrative responses for many upstream signals. Receptor-selective antagonists can then be used to identify the specific mechanism of disruption. In addition, assays covering many potential modes of endocrine activity are not yet available.

A second major limitation arises from the fact that the endocrine system is an integrated system. By definition, this integration cannot be assessed using the assays in Tiers 1, 2, and 3. For instance, cell line-based assays do not provide information on sensitive developmental stages, and cannot take into consideration bioaccumulation, absorption, distribution, metabolism, and excretion, nor can they reveal tissue-specific effects if they rely on transfected receptors and response systems. This limitation makes it essential that chemists employ *in vivo* assays in Tiers 4 and 5.

3.e. Tier 4: fish and amphibian whole animal assessment

i. Fish. In Tier 4, we use whole animal assays with fish and amphibian model systems (see Table S5:† "*Whole fish and amphibian assays*"). *In vivo* assays allow for examinations of multiple endpoints for multiple hormones, and multiple mechanisms of action. New compounds thus can be screened without prior information on suspected activity or mechanisms of action. Because all of the possible sites of action and mechanisms of endocrine action are not tested for by assays in the first three tiers, Tier 4 is needed for any chemicals that "passed" these early screens.

The advantages of the screens discussed here are that they are conducted in a physiologically intact vertebrate system, and by testing effects on developmental morphology and locomotion during the earliest life stages, the probability of capturing an adverse event is markedly heightened. Large sample sizes are easy to accommodate with amphibians and fish. Additionally, stages of development are much shorter and access to embryos for manipulation and observation is easier because they are not in a womb. An important limitation is known metabolism differences between fish and amphibians and mammals that may render comparisons across the vertebrate classes difficult. In other words, a negative result in these tests will need to be confirmed in mammalian assays (Tier 5). Additionally, there are a limited number of non-mammalian animal models available that can easily assess a large number of endocrine endpoints. That said, there are robust vertebrate non-mammalian assays that provide opportunities for relatively rapid whole animal testing at significantly lower cost than experiments with mammals.

"Rapid developmental toxicity assays" (utilizing fathead minnow, medaka, and zebrafish) are now available. As a complement to the targeted *in vitro* assays included in Tiers 2 and 3, *in vivo* rapid developmental toxicity screens provide a quick and inexpensive method to detect adverse interactions between test chemicals and a vertebrate whole animal system. The developmental toxicity assays identify changes in morphology that reflect interference with normal development of the animal's body.

The primary advantage of assays using lower vertebrates is that the embryos develop rapidly and exercise their complete repertoire of gene expression and molecular signaling during the short transition from fertilization to organogenesis. During this window of development, there is a high probability of detecting an adverse interaction between an EDC and its molecular target that manifests as developmental delays or discrete morphological abnormalities including pericardial and yolk sac edemas, curved body axis, and eye, jaw, craniofacial, fin, and/or pigmentation defects.[32–37] The developing embryo can also be monitored for a series of cardiovascular[38] and behavioral [39] endpoints (see also Table S6:† "*Factors for consideration in fish EDC studies*").

In addition to the rapid developmental profile, fish embryos are transparent and develop externally (in contrast to mammalian *in utero* development), allowing for noninvasive microscopy techniques to resolve individual cells across many developmental stages. The model can therefore be used to monitor organogenesis and the impact of chemical exposure in live animals in real time. Because of their small size, embryos can be placed robotically in individual wells containing nano-to-microliter volumes of test solution. Using 96- or 384-well plates, multiple concentrations of different candidate compounds can be tested in tandem. Indeed, to facilitate higher throughput screening, machine vision systems have been developed that can rapidly conduct quantitative morphological and behavioral analyses of hundreds of fish in minutes. Thus while perhaps not equivalent

to "High Throughput Screening" (an ambiguous concept regarding volume or endpoints) in scope, these methods can test multiple chemicals simultaneously in what might be conceived of as "medium throughput" quantities.

As a complement to developmental toxicity assays, the endocrine disrupting potential of chemicals can also be assessed through partial and full life-cycle "reproduction assays" (using medaka, or fathead minnow). Partial life-cycle assays employ short-term exposure during critical windows of sensitivity (*i.e.* sexual differentiation, gonadal development, active reproduction), whereas full life-cycle assays initiate chronic exposure with newly fertilized eggs.

Reproduction tests are well established using this model system, and have been used to assess a number of chemicals suspected of having endocrine activity. In larval fish, sexual differentiation and gonadal development are especially vulnerable to disruption by endocrine active chemicals. Assays designed to exploit one or both of these critical developmental windows begin with embryos or juvenile fish that have not begun the process of sexual differentiation, and continue through developmental stages of known sensitivity. Endpoints typically evaluated with these assays include an assessment of gonadal development, vitellogenin concentration, and phenotypic sex relative to expected genotypic sex.

In adult fish, active reproduction represents a period of sensitivity to chemicals that target the hypothalamo-pituitary-gonadal (HPG) axis. Assays designed to exploit this window of susceptibility begin with reproductively mature animals that have a successful history of reproduction, and assess apical (whole organism) endpoints following short-term (typically 21 days) exposure to a chemical.

Several transgenic zebrafish lines have been engineered to detect direct transcriptional activation of specific endocrine signaling pathways in "reporter gene assays". Some rely on tissue-specific promoters that contain, for example, estrogen response elements upstream of the fluorescent reporter gene[40] while other strains harbor multimeric promoter elements that amplify and drive reporter expression in an unbiased fashion.[41] Medaka transgenic lines have also been established, some also indicating estrogen expression. While reporter gene activation is often easily quantified, antagonism or repression is more difficult to detect (though there are methods to overcome this limitation such as screening in the presence of an agonist). [42]

These assays specifically and rapidly detect aspects of endocrine disruption. Researchers are limited, however, by the types of reporter line available and should understand that a variety of disruptions to the system may be missed because transcriptional reporter-based models are not capable of detecting non-genomic signaling.[43,44]

Finally, there are fish assays for 'non-reproductive' endocrine endpoints. While fewer assays have been developed for such endpoints in fish, components of the endocrine system (*e.g.* thyroid axis, stress axis) must be considered to fully assess endocrine disrupting potential of novel test compounds. Because vital physiological processes under endocrine control could be disrupted by endocrine active substances, fish screening assays also consider processes regulated by atrial natriuretic peptide, growth hormones, melanin-concentrating hormones, prolactin, parathyroid, somatostatin, and vasotocin hormones, among others. A comprehensive approach using modern whole genome and proteome techniques can, in theory, be utilized to assess this array of potential endocrine targets. The TiPED website (www.TiPEDinfo.com) will track these developments and make them accessible as they become practical.

A few recent studies have reported effects on the hypothalamo-pituitary-thyroid and hypothalamo-pituitary-interrenal axis in zebrafish. Thyroid development and function in zebrafish has been extensively characterized (reviewed in[45]), and thyroid hormones are known to play an important role in maintenance of homeostasis, growth, metabolism, behavior, immune function, and in the transition from larval to juvenile developmental stage. The potential long-term effects of low-dose EDC exposures on thyroid function has not been studied in detail in fish models, nor have TR-binding or transactivation assays been included as endpoints for regulatory EDC screening. Nonetheless, we propose that the following endpoints be included as markers for interference with synthesis, regulation and action of thyroid hormones: thyroid hormone levels, expression of genes involved in the thyroid axis such as TR-alpha, TR-beta, TSH and those containing TRE elements, and thyroid tissue histology.

3.e. Tier 4: fish and amphibian whole animal assessment (continued)

ii. Amphibians. As non-amniotes (lacking egg shells or fetal membranes as embryos) without barriers to chemical contaminants, amphibians are highly susceptible to contaminant exposure and are thus exceptionally good indicators of environmental disturbance. Even as larvae and adults, their moist permeable skin provides easy access for chemical contaminants to cross. By definition, as amphibians (with both a terrestrial and aquatic life), they are exposed to and susceptible to perturbations in both the terrestrial environment and the aquatic environment. Because hormones and their mechanisms of action are similar across vertebrates, amphibian studies provide insight into effects across wildlife. In addition, amphibians can be readily assessed in field and large outdoor container experiments to address the effects of endocrine disruptors on animals in the wild.

Frogs are the preferred amphibian model because their clutch sizes are large (over 10 000 per female in some cases, contrasted with a few dozen in most salamanders. See Table S7:† "*Selecting species for amphibian assays*"). Husbandry is simpler for most frogs because the larvae are herbivorous compared to larval salamanders, which require small live food.

Frogs show responses to thyroid hormones, androgens, estrogens, and corticoids, and biological markers that can detect disturbances in all four of these hormone classes have been defined and developed. More specifically, disturbances in hormone synthesis, release, transport, receptor binding, activity, and degradation can be detected. Thus, a single *in vivo* amphibian test can detect disruption of multiple hormone targets *via* multiple mechanisms of action. Several species are available with different advantages and disadvantages for endocrine disruptor screening (Table S7†).

One major benefit of the amphibian model is the dependence of metamorphosis on proper signaling of the pituitary and thyroid. Compounds that inhibit metamorphosis could do so by interfering with any aspect of thyroid hormone synthesis, transport, receptor binding, action or degradation. Similarly,

compounds that stimulate or accelerate metamorphosis can act *via* multiple mechanisms. By using an *in vivo* model, chemicals that interfere with any of these aspects of thyroid hormone function can be detected by assessing limb emergence or tail reabsorption, in addition to monitoring thyroid hormone-regulated genes. Additional *in vitro* tail tip assays can also be used to measure direct effects of chemicals on thyroid hormone metabolism, binding, and action.

To screen chemicals for androgen agonist/antagonist activity, several assays are available. Because exogenous androgens can sex-reverse some species of amphibian larvae, by monitoring sex ratio, androgen mimics can be detected. In addition, several androgen-dependent secondary sex characters can be assayed including laryngeal size, gular pouch development, and breeding glands.[46,47] There are also several behavioral assays available to examine androgen dependent reproductive behavior and functional assays that examine fertility in males.[48]

Amphibians are also useful for screening chemicals that influence estrogen levels. Exogenous estrogens or compounds that induce estrogen synthesis can sex reverse some species, and therefore distortions in the sex ratio can be used to identify these compounds. Likewise secondary sex characteristics can be used to monitor estrogenic compounds.[49] Other markers include oviductal growth and vitellogenin expression.[47] Anti-estrogens reduce these same features and can similarly be detected using these markers.

Finally, because corticoids affect growth, osmoregulation, and immune function, among other aspects of amphibian development, the effects of corticoid agonists/antagonists are more complicated to assay. Often, compounds that interfere with corticoids do so by increasing or inhibiting corticoid synthesis, which is easily monitored *in vivo* along with the assays described above. Importantly, mechanisms other than corticoid agonism/antagonism or changes in corticoid synthesis could explain many of the effects described here.[50]

3.f. Tier 5: mammalian whole animal assessment. We assume here that the chemist employing TiPED has run his/her molecule through Tiers 1–4 without detecting EDC activity. Tier 1 will have determined that the chemical being developed, based on its structure, will not fit into a receptor binding pocket or possess structural characteristics or physico-chemical properties associated with toxicity. In this case it is likely, but not certain, that the molecule will not interact with receptors for which there are computational data. Lack of EDC activity in Tier 2 confirms that the chemical actually does not bind to any of the known receptors used in the *in vitro* binding assays. Tier 3 provided additional assurance that at the subcellular level, integrative genomic pathways (which contain both known and unknown mechanisms of action) were not disrupted. This tier also allowed for assessment of metabolism using human liver cells providing some insight into potential EDC activity by metabolites. Thus working a compound through the first three tiers without detecting EDC activity will strongly reduce the likelihood that the molecule is working as an EDC *via* well-studied mechanisms.

There remain, however, important risks of false negatives. These can be further reduced through *in vivo* experiments, which because they use whole animals, include EDC effects that work through mechanisms integrating different elements of the endocrine system. Importantly, these include developmental processes which when disrupted may not manifest adverse effects until much later in life. The other advantage of whole animal experiments is that when integrative endpoints are assayed—*i.e.*, endpoints whose proper development involves multiple components of the endocrine system—they allow discoveries of EDC activity without knowledge of mechanism, including currently unknown mechanisms.

There are differences in hormones and pathways between fish and amphibian and mammalian systems. Thus, to be confident the chemical has no endocrine activity or to assess a specific endocrine system in more detail it is essential to consider mammalian whole animal assessment. These are the assays of 'last resort,' which would only be used if work in prior tiers revealed no EDC activity.

Tier 5 involves testing in mammalian models, primarily rodents. Tier 5 is not designed to replace regulatory testing but to be a focused assessment of endpoints/tissues/diseases/pathways that may have been missed by earlier tiers because they lack the complexity of mammalian development. It can also be used to shed additional light on endocrine disrupting actions identified by earlier tiers.

The mammalian models are unique in their capacity to study *in utero* exposures that involve interactions between endocrine responses in the mother, placenta and embryo/fetus. Furthermore, certain behavioral repertoires can be studied in mammals that have greater biomedical relevance, such as mating and maternal behaviors, lactation, weaning, and complex adult sociosexual behaviors.

While in many ways conserved, mammalian physiological processes differ in some ways from those of lower vertebrates. For example, some mammalian hormones such as vasopressin and oxytocin have fish orthologs in vasotocin and isotocin, but these play very different roles between the species. Mammals also have a much more complex central nervous system than fish or amphibians, and the neurological and neuroendocrine effects of EDCs could be quite different. Finally, the mode of exposure differs, with fish swimming in water contaminated by endocrine disruptors, and by ingestion of contaminated organisms lower in the food chain. Mammalian exposure, in contrast, is most commonly to be *via* ingestion in adults, or by maternal-fetal or maternal-infant transfer, the former by placental transport and the latter *via* lactation. Skin absorption and inhalation are also possible routes of exposure, and each route of exposure has its own profile in terms of rate of metabolism of a chemical. Thus, a mammalian model is necessary to verify the lack of EDC properties necessary for extrapolation to humans.

Rodent assays are not high-throughput manner because of the labor-intensiveness of husbandry and breeding, determining birth outcomes (number of pups, sex ratio, pup qualities such as birth weight and other physical parameters), culling litters to a standardized size, and monitoring postnatal development, or adult functioning of a variety of organ systems. The ability to quantify behaviors in an unbiased manner is also very labor-intensive but a critical endpoint, as neurobiological effects of EDCs may be small but pervasive and biologically relevant. As an example, effects of EDCs on reproductive behavior have often been small on the individual, but from the population

perspective may be significant. Similarly, small losses in IQ may seemingly have little consequence for an individual but be highly important for society.[51]

The choice of endpoints in mammalian models is not an easy task in the case of a chemical that has passed Tiers 1–4 without revealing any indication of endocrine disrupting activity. If that molecule does not test as positive for estrogen, androgen, thyroid, and other tested hormonal signaling pathways, then it becomes difficult to predict what specific endpoints to evaluate in a mammal.

We incorporate two approaches in Tier 5, both of which test for impacts that result from developmental exposures (fetal, neonatal, pubertal) as the individual matures and ages. The first approach focuses on a general overview of physical health, development, somatic markers such as body weight and anogenital distance (a bioassay relevant to normal masculinization), as well as monitoring of serum for indications of adverse effects. In both sexes, timing of puberty can be assessed, and in females, estrous cycles should be monitored by vaginal smears. If animals are euthanized, organs should be examined and weighed, and snap-frozen for molecular assays in addition to conventional histopathologic examination. For example, depending upon other endpoints of interest, tissues should also be fixed for use in general histopathology as well as for immunohistochemistry for selected proteins. When this is done, we recommend that gross morphological analysis using conventionally stained tissue sections not be the only endpoint of analysis. Research has shown that the *phenotype* of specific cell types (*e.g.*, specific changes in gene or protein expression within certain cell populations) is often affected by EDCs that can go undetected with gross morphological analysis to identify gross pathological changes. Blood should be collected and serum/plasma banked for hormone assays such as estradiol, testosterone, progesterone, glucocorticoid, thyroid hormones, as well as other components of blood that provide information about overall health, such as the lipid profile and markers of inflammation and tissue damage *etc.*

The second approach explicitly acknowledges that we are testing for mechanisms of EDC activity not revealed by the earlier tiers. To do this we have selected measurements that reveal perturbations of integrative endpoints, *i.e.*, endpoints whose proper development is influenced by inputs from multiple components of the endocrine system, not just a single hormone or single hormone axis (*e.g.*, the hypothalamic-pituitary-adrenocortical axis). This approach assumes that if a chemical has unknown EDC activity then that will manifest through effects on the development of one or more integrative endpoints, even if the mechanism is not known.

The integrative endpoints we have selected include: neurobehavioral, brain morphology, mammary gland, prostate gland, insulin-glucose and body weight regulation, allergic responses in airways, and hormonally related cancers (prostate, mammary gland). Table S8† provides a description of some of these integrative endpoints. We also recommend that some testing is done as animals become senescent, although this adds considerable expense and would not be included in the first set of experiments. The concern is that there is evidence that some disease outcomes due to developmental exposure to EDCs are not expressed until mid-life, which in rodents is around 18 months old. [52] Work on the endocrine disruptors vinclozolin,[53] methoxychlor [54,55] and BPA [56] are associated with an early aging phenotype, which, depending on the model, can require maintaining animals beyond young adulthood.

It is clear that these rodent studies are laborious and costly. There are, however, several points where some high-throughput assays or where simple measurements can be built in. Some endpoints can be monitored in a longitudinal manner (*e.g.* body weight), while others such as serum hormone radioimmunoassays and RT-PCR analyses for gene activity can easily be measured in bulk assays. Importantly, many assays can be run concurrently with siblings from the same litters and some assays can be run prior to the use of an animal for a specific endpoint. For example, if animals are going to be killed at 9 months of age for detailed analyses of various systems, body weight, estrous cycle, fertility, metabolic data such as glucose tolerance, and other types of data can be collected up until the time of euthanasia.).

4. Principles guiding protocol development and use

Several principles have framed the development of TiPED and will continue to guide its development in the future. Table 1 summarizes the overarching principles guiding our design of the EDC testing protocol. We briefly elaborate on each of the principles below.

4.a. Overarching principles. The first principal comes from green chemistry. Green chemists design against hazard.[57] The earlier in the design process that hazard can be discovered, the more likely it is that downstream problems will be minimized, if not avoided entirely. This can have material benefits for the chemist and his/her company.

The second principle, on current scientific understanding, contrasts our protocol with standardized approaches used in regulatory toxicology. As noted above, the standardized assays upon which traditional toxicological approaches are based are often decades old. They rarely reflect the quality and modernity of assay tools used in scientific research funded by the National Institutes of Health, including the National Institute of Environmental Health Sciences. The old approaches are insensitive and largely incapable of dealing with EDCs. Ignoring current science would result in chemists producing yet another generation of hazardous chemicals.

That said, the assays we recommend have been chosen because multiple laboratories have successfully used them. They can require specialized knowledge and skills, but are not so

Table 1 Overarching principles guiding design of TiPED

- Chemical hazard must be considered at all stages of molecular design and synthesis.
- Assays used should reflect current scientific understanding, and the protocol should be reviewed regularly to incorporate new scientific discoveries and tools.
- The assays within each tier should span a comprehensive range of EDC mechanisms of action.
- While *in silico* and *in vitro* assays offer less costly starting points, *in vivo* assays are necessary to conclude that a chemical is unlikely to have EDC activity.

arcane that only a single, or small number of, laboratories, would be capable of implementing them. The second half of the principle acknowledges the fast pace of scientific discoveries in the field of endocrine disruption. New modes of action requiring new assays will certainly be discovered. Incorporating this evolving knowledge into the protocol is essential.

The third principle, a comprehensive range of EDC mechanisms, reflects the need to look for more than one or two EDC modes of action. This is because single chemicals can act through multiple mechanisms. The absence of action through one mechanism cannot be taken as evidence of no action through another mechanism. A case in point is BPA. It is an estrogen *via* both genomic and non-genomic pathways, an anti-androgen, a thyroid hormone antagonist, and a peroxisome proliferater-activated receptor (PPAR) agonist.

The fourth principle acknowledges that the current state of *in silico* and *in vitro* assays do not sufficiently incorporate the complexity of an endocrine system functioning in a living organism, and especially that of a developing organism.

4.b. Evaluating EDC assays. In all likelihood, the chemist him/herself will not be performing the assays, but instead will be working in partnership with environmental health scientists or with a contract laboratory. Because this is not the chemist's field, yet their research is dependent upon the test findings, it is important for the chemist to have some ability to gauge the quality and reliability of the work being done. This is especially the case for EDCs because of the complexity of the science.

With this in mind, we offer a set of principles that chemists can use to select and evaluate EDC assays (Table 2). Each principle is briefly elaborated upon below.

The first principle is designed to select assays that have proven reliable among different laboratories, to have well defined performance standards and to avoid assays that test for poorly defined endpoints and hence are open to arbitrary and variable interpretation.

The second principle should guide experimental design. Negative controls are essential to establish an effect. Positive controls are needed to demonstrate that the experimental system is appropriately sensitive and free of contamination or other confounding variables. The positive control must be used at an appropriate concentration or dose to demonstrate the sensitivity of the assay in terms of being capable to detect effects of low doses of EDCs. Prior use of insensitive strains of rodents in EDC tests without positive controls has led to significant confusion in the peer-reviewed literature. Controls must be run concurrently because of the potential for temporal variation in unintended contamination (*e.g.*, changes in composition of rodent chow from batch to batch or inadvertent contamination of lab).

The third principle acknowledges a fundamental feature of endocrine disruption, that high dose effects do not necessarily predict low dose effects, *i.e.*, non-monotonicity in the dose-response curve.

The fourth principle addresses another key feature of endocrine disruption, that developmental exposures can lead to effects that are initially subtle, for example changes in epigenetic programming, but ultimately highly adverse, *e.g.*, cancer in adulthood.

The fifth principle is designed to widen the reach of the assays beyond currently known mechanisms of endocrine disruption. We have identified several *in vivo* versions of high-throughput screening that do not assume the mechanism of EDC action but instead look broadly at developmental disorders following early life exposure in fish and amphibians.

4.c. Evaluating laboratories. As with the choice and evaluation of specific assays, assessment of laboratory practice performance and capabilities in experimental environmental health science is outside the expertise of most chemists. In Table 3 we list six important criteria that should be addressed explicitly in the choice of collaborators/contract laboratories.

The first criterion specifies that the laboratory must demonstrate it can replicate the appropriate performance of the assay(s) as carried out by other laboratories and that it is capable of repeatedly performing the assay successfully.

The second criterion specifies that the laboratory must be willing to share all relevant information about the laboratory and its methods and practices, as well as all relevant data on assay performance.

The third criterion focuses on animal husbandry practices by the laboratory. Poor animal husbandry is not only unethical, it introduces additional variability in the experiments that can mask effects, making it more difficult to confirm or reject EDC activity. The laboratory should share information about its husbandry practices and benchmark those against industry standards.

The fourth criterion stipulates that power analyses should be performed in preparation for the full assay. Power analysis is a statistical tool that provides guidance, based on preliminary data, on the sample size necessary to find a statistically significant result given the magnitude of the effect and the variance inherent in the data. Use of power analysis is especially important in

Table 2 General principles for selection and evaluation of EDC assays for Green Chemists

- Each assay should be reliable, relevant, meet performance standards and use well-defined endpoints.
- Experimental design should employ concurrent negative and positive controls and blanks to confirm that the experimental system is free from contamination and that it is appropriately sensitive.
- A dynamic testing range should be established, and testing should be carried out over the full range, including high and low doses.
- Some *in vivo* tests should be structured to reveal the consequences of developmental exposures on health and function later in life, through all life stages.
- Some *in vivo* tests should not assume knowledge of the mechanism/pathway of action.

Table 3 Criteria to guide evaluation of laboratories

A lab must:

- Demonstrate intra- and inter-laboratory repeatability.
- Demonstrate transparency in reporting.
- Utilize effective, safe husbandry practices; high mortality/morbidity rates in controls are unacceptable.
- Employ power analysis of preliminary results to design methods.
- Utilize standard protocols and solutions/reagents/cultures/etc., where they are available.
- Undergo external review and audit on regular basis comparable to NSF/NIH external reviews.

in vivo studies to ensure the sample size is large enough to detect an effect but not so large that an excessive number of animals are used.

The fifth criterion addresses replicability and reliability. Standard protocols, well established in the peer-reviewed literature, should be followed in carrying out the assays and variations in assay performance must be avoided. Use of standard solutions/reagents/cultures/*etc.*, will help avoid inadvertent contamination and unexpected biological variability.

The sixth criterion—external review and audit—will provide the chemist overall assurance of the laboratory's quality.

5. Using TiPED with known EDCs: verification of methodology

When we first started the process of developing this tiered approach to screening new chemicals, we identified several known EDCs (chemicals or classes of chemicals) that work through different mechanisms and are known to have widely different effects on exposed cells, animals or humans. Using these examples, we identified a repertoire of assays that we expected would be sufficient to detect known endocrine disrupting activities. To continue this thought-exercise, we then identified published studies that determined whether the TiPED assays described above (or similar ones) have been used successfully with these six known EDCs (Table S9†).

Clearly, some of these EDCs would be identified by several computational assays in Tier 1. BPA and phthalates, for example, have been tested with both Q/SAR and molecular docking assays, and both of these methods indicate that these chemicals bind to nuclear hormone receptors. Other EDCs, such as perchlorate and atrazine, would likely "pass" the first tier. Testing BPA further with TiPED, it would also be identified as an EDC in Tiers 2, 3, 4 and 5. Thus, a chemical like BPA, with mechanisms that span several NRs, would be easily identified by this tiered screening protocol. In contrast, perchlorate and atrazine might make it to Tiers 3 or 4 before they are identified as EDCs. Yet the proposed assays are clearly robust enough that these chemicals would not make it to market, providing supportive evidence that the TiPED screens will be sufficient to identify putative EDCs.

Currently we intend to place EDC test protocol in the public domain. The institutional home for it is still to be determined, but it will likely be either an academic or government institution. Wherever it is located the "home" for TiPED will be a place where detailed protocols for assays will be found along with lists of available online databases and tools. In addition there will be trained personnel to answer questions and provide general guidance and referral to labs that can contract to do specific assays. The design and creation of the protocol has been overseen by a Scientific Advisory Committee comprised of experts from both chemistry and biology (Table S1†). Future management of the protocol will also require oversight and regular reexamination of the assays in light of scientific advancement.

The latter point is critically important—in order to avoid submitting future chemical innovation to insensitive safety tests (or, worse, giving approval to chemicals that future scientists learn to be EDCs), the assays and tiers of the protocol must be reviewed and updated on a regular basis. The continuing role of the Scientific Advisory Committee will be essential to this process and will keep the protocol on the leading edge of EDC science.

Summary and conclusions

TiPED provides tools that can guide the development of inherently safer materials by avoiding chemicals likely to disrupt the endocrine system. Using the assays in the protocol early in the design process to detect potential EDCs, chemists can choose not to pursue development of a candidate chemical that has EDC characteristics. Alternatively, they can use mechanistic data from the assays to guide redesign of the chemical, with the goal of retaining desired material characteristics but avoiding action through identified EDC mechanisms.

In an effort to prevent novel EDCs from being produced in appreciable quantities, we focused solely on scientific issues to provide chemists with a set of guiding principles and tools that will enable them to stem production of chemicals with EDC potential. The goal of this ground-up approach, termed TiPED, is to identify hazard early in the design process using a systematic series of assays that build upon one another. We wish to emphasize that this tiered protocol was not designed as a one-size-fits-all tool. Depending upon their unique situation, a chemist may have good reason to start at any point within the protocol, not necessarily with Tier 1.

A positive test at any step in the process is an indication of potential endocrine disruptor activity and thus provides the chemist an opportunity to modify the chemical under development. The endocrine disruption screening assays comprised in each tier are based on the best and most up-to-date science; collectively TiPED is designed to cover all known aspects of endocrine disruption.

Beyond serving as a tool for chemists, this paper highlights the need for a transformation in the field of toxicology, advancing this science from an exclusively reactive, analytical one to include a significantly preemptive arm. For new chemicals, and perhaps old ones as well, toxicology, at least as it is associated with commercial chemicals, has to become much more a collaborative undertaking between the chemist and scientists who can evaluate toxicity in real time both theoretically and experimentally. It is our hope that collaborative efforts such as these, which lie at the interface of endocrine disruption and green chemistry, will help lead to a new generation of inherently safer chemicals.

Acknowledgements

We are grateful to the following scientists who participated in a two day retreat at Cavallo Point, California to help us develop and refine our thinking on this investigation: Paul Anastas, Evan Beach, Amy Cannon, Chad Ellis, Julie Haack, Andreas Kortenkamp, C.J. Li, Nicolas Olea, Sarah Vogel, Adelina Voutchkova and Wim Thielemans. The work was supported by grants from NIEHS (grant no. 1R13ES016981–01), the Cedar Tree Foundation, the Johnson Family Foundation, the Kendeda Fund, the Marisla Foundation, the John Merck Fund, and the Passport Foundation. We thank Julie Jones for outstanding logistical and organizing efforts without which this collaboration

would not have been possible. Raymond Tice (NIEHS/DNTP) provided constructive comments on the manuscript.

Author contributions: K.P.O. and J.P.M. organized and directed the project. B.B., T.C., D.C., P.D., A.G, L.G., J.H., K.T., J.W., C.W., F.V., R.Z., and J.P.M. served on the TiPED Scientific Advisory Committee, developed the project concept, and contributed to specific sections in the paper; R.A., T.E., T.H., H.P., T.T., and A.M. provided scientific and technical expertise to sections of the paper relevant to their expertise; T.S. composed a first draft of the paper. K.P.O, J.P.M, T.S., S.D., L.V., A.G., and B.B. led editing of the paper.

References

1 P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, New York, USA, 2000.
2 L. N. Vandenberg, *et al*., Hormones and endocrine-disrupting chemicals: low-dose effects and non-monotonic dose responses, *Endocr. Rev*, 2012, **33**(3), 1–78.
3 R. T. Zoeller, *et al*., Endocrine-disrupting chemicals and public health protection: a statement of principles from the endocrine society, *Endocrinology*, 2012, **25**, 25.
4 W. V. Welshons, *et al*., Large effects from small exposures. I. Mechanisms for endocrine-disrupting chemicals with estrogenic activity, *Environ. Health Perspect.*, 2003, **111**(8), 994–1006.
5 P. Alonso-Magdalena, *et al*., Bisphenol-A acts as a potent estrogen via non-classical estrogen triggered pathways, *Mol. Cell. Endocrinol.*, 2012, **355**(2), 201–207.
6 J. P. Myers, R. T. Zoeller and F. S. vom Saal, A clash of old and new scientific concepts in toxicity, with important implications for public health, *Environ. Health Perspect.*, 2009, **117**(11), 1652–1655.
7 R. R. Newbold, E. Padilla-Banks and W. N. Jefferson, Adverse effects of the model environmental estrogen diethylstilbestrol are transmitted to subsequent generations, *Endocrinology*, 2006, **147**(6 Suppl), S11–S17.
8 R. R. Newbold, *et al*., Developmental exposure to endocrine disruptors and the obesity epidemic, *Reprod. Toxicol.*, 2007, **23**(3), 290–296.
9 A. M. Voutchkova, T. G. Osimitz and P. T. Anastas, Toward a comprehensive molecular design framework for reduced hazard, *Chem. Rev.*, 2010, **110**(10), 5845–5882.
10 D. P. Williams and D. J. Naisbitt, Toxicophores: groups and metabolic routes associated with increased safety risk, *Curr. Opin. Drug Discovery Dev.*, 2002, **5**(1), 104–115.
11 G. Bottegoni, *et al*., Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking, *J. Med. Chem.*, 2009, **52**(2), 397–406.
12 C. Hansch and T. Fujita, Rho-sigma-pi analysis. A method for the correlation of biological activity and chemical structure, *J. Am. Chem. Soc.*, 1964, **86**, 1616–1626.
13 E. Papa, S. Kovarich and P. Gramatica, QSAR modeling and prediction of the endocrine-disrupting potencies of brominated flame retardants, *Chem. Res. Toxicol.*, 2010, **23**(5), 946–954.
14 L. M. Shi, *et al*., QSAR models using a large diverse set of estrogens, *J. Chem. Inf. Comput. Sci.*, 2001, **41**(1), 186–195.
15 J. H. van Drie, in *Computational Medicinal Chemistry for Drug Discovery*, ed. P. Bultinck, *et al*., CRC Press, New York, USA, 2005.
16 W. Yang, *et al*., Insights into the structural and conformational requirements of polybrominated diphenyl ethers and metabolites as potential estrogens based on molecular docking, *Chemosphere*, 2011, **84**(3), 328–335.
17 W. H. Yang, *et al*., Exploring the binding features of polybrominated diphenyl ethers as estrogen receptor antagonists: docking studies, *SAR QSAR Environ. Res.*, 2010, **21**(3–4), 351–367.
18 W. Yang, *et al*., Anti-androgen activity of polybrominated diphenyl ethers determined by comparative molecular similarity indices and molecular docking, *Chemosphere*, 2009, **75**(9), 1159–1164.
19 S. J. Park, I. Kufareva and R. Abagyan, Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles, *J. Comput. Aided Mol. Des.*, 2010, **24**(5), 459–471.
20 D. J. Dix, *et al*., The TOXCAST Program for prioritizing toxicity testing of environmental chemicals, *Toxicol. Sci.*, 2007, **95**(1), 5–12.
21 R. J. Kavlock, C. P. Austin and R. R. Tice, Toxicity testing in the 21st century: implications for human health risk assessment, *Risk Anal.*, 2009, **29**(4), 485–487; discussion 492–7.
22 N. Malo, *et al*., Experimental design and statistical methods for improved hit detection in high-throughput screening, *J. Biomol. Screen.*, 2010, **15**(8), 990–1000.
23 M. Xia, *et al*., Identification of compounds that potentiate CREB signaling as possible enhancers of long-term memory, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(7), 2412–2417.
24 J. Inglese, *et al*., Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(31), 11473–8.
25 R. Huang, *et al*., Chemical genomics profiling of environmental chemical modulation of human nuclear receptors, *Environ. Health Perspect.*, 2011, **119**(8), 1142–1148.
26 W. V. Welshons, *et al*., Large effects from small exposures: I. Mechanisms for endocrine-disrupting chemicals with estrogenic activity, *Environ. Health Perspect.*, 2003, **111**, 994–1006.
27 F. Zucco, *et al*., Toxicology investigations with cell culture systems: 20 years after, *Toxicol. In Vitro*, 2004, **18**(2), 153–163.
28 K. Okud, *et al*., Novel pathway of metabolic activation of bisphenol A-related compounds for estrogenic activity, *Drug Metab. Dispos.*, 2011, **39**(9), 1696–1703.
29 C. H. Yang, *et al*., Sulfation of selected mono-hydroxyflavones by sulfotransferases *in vitro*: a species and gender comparison, *J. Pharm. Pharmacol.*, 2011, **63**(7), 967–970.
30 G. D. Charles, *et al*., Incorporation of S-9 activation into an ER-alpha transactivation assay, *Reprod. Toxicol.*, 2000, **14**(3), 207–216.
31 K. P. Miller, R. K. Gupta and J. A. Flaws, Methoxychlor metabolites may cause ovarian toxicity through estrogen-regulated pathways, *Toxicol. Sci.*, 2006, **93**(1), 180–188.
32 M. Shi and E. M. Faustman, Development and characterization of a morphological scoring system for medaka (*Oryzias latipes*) embryo culture, *Aquat. Toxicol.*, 1989, **15**(2), 127–140.
33 M. W. Carney, *et al*., Differential developmental toxicity of naphthoic acid isomers in medaka (*Oryzias latipes*) embryos, *Mar. Pollut. Bull.*, 2008, **57**(6–12), 255–266.
34 G. D. Marty, *et al*., Age-dependent changes in toxicity of N-nitroso compounds to Japanese medaka (*Oryzias latipes*) embryos, *Aquat. Toxicol.*, 1990, **17**(1), 45–62.
35 G. Laban, *et al*., The effects of silver nanoparticles on fathead minnow (*Pimephales promelas*) embryos, *Ecotoxicology*, 2010, **19**(1), 185–195.
36 K. Herrmann, Teratogenic effects of retinoic acid and related substances on the early development of the zebrafish (*Brachydanio rerio*) as assessed by a novel scoring system, *Toxicol. in Vitro*, 1995, **9**(3), 267–283.
37 M. Strmac and T. Braunbeck, Effects of triphenyltin acetate on survival, hatching success, and liver ultrastructure of early life stages of zebrafish (*Danio rerio*), *Ecotoxicol. Environ. Saf.*, 1999, **44**(1), 25–39.
38 P. V. Asharani, *et al*., Comparison of the toxicity of silver, gold and platinum nanoparticles in developing zebrafish embryos, *Nanotoxicology*, 2011, **5**(1), 43–54.
39 J. Chen, *et al*., Trimethyltin chloride (TMT) neurobehavioral toxicity in embryonic zebrafish, *Neurotoxicol. Teratol*, 2011, 721–726.
40 H. Chen, *et al*., Generation of a fluorescent transgenic zebrafish for detection of environmental estrogens, *Aquat. Toxicol.*, 2010, **96**(1), 53–61.
41 D. A. Gorelick and M. E. Halpern, Visualization of estrogen receptor transcriptional activation in zebrafish, *Endocrinology*, 2011, **152**(7), 2690–2703.
42 X. Terrien, *et al*., Generation of fluorescent zebrafish to study endocrine disruption and potential cross-talk between thyroid hormone and corticosteroids, *Aquat. Toxicol.*, 2011, **105**(1–2), 13–20.
43 P. Damdimopoulou and E. Treuter, Reporter zebrafish: endocrine disruption meets estrogen signaling, *Endocrinology*, 2011, **152**(7), 2542–2545.
44 E. Silva, A. Kabil and A. Kortenkamp, Cross-talk between non-genomic and genomic signalling pathways–distinct effect profiles of environmental estrogens, *Toxicol. Appl. Pharmacol.*, 2010, **245**(2), 160–170.
45 P. Porazzi, *et al*., Thyroid gland development and function in the zebrafish model, *Mol. Cell. Endocrinol.*, 2009, **312**(1–2), 14–23.
46 T. B. Hayes, *et al*., Atrazine induces complete feminization and chemical castration in male African clawed frogs (*Xenopus laevis*), *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**(10), 4612–4617.
47 J. H. van Wyk, E. J. Pool and A. J. Leslie, The effects of anti-androgenic and estrogenic disrupting contaminants on breeding gland (nuptial pad) morphology, plasma testosterone levels, and plasma vitellogenin levels in

male *Xenopus laevis* (African clawed frog), *Arch. Environ. Contam. Toxicol.*, 2003, **44**(2), 247–256.
48 T. Behrends, *et al*., Mate calling behavior of male South African clawed frogs (*Xenopus laevis*) is suppressed by the antiandrogenic endocrine disrupting compound flutamide, *Gen. Comp. Endocrinol.*, 2010, **168**(2), 269–274.
49 T. B. Hayes and K. P. Menendez, The effect of sex steroids on primary and secondary sex differentiation in the sexually dichromatic reedfrog (*Hyperolius argus*: Hyperolidae) from the Arabuko Sokoke Forest of Kenya, *Gen. Comp. Endocrinol.*, 1999, **115**(2), 188–199.
50 T. B. Hayes, *et al*., Pesticide mixtures, endocrine disruption, and amphibian declines: are we underestimating the impact?, *Environ. Health Perspect.*, 2006, **114**(Suppl 1), 40–50.
51 B. Weiss, Endocrine disruptors and sexually dimorphic behaviors: a question of heads and tails, *Neurotoxicology*, 1997, **18**(2), 581–586.
52 R. R. Newbold, *et al*., Uterine adenocarcinoma in mice treated neonatally with genistein, *Cancer Res.*, 2001, **61**(11), 4325–4328.
53 M. D. Anway, *et al*., Epigenetic transgenerational actions of endocrine disruptors and male fertility, *Science*, 2005, **308**(5727), 1466–1469.
54 A. E. Armenti, *et al*., Developmental methoxychlor exposure affects multiple reproductive parameters and ovarian folliculogenesis and gene expression in adult rats, *Toxicol. Appl. Pharmacol.*, 2008, **233**(2), 286–296.
55 A. C. Gore, *et al*., Early life exposure to endocrine-disrupting chemicals causes lifelong molecular reprogramming of the hypothalamus and premature reproductive aging, *Mol. Endocrinol.*, 2011, 2157–2168.
56 S. M. Ho, *et al*., Developmental exposure to estradiol and bisphenol A increases susceptibility to prostate carcinogenesis and epigenetically regulates phosphodiesterase type 4 variant 4, *Cancer Res.*, 2006, **66**(11), 5624–5632.
57 N. D. Anastas and J. C. Warner, The incorporation of hazard reduction as a chemical design criterion in green chemistry, *Chem. Health Safety*, 2005, **12**(2), 9–13.