

Special Series

Current Limitations and Recommendations to Improve Testing for the Environmental Assessment of Endocrine Active Substances

Katherine K Coady,*† Ronald C Biever,‡ Nancy D Denslow,§ Melanie Gross,|| Patrick D Guiney,# Henrik Holbech,†† Natalie K Karouna-Renier,‡‡ Ioanna Katsiadaki,§§ Hank Krueger,|||| Steven L Levine,## Gerd Maack,††† Mike Williams,‡‡‡ Jeffrey C Wolf,§§§ and Gerald T Ankley|||||

†The Dow Chemical Company, Toxicology and Environmental Research and Consulting, Midland, Michigan, USA

‡Smithers Viscient Laboratories, Wareham, Massachusetts, USA

§Department of Physiological Sciences and Center for Environmental and Human Toxicology, University of Florida, Gainesville, Florida, USA

||wca, Faringdon, United Kingdom

#Molecular and Environmental Toxicology Center, University of Wisconsin, Madison, Madison, Wisconsin, USA

††Department of Biology, University of Southern Denmark, Odense M, Denmark

‡‡USGS Patuxent Wildlife Research Center, Beltsville, Maryland

§§Centre for Environment Fisheries and Aquaculture Science, Dorset, United Kingdom

||||Wildlife International, Division of EAG Laboratories, Easton, Maryland, USA

##Global Regulatory Sciences, Monsanto Company, St Louis, Missouri, USA

†††German Environment Agency, Dessau-Roßlau, Germany

‡‡‡CSIRO Land and Water, South Australia, Australia

§§§Experimental Pathology Laboratories, Sterling, Virginia, USA

|||||US Environmental Protection Agency, Duluth, Minnesota

EDITORS NOTE:

This is one of 1 of 5 articles generated from the SETAC Pellston Workshop “Ecotoxicological Hazard and Risk Assessment Approaches for Endocrine-Active Substances (EHRA)” (February 2016, Pensacola, Florida, USA). The primary aim of the workshop was to provide objective advice, based on current scientific understanding, to regulators and policy makers, whether in industry, government, or academia. The goal is to make considered, informed decisions on whether to select an ecotoxicological hazard- or risk-based approach for regulating a given endocrine disrupting substance under evaluation.

ABSTRACT

In the present study, existing regulatory frameworks and test systems for assessing potential endocrine active chemicals are described, and associated challenges are discussed, along with proposed approaches to address these challenges. Regulatory frameworks vary somewhat across geographies, but all basically evaluate whether a chemical possesses endocrine activity and whether this activity can result in adverse outcomes either to humans or to the environment. Current test systems include *in silico*, *in vitro*, and *in vivo* techniques focused on detecting potential endocrine activity, and *in vivo* tests that collect apical data to detect possible adverse effects. These test systems are currently designed to robustly assess endocrine activity and/or adverse effects in the estrogen, androgen, and thyroid hormone signaling pathways; however, there are some limitations of current test systems for evaluating endocrine hazard and risk. These limitations include a lack of certainty regarding: 1) adequately sensitive species and life stages; 2) mechanistic endpoints that are diagnostic for endocrine pathways of concern; and 3) the linkage between mechanistic responses and apical, adverse outcomes. Furthermore, some existing test methods are resource intensive with regard to time, cost, and use of animals. However, based on recent experiences, there are opportunities to improve approaches to and guidance for existing test methods and to reduce uncertainty. For example, *in vitro* high-throughput screening could be used to prioritize chemicals for testing and provide insights as to the most appropriate assays for characterizing hazard and risk. Other recommendations include adding endpoints for elucidating connections between mechanistic effects and adverse outcomes, identifying potentially sensitive taxa for which test methods currently do not exist, and addressing key endocrine pathways of possible concern in addition to those associated with estrogen, androgen, and thyroid signaling. *Integr Environ Assess Manag* 2017;13:302–316. © 2016 The Authors. *Integrated Environmental Assessment and Management* published by Wiley Periodicals, Inc. on behalf of Society of Environmental Toxicology & Chemistry (SETAC)

Keywords: Risk and hazard assessment Endocrine disruption High-throughput assays Regulatory tests

This article includes online-only Supplemental Data.

* Address correspondence to kcoady@dow.com

Published 28 October 2016 on wileyonlinelibrary.com/journal/ieam.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

INTRODUCTION

Since the late 1990s, individual countries and international organizations, including Japan, the United States, the European Union, and the Organisation for Economic Cooperation and Development (OECD), have initiated

programs for assessing the potential impacts of endocrine active substances (EASs) to human health and wildlife. Although these programs may have originally been developed independently, their parallel aims have often resulted in harmonization of test methods (OECD 2010a).

The situation in the United States is broadly illustrative of the evolution of EAS screening and testing. In 1996, in response to an increasing number of publications, public pressure, and media focus, the US Congress mandated that the US Environmental Protection Agency (USEPA) develop a screening program to “determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect” (21 U.S.C. § 346a (p)(1-7)). The initial mandate for assessing estrogen (E)-related effects on humans was quickly expanded to include wildlife and the androgen (A) and thyroid (T) signaling pathways. As a result of this legislation, the USEPA created the Endocrine Disruptor Screening Program (EDSP), which utilizes a 2-tiered framework, with the 1st tier assessing the potential for a substance to interact with EAT pathways in vertebrates. Tier 1 includes a battery of 11 *in vitro* and *in vivo* assays that are interpreted using a weight-of-evidence approach to determine the potential for endocrine activity of a test substance. In Tier 2 testing, adverse effects and dose–response relationships of compounds identified as potentially active in Tier 1 are determined using additional mammalian and nonmammalian animal models. The USEPA has also been moving toward computational models and *in vitro* high-throughput screening (HTS) assays to help prioritize and screen chemicals for endocrine activity. For more information on the USEPA’s Endocrine Disruptor Screening Program in the 21st century, please visit <https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-21st-century>.

Concurrent with the initiation of the EDSP program in the United States, the OECD established a Special Activity on Endocrine Disrupter Testing and Assessment to coordinate the development of test guidelines to detect endocrine disruptors and to harmonize hazard and risk characterization approaches (OECD 2010a). This initial effort culminated in the release in 2002 (updated in 2012) of the Conceptual Framework (CF) for the Testing and Assessment of Endocrine Disrupting Chemicals, which outlined a 5-level categorization (OECD 2012a). Recently, the European Commission issued a draft proposal for the identification of endocrine disrupting chemicals, which are criteria that are intended to apply horizontally to various chemical regulations within the European Union. For more information on the European Commission’s Endocrine Disruptors Policy, please visit http://ec.europa.eu/health/endocrine_disruptors/policy/index_en.htm.

In 1998, the Ministry of Environment in Japan initiated their Strategic Programs on Environmental Endocrine Disruptors, which outlined policies and measures for studying and assessing risks associated with EAS perturbation (<https://www.env.go.jp/en/chemi/ed.html>). This program was revised in 2000; then in 2005, the Ministry of Environment released a new framework, “Perspectives on Endocrine Disrupting

Effects of Substances—ExTEND2005,” which promoted the conduct of fundamental studies on endocrine disruption and development of new test methods in cooperation with OECD and other international groups. ExTEND2005 (Enhanced Task on Endocrine Disruption) was replaced in 2010 by EXTEND (Extended Tasks on Endocrine Disruption) 2010, to accelerate the establishment and implementation of assessment methods and to identify priority chemicals and areas of research. In contrast with the EDSP in the United States, neither the Japanese programs nor the OECD CF are prescriptive in the context of regulatory activity, but, rather, they outline information and tools or research for evaluating the hazards of EASs and informing risk assessment.

In the time since the US, OECD, European Union, and Japanese programs were first created, availability of test systems for assessing potential endocrine activity and disruption has substantially increased (e.g., see <http://www.oecd.org/chemicalsafety/testing/seriesontestingandassessmentteststingforendocrinedisrupters.htm> for a list of OECD guidance documents for assessing endocrine disruptors). Although these newer (and still emerging) methodologies often represent a significant advancement and improvement over historic toxicity test methods in the context of EAS assessment, they nonetheless have potential shortcomings, including: (1) large resource requirements; (2) uncertainty relating to data interpretation in the context of linking mechanistic to apical data, accounting for non–endocrine-related toxicity, among others; (3) uncertainty in design aspects relative to the selection of most relevant species, strains, life stages, and endpoints; and (4) limitations in available technical expertise and training needed to conduct the tests. Later we examine the current test systems for assessing EASs, illuminate potential limitations and weaknesses associated with these test systems, and propose improvements and future research that would move the global evaluation of EASs forward, to better inform the assessment of both hazard and risk associated with these compounds.

The emphasis of the Pellston workshop that served as the genesis of the present study was on ecological hazard and risk assessment, which generally is focused on adverse effects at the population level, often for a large number of species. Although this differs from the individual-level emphasis in human health assessments, it is important to recognize that screening and testing programs for EASs are inherently integrative across, at least, vertebrate species. For example, many of the *in silico* and *in vitro* systems used for hazard assessment are based on mammalian (primarily human and rodent) data or test systems, or both. Further, some of the *in vivo* nonmammalian assays used for screening for endocrine activity (e.g., in EDSP Tier 1) also are intended to serve as “triggers” for possible higher-tier testing in mammalian systems (USEPA 1998). This is justified by the strong phylogenetic conservation of the hypothalamic–pituitary–gonadal/thyroidal (HPG/T) axes across vertebrate species. In any instance, it is not feasible to consider ecological hazard and risk assessment for EASs in isolation from efforts focused on human health.

CURRENT TEST SYSTEMS FOR ASSESSING ENDOCRINE ACTIVITY

In silico methods

A practical 1st step for examining a substance's potential for endocrine activity, especially for substances with little toxicity or mode of action data, is to collect available information based on reliable quantitative structure–activity relationships (QSARs) and read-across approaches from chemical analogues. An example tool for this type of an assessment is the Estrogen Receptor (ER) Expert System used by the EDSP, which is a database compiled into a decision tree for determining ER binding potential (http://archive.epa.gov/med/med_archive_03/web/html/er.html). The publicly available Oasis LMC QSAR (<http://oasis-lmc.org/>) contains ER and androgen receptor (AR) binding affinity QSARs, as well as models to predict chemical effects on cytochrome P450 aromatase (CYP19), an enzyme responsible for a key step in the biosynthesis of steroid hormones. The OECD Toolbox, which is publicly available freeware developed with the scientific and financial assistance of OECD and the European Union, is also an *in silico* QSAR platform for filling gaps in toxicity data and grouping substances into chemical categories (<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>). The Danish QSAR Database (<http://qsar.food.dtu.dk>) is another example of publicly available freeware that includes estimates from more than 200 QSARs, including more than 600 000 substances and models for ER, AR, thyroid receptor, and pregnane X receptor binding. These QSAR-*in silico* tools are useful for preliminary data collection steps to guide subsequent *in vitro* and *in vivo* screening and testing for potential endocrine activity. The OECD provides general guidance on the validation principles and potential regulatory use of QSAR tools (OECD 69; Supplemental Table S1).

In vitro assays

Currently, the EDSP Tier 1 battery includes a total of 5 *in vitro* assays that assess ER and AR competitive binding, ER transcriptional activation, and effects on steroidogenesis and aromatase. The OECD CF for Level 2 suggests the use of the same assays, but also provides a basis for expanding the number and types of *in vitro* systems that could be considered as they are developed and validated (OECD 2012a). A complete description of the various *in vitro* EDSP assays, as well as suggestions pertaining to their staging and conduct, is provided by Borgert et al. (2011). In addition, Scholz et al. (2013) provide an overview of several *in vitro* tests for the assessment of EASs in fish and amphibians.

The USEPA ToxCast™ program uses a suite of more than 700 HTS *in vitro* assays that cover approximately 300 signaling pathways and provide the basis for a framework to assist in ranking and prioritizing chemicals for future testing (<http://www.epa.gov/chemical-research/toxicity-forecasting>). A subset of the ToxCast assays evaluates potential impacts on EAT signaling pathways. The EPA EDSP recently proposed that this

subset could be used to identify chemicals that display specific endocrine activities of concern and that this information, in conjunction with rapid exposure assessment techniques, could be used for prioritization of chemicals for Tier 1 testing. A limited practical demonstration of this approach that uses approximately 20 HTS assays focused on chemicals that interact with mammalian (primarily human) ER isoforms has been described, and recently it was reviewed by an external advisory group to the EDSP (Browne et al. 2015; Federal Register Notice 2015).

In vivo methods

Some *in vivo* assays (e.g., those at OECD CF Level 3 or EDSP Tier 1) primarily provide screening for possible endocrine activity. They are designed to provide a qualitative (yes or no) indication of perturbed EAT signaling as a consequence of direct interactions with a receptor or enzyme. A prototypical example is induction of vitellogenin (VTG; egg yolk protein precursor) in male fish by ER agonists, an endpoint included in several fish assays, including those described in OECD TG 229 and 230 (Supplemental Table S1). Indirect effects such as changes in sex steroid synthesis or thyroid hormone production also may be detected in some of the shorter-term *in vivo* assays. These mechanism-oriented assays do not generally expose organisms for a large proportion of their life cycle, and therefore are incapable of revealing the full spectrum of possible effects. Assays at this level are designed primarily for hazard identification in the context of indicating perturbation of specific HPG/T pathways. A positive outcome in these screening assays indicates a possibility for adverse effects in reproductive or developmental endpoints, or both, in longer-term tests (e.g., OECD CF Levels 4 and 5 or EDSP Tier 2). The results from these *in vivo* screens are used in a weight-of-evidence analysis to decide whether and how higher-tier *in vivo* tests should be performed. Some guideline examples of *in vivo* endocrine screening studies are the 21-day Fish Assay (OECD TG 230), the Fish Short-Term Reproduction Assay (FSTRA; OECD 229; OPPTS 890.1350), the Amphibian Metamorphosis Assay (AMA; OECD TG 231; OPPTS 890.1100), the Hershberger Bioassay in Rats (OECD TG 441; OPPTS 890.1400), and the Rodent Uterotrophic Assay (OECD TG 440; OPPTS 890.1600). These test guidelines are fully listed in Supplemental Table S1.

Slightly longer-term and/or more comprehensive tests (in terms of endpoints) at the OECD CF Level 4 typically are sensitive to more than 1 mode of action and include apical endpoints potentially suitable both for hazard and for risk assessment. Because they have numerous endpoints, the criteria for a positive result in terms of endocrine activity can be more complex than at lower levels. Because of the number and diversity of endpoints evaluated, Level 4 tests may also help to determine the relative sensitivity of endocrine-mediated effects as compared with effects of general toxicity (e.g., Ankley and Jensen 2014). Some Level 4 assays can provide data on adverse effects, which may be sufficient for identifying points of departure for hazard or risk assessment, or both.

However, most do not provide more comprehensive information about possible endocrine disrupting effects at multiple life stages, such as those obtainable from life-cycle experiments (OECD CF Level 5). Some guideline examples of OECD CF Level 4 studies include the Fish Sexual Development Test (FSDT; OECD 234), the Larval Amphibian Growth and Development Assay (OECD 241; OCSPP 890.2300), and the Pubertal Development and Thyroid Function Assay in Peripubertal Male and Female Rats (US EPA OPPTS 890.1500 and 890.1450, respectively). These test guidelines are fully listed in Supplemental Table S1.

Full life-cycle and multiple generational studies at the OECD CF Level 5 provide data on adverse effects and are especially useful for risk assessment because they add to the weight of evidence concerning potential impacts in humans and vertebrate wildlife. The effects observed may be caused by endocrine disruption or other mechanisms. Life-cycle and multigenerational tests include the evaluation of longer-term exposures during multiple windows of potential susceptibility to EASs, including maternal transfer to offspring, early development, sexual differentiation, and active reproduction; thus, there is a higher level of confidence about negative results. Some guideline examples of OECD CF Level 5 studies are the Medaka Extended One Generation Reproduction Test (OECD 240, OCSPP 890.2200) and the Extended One-Generation Reproductive Toxicity Study in Rats (OECD 443). These test guidelines are fully listed in Supplemental Table S1.

No specific test guidelines exist for characterizing endocrine activity in invertebrates; however, several tests could at least partially indicate perturbations of endocrine function through evaluation of apical endpoints. The Enchytraeid Reproduction Test (OECD 220) and the Earthworm Reproduction Test (OECD 222) broadly represent the Phylum Annelida. Arthropods are represented by the *Daphnia magna* reproduction test (OECD TG 211), the Mysid Chronic Toxicity Test (OPPTS 850.1350), the Developmental Toxicity to Dipteran Dung Flies (OECD 228), the Predatory Mite Reproduction Test in Soil (OECD 226), the Collembolan Reproduction Test in Soil (OECD 232), the Sediment-Water Chironomid Toxicity Using Spiked Sediment/Water (OECD 218 and OECD 219, respectively), and the Sediment-Water Chironomid Life-Cycle Toxicity Test Using Spiked Water or Spiked Sediment (OECD 233). In addition, 2 molluscan test guidelines on reproductive toxicity (*Lymnaea stagnalis* [OECD 243] and *Potamopyrgus antipodarum* [OECD 242]) were recently adopted by the OECD. These test guidelines are fully listed in Supplemental Table S1.

RESOURCE DEMANDS FOR ENDOCRINE SCREENING AND TESTING

One of the major issues in the context of the routine use of in vivo EDSP Tier 1/OECD CF 3/4 assays involves resources. This is particularly true when large numbers of chemicals need to be evaluated, such as the 10 000+ substances under consideration in the United States. Although slight differences exist, guideline studies required or recommended by the EPA, OECD, European Union, and Japanese programs

for EAS screening and testing are essentially the same, so the following discussion about resources, although focused on the EDSP as an illustration, is applicable to programs throughout the world. The cost of conducting the entire battery of 11 EDSP Tier 1 screening assays was estimated originally at US\$500 000 to US\$800 000 per chemical, with current estimates at closer to \$1 million per chemical. Following the 1st round of EDSP Tier 1 testing, The USEPA recommended further testing for 18 chemicals of the 52 (~35%) evaluated in the Tier 1 battery (USEPA 2016). Individual nonmammalian EDSP Tier 2 (OECD Level 4/5) tests have considerably larger projected costs per study, with estimates in the order of US\$300 000 to US\$350 000, US \$400 000 to US\$450 000, and US\$300 000 to US\$650 000, respectively, for the Larval Amphibian Growth and Development Assay, the Medaka Extended One Generation Reproduction Test, and the Avian Two-Generation Toxicity Test. Thus, if required, higher-tier tests would quickly and significantly add to the costs of EAS assessment on a per-chemical basis.

Guideline tests also can take a significant amount of time for scheduling and completion. For example, to conduct the FSTRA (OECD 229), there may be a wait time to rear sufficient numbers of fish from in-house cultures to the appropriate age for testing (i.e., 4.5–6 months old). In addition to the actual study period (i.e., a minimum 14-d pre-exposure phase followed by a 21-d exposure), considerable data collection and analysis occurs after the end of the FSTRA exposure, with biochemical analyses and histopathology requiring an additional 4 to 6 weeks. Some test chemicals also possess challenging physicochemical properties that may require nonroutine approaches for test solution preparation for aquatic testing. For example, the desire to avoid using solvents in testing can mean the development of solvent-free exposure systems for aquatic tests, which can add time and resources to the program.

Finally, EAS test batteries can require a significant number of animals. For example, the EDSP Tier 1 battery of mammalian and nonmammalian assays uses ~600 animals per test substance (Bishop and Willett 2014). Additional animals are also often required for dose range-finding studies, additional groups in pre-exposure to ensure sufficient numbers meet test performance criteria, and experiments that need to be repeated when a first attempt does not meet performance criteria.

Resource issues may become less problematic as information is gathered relative to endpoints or tests that are not informative or, possibly, redundant for the purposes of EAS screening. For example, Ankley and Gray (2013) evaluated data from method validation studies with 12 model EASs that have activity as ER agonists, AR agonists, AR antagonists, or inhibitors of different steroidogenic enzymes. All the chemicals had been tested in the FSTRA and in 1 or more of the 4 in vivo EDSP Tier 1 screens with rats. In most cases there was high concordance between the fish and the rat assays with respect to identifying chemicals that impacted specific endocrine pathways of concern, reflecting strong

structural and functional conservation of the HPG axis across vertebrates. The model EASs produced positive pathway-specific responses in the fish and 1 or more of the rat assays. However, some assays were clearly superior to others in terms of detecting specific pathways; for example, the effects of inhibitors of steroid hormone synthesis were most obvious in the FSTRA, whereas the activity of AR antagonists was clearest in the Hershberger and male pubertal assays. Based on this analysis, it appears possible to use just 2 of the current Tier 1 tests, the FSTRA and the male pubertal assay, to ensure full coverage of HPG axis pathways of concern. For example, Ankley and Gray (2013) proposed that these 2 tests could serve as initial “gatekeeper” assays, following which chemicals with negative findings may be exempt from further testing or, if positive, subjected to additional, confirmatory analyses with other relevant EAS assays. This would greatly enhance throughput of chemicals through initial testing, both in terms of resource utilization and timing.

Greater use of HTS data also should help address resource limitations relative to EAS screening and testing. As noted earlier, the EDSP is utilizing data from a battery of 18 ER-oriented HTS assays integrated into a computational model as a basis for prioritizing chemicals for more expensive and intensive Tier 1 testing (Browne et al. 2015). The accuracy of the ER computational model was 84% to 100% for predicting uterotrophic responses (Browne et al. 2015). This ER computational model was applied to 1812 commercial and environmental chemicals (which included 45 ER-positive and -negative reference chemicals), and a total of 111 chemicals (6.1%) were predicted to be strongly ER active (Judson et al. 2015). Expansion of this basic approach to encompass a suite of other endpoints relevant to perturbation of the HPG/T axes, as well as other potential toxicity pathways, has significant promise. As shown through the EDSP ER demonstration project, this type of approach could be used to prioritize chemicals for testing, based on the degree of their interaction with HPG/T targets of concern. In addition to the HTS assays for ER interactions, several existing ToxCast assays capture chemical interactions with the AR and enzymes involved in steroid synthesis such as CYP19 (Karmaus et al. 2016). High-throughput screening of chemicals for effects on steroidogenesis using human H295R adrenocortical carcinoma cells indicated that of the 2060 chemicals evaluated, 411 (20%) showed effects on at least 1 hormone in the steroidogenesis pathway evaluated (Karmaus et al. 2016). Additional assays amenable to HTS are being developed for other endocrine targets, including the HPT axis (Paul-Friedman et al. 2016). Although most existing HTS assays relevant to the HPG/T axes are based on mammalian systems, the high degree of cross-species conservation of structural and functional aspects of key endpoints (e.g., ER, AR) suggests that these systems should also be useful for nonmammalian vertebrates, at least at the level of screening and prioritizing chemicals for endocrine activity (LaLone et al. 2013; Ankley et al. 2016).

As a simple example or proof of concept of the application of HTS data, an analysis of the activity of the model AR

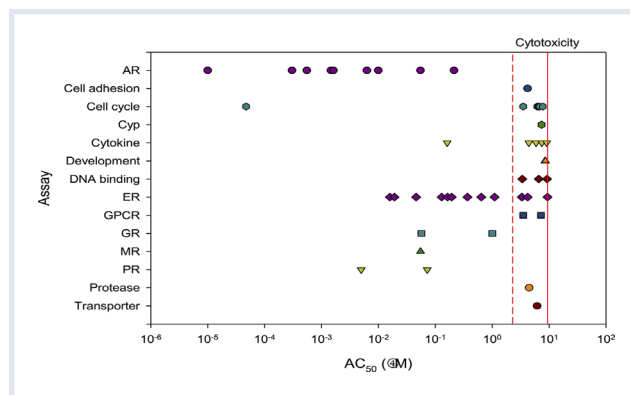


Figure 1. ToxCast data for 17 β -trenbolone (data download February 2016). Molecular targets of the ToxCast assays that showed positive responses are listed on the y-axis. The points on the graph indicate AC₅₀ values that correspond to the concentrations listed on the x-axis. The dashed line indicates the lower cytotoxicity limit, and the solid line indicates the median cytotoxicity response. AR = androgen receptor; Cyp = cytochrome P450; ER = estrogen receptor; GPCR = G protein-coupled receptor; GR = glucocorticoid receptor; MR = mineralocorticoid receptor; PR = progesterone receptor.

agonist 17 β -trenbolone (TRB) was conducted using data from the ToxCast HTS suite, and it found that the androgen was positive in 9 of the 10 assays designed to detect AR interactions (Figure 1). TRB also displayed some degree of activity in ToxCast assays with the progesterone receptors and ERs, which is consistent with the available knowledge relating to in vivo behavior of TRB in vertebrates, including fish. If TRB were an “unknown” chemical that exhibited this profile, it would be prioritized above other chemicals not displaying these types of interactions for further in vivo testing.

An additional use of HTS data in terms of optimizing resources would be to help select the individual assays and endpoints best suited for the generation of higher-level in vivo data to support regulatory decision making for EASs as opposed to conducting, for example, the entire battery of EDSP Tier 1/2 (or OECD Levels 3–5) assays. For example, based on the TRB analysis, one would conclude that invertebrate tests would likely not be required for a chemical with this profile, because most invertebrates do not appear to have a functional ER or AR (LaLone et al. 2013; Ankley et al. 2016). Instead, it would be most appropriate to consider using ER- and AR-mediated developmental and reproductive endpoints, and test designs such as the FSTRA or FSdT. Given the apparent sensitivity of the FSdT to androgens (Knacker et al. 2010), one might decide that this would be the optimal choice for further in vivo testing of the chemical of concern.

INTERPRETATION OF RESULTS FROM ENDOCRINE SCREENING AND TESTING

Assessments of environmental risk and hazard for EASs are unique compared with those historically conducted for other classes of chemicals, in that concerns raised relate to not only possible individual- or population-level adverse effects (Marty et al. this issue), but also the mechanistic basis via

which these effects occur. This is the foundation of legislated mandates behind most EAS screening and testing programs, and has necessitated the development of both screening-level assays and endpoints that enable identification of chemical perturbation of a given molecular target and tests that identify the possible negative consequences of this. This has sometimes resulted in a conceptual “disconnect” between the results of lower-tier and lower-level mechanistic assays and data from studies focused on apical responses, which often do not include collection of extensive mechanistic data. As a consequence, there may be no assurance that endocrine activity of a chemical determined *in vitro* or in short-term *in vivo* assays (EDSP Tier 1; OECD CF Tiers 2/3) is indeed the basis of adverse apical responses measured in longer-term assays (EDSP Tier 2; OECD CF Tiers 4/5). Some EAS assays such as the FSDT (OECD 234) feature the simultaneous collection of both mechanistic and apical data that can help inform this linkage (Ankley and Jensen 2014), but most existing endocrine assays do not, leaving the association between pathway perturbation and downstream responses uncertain. A conceptual framework that can help bridge this gap is the adverse outcome pathway (AOP; Ankley et al. 2010), which aims to depict causal, not correlational, linkages between perturbation of a molecular initiating event (MIE), such as receptor activation or inhibition, and the subsequent cascade of biological responses (or key events) that culminate in negative impacts on individuals or even populations. Various publications have described AOPs and AOP networks that could support EAS hazard and/or risk assessment in the context for establishing linkages across biological levels of organization (e.g., Ankley et al. 2009; Knapen et al. 2015).

For risk assessment it may not be imperative to establish and evaluate intermediate key events linking an MIE to an adverse outcome in an AOP, but this process becomes critical when conducting hazard assessments in which decisions may be based on linkage of a specific MIE perturbation (e.g., ER activation) to a discrete population-relevant adverse effect (e.g., reduced fecundity). Becker et al. (2015) discuss the application of Bradford-Hill considerations in assessing the strength of an AOP. Results from the approach outlined by Becker et al. (2015) can guide researchers toward additional data collection to either strengthen an existing AOP or investigate other lines of evidence that may be contributing to observed adverse effects.

For study designs that assess the potential for endocrine activity, it may be advisable to collect additional endpoints to better characterize the MIE and intermediate key events associated with adverse outcomes. Primary target organs of EASs in vertebrates include multiple segments of the HPG/T axes including the hypothalamus-pituitary plus the primary endocrine glands: the thyroids, adrenal-type tissues, parathyroids, pineal glands, pancreatic islets, and gonads. Other tissues and organs such as the liver, heart, and adipose tissue have secondary endocrine functions that may also be targeted by EASs (Schug et al. 2013). Yet, many studies focus on a limited number of targets, most commonly the

thyroid and gonads for analysis of biochemical or molecular endpoints. Only limited information is available on the effects of EASs on other potential targets, such as the adrenals (interrenal and chromaffin tissue in fish), which have been identified as commonly affected and vulnerable endocrine organs (Bergman et al. 2013). Adrenals and interrenals play a critical role during development but are also active in metabolism and, therefore, present a particularly effective target for assessing effects of EASs. Coady et al. (2014) suggest various additional tissue sampling procedures to maximize the mechanistic information that would be available for assessment following the conduct of an FSTRA or an AMA. For example, liver and kidney tissue could be preserved for potential histopathological investigations to assess whether adverse effects are due to systemic toxicity. In addition, any remaining fish blood plasma (i.e., not used to quantify circulating levels of VTG) could be analyzed for changes in sex steroid concentrations, which can be especially informative for some chemicals that cause endocrine disruption through inhibition of steroid synthesis (Ankley et al. 2009; Mihaich et al. this issue). Preserving additional tissues and fluids for potential future investigations and storing surplus animals from long-term studies can be particularly useful for collecting data on additional mechanistic and histopathological endpoints. This approach also has positive ethical aspects in regard to animal testing because it optimizes the use of the individual organisms.

Utilization of an AOP-based framework to establish credible linkages between mechanistic and diagnostic measurements and relevant apical outcomes represents 1 approach for effectively discerning whether endocrine perturbation by a given test chemical may actually be responsible for observed adverse effects in a test. A complementary approach to the evaluation of potential MIE involves the use of HTS data, such as that collected by the ToxCast program, which could provide insights as to potential nonendocrine biological activity of unknown chemicals. In other instances, HTS data for an unknown chemical may produce responses indicative of perturbation of MIEs associated with multiple toxicity pathways, only some of which involve endocrine function (Mihaich et al. this issue).

TEST SPECIES AND LIFE-STAGE SELECTION FOR ENDOCRINE SCREENING AND TESTING

The current internationally approved *in vivo* test guidelines addressing endocrine effects have been developed in a few model test species selected primarily for pragmatic rather than ecological reasons. Therefore, the extrapolation of test data to all other species, even within a given phylogenetic group, may not always be appropriate. For example, in the case of amphibians, only 1 species, *Xenopus laevis*, is used in guideline toxicity tests. This species is aquatic in all life stages, and therefore it might not be adequately representative of amphibians with different life strategies. Similarly, in avian reproduction testing, only precocial species (i.e., mallard duck and quail, which are well-developed at hatch) are used in standard avian toxicity tests. The altricial strategy

in avian development, which includes most songbirds (passerines) and birds of prey, is not represented in standard avian reproduction test guidelines. Altricial hatchlings depend on intensive parental care, which may enhance their sensitivity to EASs that affect behavior (Jaspers 2015). Thus, it is important to consider the genetic, physiological, and ecological traits of laboratory species used in EAS screening and testing, and how these species may or may not represent wild populations when attempting to extrapolate laboratory findings to environmental populations and communities (Segner 2011).

In other instances, test guidelines may be completely lacking for potentially sensitive phyla. For example, guideline tests on reptiles have not yet been developed, although possible endocrine-mediated adverse effects have been observed in the field (Guillette et al. 1999), and temperature-dependent sex determination could make some species in this class particularly vulnerable to endocrine disruption during development (Bergeron and Crews 1998). Test guidelines that can screen for and identify endocrine activity in invertebrates are somewhat lacking. The limitations of the invertebrate tests in OECD CF are well described in footnote 4 of the CF document (OECD 2012a, p. 4): "At present, the available invertebrate assays solely involve apical endpoints which are able to respond to some endocrine disruptors and some non-EDs. Those in Level 4 are partial life-cycle tests, while those in Level 5 are full- or multiple life-cycle tests." The 2 new mollusk reproduction tests do not change this overall picture because they do not include endocrine-specific endpoints, although they cover endocrine-mediated reproductive effects in mollusks such as aquatic snails.

Fish are by far the best represented nonmammalian vertebrate taxa in terms of EAS assays. However, the 3 main model species used in fish EAS guideline toxicity tests represent warm freshwater, fractional spawning teleost fish with relatively short generation times (Supplemental Table S2). They do not represent marine species, pelagic species, cartilaginous fish, and the diversity of reproductive strategies occurring among fish including viviparity (live breeding), seasonal spawners, and species with extremely high fecundity. For example, in Atlantic cod (*Gadus morhua*), a single female spawns millions of eggs, so a very minor reduction in fecundity could potentially be less impactful at the level of the population in comparison with fish species that produce larger and fewer eggs. In viviparous fish, such as the eelpout (*Zoarces viviparus*), the exposure scenario of the embryos is different from the oviparous test species. Chemicals could accumulate because of maternal transfer, affecting the sensitivity of the species and the metabolism of the chemical. For example, malformations in eelpout embryos after maternal exposure to 17 α -ethynylestradiol have been reported (Morthorst et al. 2014).

Among the commonly used small fish species in EAS screening, there are life history and morphological differences that can cause differential sensitivity to perturbation of different endocrine pathways (Supplemental Table S2). For

example, zebrafish gonads initially develop as ovaries, but then in male fish, the ovarian tissue degenerates and the testis develop (Takahashi 1977; Maack and Segner 2003). This period of juvenile hermaphroditism in the zebrafish may explain the increased sensitivity of the sex ratio endpoint for this species in comparison with other fish species exposed to AR agonists such as TRB during critical developmental windows. For example, concentrations around 10 ng TRB per liter cause irreversible phenotypic sex reversal in zebrafish when exposed during early life stages (Holbech et al. 2006; Örn et al. 2006; Morthorst et al. 2010), whereas the same exposure concentrations do not cause similar effects in Japanese medaka or Western mosquitofish (Örn et al. 2006; Sone et al. 2005). However, in assays focused on reproductive success, there can be limitations to using the zebrafish because sexually mature male zebrafish do not have androgen-responsive secondary sex characteristics as do, for example, fathead minnows and Japanese medaka. In particular, the observation of female fish with male secondary sex characteristics is diagnostic for AR agonists (Ankley and Jensen 2014; Borgert et al. 2014). Thus, test species and life-stage selection are important elements to consider for endocrine screening and testing (Parrott et al. this issue). With preliminary information on the endocrine mode of action from QSAR predictions or HTS in vitro assays, more informed decisions on species selection and life stage can be made before in vivo testing; however, in many cases, additional information on the natural history and ecological traits of species of concern are needed to understand how endocrine effects observed in laboratory test species will apply toward species in the field.

To increase output of information regarding sensitive life stages with a minimum investment of effort, it may be advisable to combine standard study designs. This could entail, for example, adding a fish early-life-stage test (OECD 210) to a fish full-life-cycle test. This combination would allow for the assessment of maternal transfer to the F1 generation if deemed necessary based on knowledge of physicochemical properties of a test substance, such as the propensity to bioaccumulate. Depending on the specific question to be addressed, the F1 generation could be exposed to the same concentration as the F0 generation or reared under control conditions. Analogously, the FSTRA (OECD 229; OPPTS 890.1350) could be combined with the FSDT (OECD 234). This approach would allow for the assessment of both the activational (e.g., reproduction) and the organizational (e.g., sexual development) aspects of the fish life cycle, without actually conducting a fish full-life-cycle test.

TEST DESIGN CONSIDERATIONS FOR ENHANCING CURRENT ENDOCRINE ACTIVE SUBSTANCE ASSAYS

Test sensitivity and power

From a statistical perspective, a design limitation that is particularly relevant to EAS testing is the complexity of the current in vivo assays in terms of the number and variety of

endpoints, animal groupings, and sampling schedules. For example, both fish and avian multigenerational studies can have greater than 40 distinct endpoints distributed among the sexes and generations. This can result in hundreds of statistical comparisons, the breadth of which often requires the simultaneous use of several different statistical methods. Additionally and importantly, the large number of endpoints tends to greatly increase the likelihood of Type 1 errors. Another potential design limitation relevant to aquatic animal studies involves the statistical unit. If animals are to be housed in groups of 2 or more individuals, then these test subjects are considered to be interdependent, and consequently the unit of analysis will be, with few exceptions, the contained group rather than the individual. Statistically, this requirement can be managed by increasing the number of replicates relative to the group sizes. The inclusion of multiple replicates not only improves the power of the assay, it instills confidence to the findings and avoids the potential influence of “tank effect” on study outcome. However, for a large study, this can cause the number of tanks (and tank maintenance) to become untenable.

Although it is evident that some of the aforementioned design limitations cannot be easily rectified, others can be avoided by judicious planning. For example, issues of unintended bias can be mitigated by the masking of group identities and randomization of all study variables that cannot be held constant (e.g., Wolf et al. 2015). Power analysis can be performed a priori to estimate the optimal number of replicates and group sizes to be used for a given study. If a no-observable effect concentration is to be determined, then a power analysis should be performed during the study design phase to ensure sufficient power to find a toxicologically important effect. If a point estimate of an effect concentration is to be estimated, then a sensitivity analysis should be conducted to determine that an $x\%$ effect can be estimated, if it occurs, with a confidence interval that is not so wide as to render the estimate meaningless. For example, in a power analysis conducted during the validation of the FSDT (OECD 234), a minimum of 4 replicate tanks containing 30 embryos each was suggested for achieving sufficient power to detect 15% to 25% phenotypic sex reversal (when the genetic sex of the fishes was known). The same design would have sufficient power to detect a 31% change in sex ratio when the genetic sex of the fishes is not known (OECD 2012b). Such power analyses should be based on relevant historical data, information gleaned from published literature, or if neither is available, on the results of a preliminary range-finding study. Some reports of historical control data for endocrine screening studies (i.e., FSTRA and AMA) are now available in the peer-reviewed literature (Coady et al. 2014; Schapaugh et al. 2015), and they could be productively used to further optimize design of the assays.

Concentration selection

Testing for EASs is historically unique in that there is concern not only for adverse effects but also for the AOP via which these effects occur. In this context, the selection of

appropriate doses and concentrations is imperative for examining potential endocrine activity and effects that occur apart from potential systemic toxicity. For in vitro assays, the analogue for systemic toxicity observed in animal assays is cytotoxicity. Currently, in vitro EDSP and OECD guidelines allow up to 20% cytotoxicity, which is a level that could potentially confound results. The level of allowable cytotoxicity could be lowered to address this potential issue. For example, for cell-based assays, the Interagency Coordinating Committee on the Validation of Alternative Methods has recommended that only concentrations that do not cause greater than 10% cytotoxicity should be included in the analysis (Interagency Coordinating Committee on the Validation of Alternative Methods and National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods 2003).

In the USEPA guideline for the FSTRA (OPPTS 890.1350), the criteria for the highest test concentration is either the maximum tolerated concentration (defined as one-third the 96-h LC50), the limit of solubility, or 100 mg test chemical per liter. In a review of the first 52 chemicals that were evaluated in the Tier 1 EDSP (USEPA 2013), there were several examples where in vivo overt toxicity at the highest concentration in the FSTRA confounded the interpretation of effects. The USEPA concluded that “the use of a 1/3 the LC50 value sometimes resulted in substantial mortality as well as effects on physical appearance, behavior and/or changes in body weight” (USEPA 2013, p. 97 of 160). Changing the current guidance to using one-tenth the LC50 and/or other available data to set the highest test concentration has been recommended as a solution (Wheeler et al. 2013) and has already been implemented in endocrine assays with fish, for example, the Medaka Extended One Generation Reproduction Test (OECD 240, OCSPP 890.2200) and the FSDT (OECD 234). However, the test concentration setting for these higher-tier tests encompassing all life stages should undoubtedly use all available chronic data and targeting range-finding strategies.

A toxicokinetic (TK) approach could also be used to guide concentration selection in endocrine screening and testing. Application of TK approaches for dose setting in ecotoxicology studies is rare. However, this approach has a long history of application in drug development and has become more common across mammalian toxicology in an effort to harmonize an approach for identifying a maximum tolerated dose (European Centre for Ecotoxicology and Toxicology of Chemicals 1996; USEPA 2003). In brief, the aim of this approach is to identify the highest level that does not exceed linear pharmacokinetics in plasma. The use of excessively high doses or concentrations can create TK processes that result in a systemic exposure to either parent compound or metabolites that no longer increases with increasing dose, or a systemic exposure that hugely increases with a small increase in dose. Exceedingly high doses and associated nonlinear TK will likely produce results that are not relevant for hazard or risk assessments. Consequently, the driver for adoption of the TK approach for dose setting in mammalian toxicology studies has been to improve testing and

assessment methodologies to investigate a specific mode of action (e.g., carcinogenicity or endocrine disruption) and not confound the interpretation for human relevance (Boobis et al. 2008). Creton et al. (2012) discuss practical examples that underscore the role TK can play in improving study design and interpretation through selection of appropriate high doses. One of the barriers to the application of TK in ecotoxicology would be development of improved sampling methods and analytical capabilities, but this would be a highly productive area of research for the aquatic and wildlife toxicology community in general.

In addition to setting the high concentration and dose in an EAS test system, the selection of the range of doses or test concentrations and the spacing of intervals is an important consideration. In the FSTRA and the AMA, for example, a minimum of 3 concentrations (plus negative control) are recommended (see OECD 229, OECD 231, OPPTS 890.1100, and OPPTS 890.1350). A range of spacing factors between 3.2 and 10 are suggested, with a minimum test concentration differential between the highest and lowest of 1 order of magnitude. The spacing of test concentrations in any test can be affected by the limit of quantification of the analytical capability available for the test substance of interest. Difficulties can arise when water solubility and/or biological effects occur at low levels that approach the limits of quantification. Using a 10-fold spacing between concentration levels may lead to difficulties in quantifying the lowest concentration level, and thus closer spacing may be desired. In contrast, the maximum dose separation of 10-fold may be preferable to ensure that there are at least 2 concentrations in the FSTRA and the AMA that are free from overt toxicity if the highest concentration is affected in this way (Coady et al. 2014). Furthermore, using a dose separation of up to 10-fold enables the inclusion of lower concentrations that may have greater environmental relevance. Including low, environmentally relevant concentrations in study designs can increase the confidence in the study findings in regard to their applicability for environmental risk assessments. Furthermore, including additional concentration groups in the study design (e.g., testing 5 concentrations rather than the minimum requirement of 3) could enhance the test design for determining no-observable effect concentration and/or effect concentration values for use in environmental risk assessments. However, this approach should be balanced against animal use considerations and the intended purpose of a particular assay or test (screening or concentration–response derivation for risk assessment).

As is evident from this discussion, there is a pressing need for more guidance for concentration selection in EAS screening and testing. The specific challenge of identifying potential endocrine activity in the absence of the confounding effects of systemic toxicity requires careful consideration. A concentration setting guidance document developed through collaborative and harmonized efforts could reduce uncertainty and guide more refined endocrine assessments in the future.

Analytical confirmation and aquatic delivery of test substances

Measuring concentrations of test chemicals in dosing media is a critical factor in determining the acceptability of *in vivo* screening assays and definitive tests for EASs, particularly when attempting to translate laboratory data into an environmental setting. Because of the extreme potency of some EASs (e.g., tributyltin, steroids such as TRB or 17 α -ethynylestradiol), this can present challenges for both analytical method development and exposure of test organisms. In some instances, low, environmentally relevant concentrations can be problematic when no-observable effect concentrations are approaching (or lower than) the analytical limits of quantification (Mensink et al. 2002; Ankey et al. 2003). Steroids such as TRB can also be composed of stereoisomers. Accounting for this in analytical method development is especially pertinent when multiple stereoisomers display endocrine activity (e.g., Jensen et al. 2006).

Managing analytical challenges may require significant resources (including time) to develop the necessary techniques, validate appropriate laboratory and analytical methods, and understand exposure dosing and conditions, a necessity that often is underappreciated in EAS testing. This includes consideration of possible changes in the efficiency of chemical delivery over the course of longer-term assays, where, for example, changes in the microbial degradation of chemicals can cause significant temporal decreases in chemical concentration. In addition, major metabolites formed during *in vivo* exposures should be screened for, and potentially identified, in solutions and biological samples (tissues from exposed organisms). For some chemicals, it is the metabolites rather than the parent chemical that possess significant endocrine activity. This information can help to inform the interpretation of endpoints and future biological assays. Finally, robust quality assurance and control procedures are necessary for establishing confidence in test chemical measurement (and isomers and metabolites identification), particularly in longer-term studies; this encompasses many considerations, including identification, quantification (e.g., 4-point mass spectrometry identification), recovery, and reproducibility. All of these factors are critical drivers in the overall limit of quantification value for an analytical method, as well as for continually improving analytical technologies.

The use of organic cosolvents historically has been an acceptable means to deliver test substances in aquatic testing (OECD 2000). However, there are many advantages in using techniques that eliminate the use of cosolvents, including decreasing the possibility of testing above the functional limit of water solubility, reducing the potential for low dissolved oxygen levels caused by increased microbial biomass (because of microbial metabolism of some solvents), reducing the use of animals (elimination of a solvent control group), and decreasing the potential for toxicological interactions between a cosolvent and a test chemical. Elimination of cosolvent use through techniques such as

generator columns for solids or liquid–liquid saturators for liquids has become a recommended option for EAS testing (e.g., see OECD 229, OPPTS 890.1350, OECD 321, OPPTS 890.1100, OECD 240, and OCSPP 890.2200). Although this can supply superior test data, additional time may be required when testing to optimize appropriate solvent-free delivery techniques associated with column size, water flow rate, chemical loading regimens, among others. It may also be necessary to accept that the exposure conditions using solvent-free techniques may be more variable than might be achieved using solvent delivery techniques.

LABORATORY CAPABILITIES FOR ENDOCRINE SCREENING AND TESTING

The relatively recent development in the field of ecotoxicology of regulatory test protocols that include mechanistic endpoints, that is, those indicative of endocrine function, requires a level of scientific and technical expertise beyond that previously needed in most contract testing situations. Further, these types of mechanistic measurements often have been validated in only a small number of laboratories or by a few people, early in the adoption of new methodology, which means there often is no or only limited historical control data to evaluate the expertise of the individual or team learning the new methodology.

The key to developing and learning new techniques and methods is clear communication among researchers who develop methods and those who implement or transfer the methods. The communication can take many forms, such as the inclusion of bold text warnings on portions of methods that must be strictly followed, providing mean and SD targets for specific endocrine endpoints, and/or video training that clearly demonstrates a critical step or technique for a procedure. For some methods, transfer of technical information also can be facilitated through hands-on workshops. As new methods develop and are successfully implemented across other laboratories, it becomes important to periodically review the error rate and method or endpoint precision and share historical control databases (Coady et al. 2014; Schapaugh et al. 2015).

Many endpoints that are informative of the endocrine mechanism of action (e.g., VTG protein or mRNA levels, gonadal histopathology, and plasma sex steroids) may be prone to measurement error because of the technical expertise and expert judgment required for providing quality data. It is illustrative to consider VTG measurements as an example of the challenges that can occur when requiring testing laboratories to adopt nonroutine endpoints. The immunoassays typically used for VTG protein determination in samples from EAS tests with fish (e.g., OECD TG 229, 230) have not been routinely used in most testing laboratories. Commercial kits are available for VTGs that could be used as means of normalizing reported values; however, these exist only for a limited number of species and, probably most importantly, can be costly when processing a large number of samples. This has resulted in the development of many in-house assays, some of which have not followed robust

validation procedures, which can cause unanticipated variability in measured VTG values across different laboratories. Another potential issue relative to VTG measurements is that of basic immunoassay performance over time even within a given laboratory, which can be compromised by changes in the availability of specific (i.e., antibodies or labels) and nonspecific (i.e., serum albumins or extraction solvents) reagents. As an example, average plasma VTG protein concentrations in female fathead minnows measured via a commercially available homologous ELISA kit (Biosense Fathead Minnow Vitellogenin ELISA kit; Bergen, Norway) varied more than 10-fold across 11 different FSTRA studies conducted in the same laboratory (Table 1). This likely was due to multiple factors, including inherent variability of VTG levels among individual female fish (e.g., Jensen et al. 2001), the variability in the ELISA kit performance, and the variability associated with different laboratory technicians conducting the assay. Based on a post hoc power analysis, the power to detect 20% to 80% decreases in female VTG varied from 5% to 99% across 11 different studies (Table 1). This analysis indicates that additional replication in the test design and/or an alternative, possibly less variable method of measuring VTG concentrations (e.g., Zhang et al. 2004) may be called for to strengthen the power of the VTG endpoint in EAS assays with fish.

Histopathological assessment of gonadal tissue in fish, amphibians, or birds, or of thyroid tissue in amphibians or birds can be another problematic endpoint. In this regard, although technical issues (e.g., sample dissection, preservation, staining, among others) can lead to poor-quality material, a key problem often encountered is the misinterpretation of pathological findings because of insufficient training in lower animal endocrine toxicologic pathology (Wolf et al. 2015). To this end, more guidelines on best practices are needed, especially involving species for which this has not been a routine endpoint in the past. Although a number of high-quality histopathology guidance documents have been generated by the OECD and USEPA (e.g., OECD 2010b), these tend to be static references that might benefit from online publication in a database or atlas format that can be continuously updated. An additional recommendation for this endpoint is the inclusion of pathology peer review (evaluation of a subset of the slides by a 2nd pathologist) as a quality-control measure, particularly for instances in which the results are equivocal, unanticipated, and/or controversial.

KNOWLEDGE GAPS IN ENDOCRINE SCREENING AND TESTING

A significant data gap for understanding EAS hazards is the lack of fundamental knowledge concerning endocrine pathways for many invertebrate species. Part of the difficulty in addressing this is related to the vast number of invertebrate species that exist, combined with the great diversity this group displays in the endocrine control of growth, development, and reproduction. Because of this lack of understanding, an unfortunate trend in the field has been for some to assume that the same indicators of endocrine activity in

Table 1. Mean, SD, and Percent Power Associated with Measurement of Female Fathead Minnow Vitellogenin Plasma Levels (Using ELISAs) in 11 Different Fish Short-Term Reproduction Assays

	Study No.										
	1	2	3	4	5	6	7	8	9	10	11
N	4	4	4	4	4	4	4	4	4	4	4
Mean (±SD) VTG in control females	27.6 ± 21.5	43.0 ± 16.1	107 ± 38.0	154 ± 36.2	66.7 ± 19.0	67.2 ± 47.3	33.5 ± 8.48	66.3 ± 22.9	70.5 ± 77.7	12.2 ± 4.17	22.9 ± 9.49
Power to detect a 20% decrease	5.9	9.0	9.4	16.1	12.2	6.1	14.4	9.8	5.4	9.9	8.2
Power to detect a 40% decrease	8.7	23.3	25.5	54.2	38.5	9.6	47.6	27.1	6.8	27.4	19.6
Power to detect a 60% decrease	13.9	48.7	53.1	89.6	74.3	16.1	84.4	56.1	9.2	56.6	40.6
Power to detect a 80% decrease	21.9	75.4	79.8	99.2	94.7	26.0	98.2	82.6	12.8	83.2	65.8

Power analyses were conducted in SAS 9.2 PROC GLMPower (SAS Institute, Cary, North Carolina, USA). VTG = vitellogenin.

vertebrates (e.g., VTG induction by estrogens in fish) also apply to invertebrates, which often is not the case. For example, the transcriptomic response of the *Vtg2* gene in *Daphnia magna* is not elevated in response to chemicals with known estrogenic modes of action in vertebrates (Hannas et al. 2011), and there is no valid evidence that vertebrate sex steroids have endocrine or reproductive roles in mollusks (Scott 2013). Consequently, there is a pressing need for research to support development of invertebrate-specific EAS screens and tests. As a 1st step in understanding the endocrine system of these ecologically important phyla, it is imperative to molecularly and functionally characterize the numerous nuclear receptors that are present. Recently available high-throughput sequencing (e.g., oyster genome) and genome mining tools can speed up this process and develop useful assays for hazard and risk assessment (Zhang et al. 2012).

Another significant gap in understanding the effects of endocrine disruptors is in the consideration of other relevant endocrine pathways in vertebrates. From an HPG perspective, most attention has been on soluble nuclear hormone receptors, but little attention, if any, has been given to membrane receptors for sex steroid hormones, including ERs, ARs, and progesterin receptors. These receptors have been described for mammals and fish, and have been shown to regulate reproduction through oocyte maturation and sperm motility, among other mechanisms (Thomas et al. 2006). Further, work that originated first in Europe has shown that reproductive effects from exposure to progestins may be a major concern for fish (Zeilinger et al. 2009), frogs (Säffholm et al. 2012), and birds (Tell et al. 1999). Synthetic progestins, such as levonorgestrel, used in the birth control pill are found in the environment (Lopez de Alda et al. 2002). Concern has also been raised about glucocorticoids in the environment and their action on the hypothalamic-pituitary-interrenal axis in aquatic vertebrates. Fish, like other vertebrates, produce endogenous corticosteroid hormones such as cortisol in interrenal cells located in the head kidney (Mommensen et al. 1999). As in other vertebrates this natural corticosteroid binds to glucocorticoid receptors, which work as transcription factors to regulate genes involved with glucose metabolism, stress response, the immune system, blood pressure, and osmoregulation (Weyts et al. 1999; Aluru and Vijayan 2009), among other endpoints. Glucocorticoids are used for human and veterinary purposes to treat a broad group of disorders including skin allergies, asthma, and rheumatic disease, among others, and can occur in the environment at nanograms to micrograms per liter (ng/L to µg/L) concentrations. Synthetic glucocorticoids can have a variety of in vivo effects in fish from altering plasma glucose levels, to destabilizing the hypothalamic-pituitary-interrenal axis (Kugathas and Sumpter 2011; LaLone et al. 2012; Nesan and Vijayan 2013; Nakayama et al. 2014), and even depressing reproduction (Schreck 2010) by decreasing VTG synthesis in the liver and consequently egg production.

New sensitive analytical methods using nontargeted approaches have identified many more contaminants than

previously anticipated in receiving waters. When these methods are coupled with bioanalytical tools to measure biological effects in surface waters, it is clear that contaminants exist at concentrations high enough to affect the function of several endocrine-related receptors (Escher et al. 2014). Among some of the activities measured are the nuclear factor erythroid-related factor 2 involved with oxidative stress, peroxisome proliferator-activated receptors α and γ involved in fatty acid metabolism, aryl hydrocarbon receptor involved in CYP induction, retinoic acid receptor (RAR) isoforms (RAR α , RAR β , RAR γ) involved with retinoic acid-related events, retinoid-related orphan receptor β , and retinoic-X receptor β involved in thyroid signaling (Escher et al. 2014), all of which could interfere with endocrine-related endpoints. Several of these receptors have been studied in relation to fish reproduction (peroxisome proliferator-activated receptors: Cheshenko et al. 2008; RARs: Lubzens et al. 2010; retinoic-X receptor: Habibi et al. 2012). Although 2 of these receptors are often associated with non-endocrine-mediated toxicity, aryl hydrocarbon receptor and Nrf2, they perhaps should not be discounted because significant cross talk between ER and aryl hydrocarbon receptor has been described in the literature that shows disruption of VTG production in fish (Bemanian et al. 2004), and between aryl hydrocarbon receptor and Nrf2 that is related to developmental embryo toxicity (Rousseau et al. 2015). Developing screening assays for activity through these receptors may be beneficial when trying to decipher the endocrine-related effects on whole organisms from exposures to individual chemicals that appear not to follow canonical AOPs.

In addition, the close relationship between the endocrine and the immune systems of organisms needs to be better evaluated to help explain phenomena such as morbidity in fish and disease outbreaks in avian populations exposed to complex mixtures that include EASs. The head kidney is involved in the production of cortisol and catecholamines, as well as hematopoiesis and other immune functions (Weyts et al. 1999). Sex steroids are known to play immunomodulatory roles in fish, and exposure to contaminants may make fish more sensitive to pathogens (reviewed by Milla et al. 2011). Similarly, the thymus, which plays a critical role in the immune response of higher vertebrates, is a known target of endogenous estrogens (Zoller and Kersh 2006). For example, exposure of European sea bass (*Dicentrarchus labrax*) to exogenous estrogens was recently reported to affect thymic growth and regionalization in juveniles, which may have persistent consequences for the immune function of adults (Seemann et al. 2015).

More research is also required to better define AOPs for various classes of chemicals commonly tested for possible endocrine activity, and current protocols and test guidelines should be strengthened to help differentiate between endocrine and nonendocrine modes of action (Mihaich et al. this issue). Also, a better understanding is needed concerning the effects of nonchemical stressors on endocrine function. For example, in female rainbow trout

(*Oncorhynchus mykiss*), confinement stress decreases VTG levels resulting in reduced egg size and significantly lower survival rates for progeny (Campbell et al. 1994). The inability to differentiate between the effects of physical or social stress and those caused by an EAS may greatly confound data interpretation and require the evaluation of additional endpoints to ensure the true cause is understood. Strengthening the understanding of AOPs should enable the use of molecular endpoints in screening assays to classify test chemical modes of action, thereby directing subsequent testing and potentially eliminating some of the high-cost assays that are now required.

Finally, many HTS assays that are (or could be) used to categorize and screen chemicals are based on mammalian (largely human) nuclear receptors and associated assays (e.g., Browne et al. 2015). To support use of these types of assays to predict endocrine functional endpoints in non-mammalian vertebrates, there is a need to conduct systematic research to examine cross-species structural and functional conservation of key MIEs, and develop, where necessary, HTS reporter assays with, for example, fish receptors (e.g., Ankley et al. 2016). A few assays already exist for some receptors that illustrate the efficacy of using this approach (Liu et al. 2005; Sabo-Attwood et al. 2007), and there is a whole fish embryo bioassay for estrogenic chemicals (Brion et al. 2012). More research should be devoted to developing more of these types of assays and making them commercially available.

CONCLUSION

There has been substantial scientific advancement through the development and implementation of the testing and hazard and risk assessment approaches to evaluate potential adverse effects through an endocrine mechanism. In particular, the existing test systems and frameworks that have been developed for assessing EASs via interaction with the EAT pathways are relatively comprehensive for identifying and assessing potential endocrine effects. However, opportunities now exist to retrospectively examine the lessons learned from the recent implementation of these efforts to improve the reliability and relevance of endocrine assessments. Priority areas that were identified included:

- leveraging information to the extent possible from HTS assays to prioritize and inform testing programs;
- utilizing *in silico* and *in vitro* data, and experience with existing assays as a basis for modifying screening and testing frameworks to optimize resource use;
- developing additional approaches to address species sensitivity, sensitive life stages, and critical endpoints to improve the predictive ability to detect an adverse effect at the population level; and
- identifying gaps that can be addressed by research to improve testing paradigms.

It is impossible to address every uncertainty, and currently available test systems are relatively robust for addressing

endocrine modalities that operate through the EAT pathways. However, implementing the recommendations outlined in the present study will move us in the direction toward reducing uncertainty in testing and assessment approaches that evaluate potential adverse effects exerted through an endocrine mechanism.

Acknowledgment—We thank John Green and Lisa McFadden for assistance in providing statistical support, and David Dreier for assistance in organizing and graphing the ToxCast results for TRB.

Disclaimer—The views and statements expressed in this paper are those of the authors. The views or statements expressed in this publication do not necessarily represent the views of the organisations to which the authors are affiliated, and those organisations cannot accept any responsibility for such views or statements. The contents of this manuscript neither constitute nor necessarily reflect USEPA policy. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US government. Nancy D Denslow is one of the founders, has stock, and sits on the Board of Directors of Banyan Biomarkers, which is developing assays for human traumatic brain injury.

Data availability—Raw data for the power analysis example in Table 1 can be provided through request from the corresponding author (kcoady@dow.com). High-throughput data (used as an example in Figure 1) can be accessed at USEPA's ToxCast dashboard (<https://actor.epa.gov/dashboard/>).

SUPPLEMENTAL DATA

Supplemental Table S1. List of in vivo test guidelines and guidance documents relevant to assessment of endocrine active substances.

Supplemental Table S2. Fish species commonly used in endocrine active substance screening and testing.

REFERENCES

- Aluru N, Vijayan MM. 2009. Stress transcriptomics in fish: A role for genomic cortisol signaling. *Gen Comp Endocrinol* 164:142–150.
- Ankley GT, Bencic D, Breen M, Collette TW, Conolly R, Denslow ND, Edwards S, Ekman DR, Jensen KM, Lazorchak J, et al. 2009. Endocrine disrupting chemicals in fish: Developing exposure indicators and predictive models of effects based on mechanisms of action. *Aquat Toxicol* 92:168–178.
- Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrano JA. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29: 730–741.
- Ankley GT, Gray LE. 2013. Cross-species conservation of endocrine pathways: A critical analysis of Tier 1 fish and rat screening assays with 12 model chemicals. *Environ Toxicol Chem* 32:1084–1087.
- Ankley GT, Jensen KM. 2014. A novel framework for interpretation of data from the fish short term reproduction assay (FSTRA) for the detection of endocrine-disrupting chemicals. *Environ Toxicol Chem* 33:2529–2540.
- Ankley GT, Jensen KM, Makynen EA, Kahl MD, Korte JJ, Hornung MW, Henry TR, Denny JS, Leino RL, Wilson VS, et al. 2003. Effects of the androgenic growth promoter 17-beta-trenbolone on fecundity and reproductive endocrinology of the fathead minnow. *Environ Toxicol Chem* 22: 1350–1360.
- Ankley G, LaLone C, Gray LE, Villeneuve D, Hornung M. 2016. Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems. *Environ Toxicol Chem* 35:2806–2816.
- Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, Van Der Burg B, Villeneuve DL, et al. 2015. Increasing scientific confidence in adverse outcome pathways: Application of tailored Bradford-Hill considerations for evaluating weight of evidence. *Regul Toxicol Pharmacol* 72:514–537.
- Bemanian V, Male R, Goksøyr A. 2004. The aryl hydrocarbon receptor-mediated disruption of vitellogenin synthesis in the fish liver: Cross-talk between AHR and ER α -signalling pathways. *Comp Hepatol* 3:2–15.
- Bergeron JM, Crews D. 1998. Effects of estrogenic compounds in reptiles: Turtles. In: Kendall RJ, Dickerson RL, Giesy JP, Suk WP, editors. Principles and processes for evaluating endocrine disruption in wildlife. 1st ed. Pensacola (FL): Society of Environmental Toxicology and Chemistry Press. p 291–300.
- Bergman A, Heindel JJ, Jobling S, Kidd KA, Zoeller RT, editors. 2013. State of the Science of Endocrine Disrupting Chemicals 2012. Geneva (CH): United Nations Environment Programme and the World Health Organization. 29 p.
- Bishop PL, Willett CE. 2014. The use and acceptance of other scientifically relevant information (OSRI) in the U.S. Environmental Protection Agency (EPA) Endocrine Disruptor Screening Program. *Birth Defects Res B Dev Reprod Toxicol* 101:3–22.
- Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirwat M, Schlatter J, Seed J, Vickers C. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. *Crit Rev Toxicol* 38:87–96.
- Borgert CJ, Mihaich EM, Quill TF, Marty MS, Levine SL, Becker RA. 2011. Evaluation of EPA's Tier 1 Endocrine Screening Battery and recommendations for improving the interpretation of screening results. *Regul Toxicol Pharmacol* 59:397–411.
- Borgert CJ, Stuchal LD, Mihaich EM, Becker RA, Bentley KS, Brausch JM, Coady K, Geter CR, Gordon E, Guiney PD, et al. 2014. Relevance weighting of tier 1 endocrine screening endpoints by rank order. *Birth Defects Res B Dev Reprod Toxicol* 101:90–113.
- Brion F, Le Page Y, Piccini B, Cardoso O, Tong SK, Chung BC, Kah O. 2012. Screening estrogenic activities of chemicals or mixtures in vivo using transgenic (cyp19a1b-GFP) zebrafish embryos. *PLoS One* 7:e36069. DOI: 10.1372/journal.pone.0036069
- Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening chemical for estrogen receptor bioactivity using a computational model. *Environ Sci Technol* 49:8804–8814.
- Campbell PM, Pottinger TG, Sumpter JP. 1994. Preliminary evidence that chronic confinement stress reduces the quality of gametes produced by brown and rainbow trout. *Aquaculture* 120:151–169.
- Cheshenko K, Pakdel F, Segner H, Kah O, Eggen RI. 2008. Interference of endocrine disrupting chemicals with aromatase CYP19 expression or activity, and consequences for reproduction of teleost fish. *Gen Comp Endocrinol* 155:31–62.
- Coady KK, Lehman CM, Currie RJ, Marino TA. 2014. Challenges and approaches to conducting and interpreting the Amphibian Metamorphosis Assay and the Fish Short-Term Reproduction Assay. *Birth Defects Res B Dev Reprod Toxicol* 101:80–89.
- Creton S, Saghir SA, Bartels MJ, Billington R, Bus JS, Davies W, Dent MP, Hawksworth GM, Parry S, Travis KZ. 2012. Use of toxicokinetics to support chemical evaluation: Informing high dose selection and study interpretation. *Regul Toxicol Pharmacol* 62:241–247.
- Escher BI, Allinson M, Altenburger R, Bain PA, Balaguer P, Busch W, Crago J, Denslow ND, Dopp E, Hilscherova K, et al. 2014. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. *Environ Sci Technol* 48:1940–1956.
- European Centre for Ecotoxicology and Toxicology of Chemicals. 1996. Practical concepts for dose selection in chronic toxicity and carcinogenicity studies in rodents. Brussels (BE): ECETOC. Monograph 25, 0773-6347-25. 55 p.

- Federal Register Notice. 2015. Use of high throughput assays and computational tools; Endocrine Disruptor Screening Program; Notice of availability and opportunity for comment. 80 FR 35350. EPA-HQ-OPPT-2015-0305, FRL-9928-69. EPA 2015-15182. p 35350–35355. [cited 2015 June 19]. <https://federalregister.gov/a/2015-15182>
- Guillette Jr LJ, Brock JW, Rooney AA, Woodward AR. 1999. Serum concentrations of various environmental contaminants and their relationship to sex steroid concentrations and phallus size in juvenile American alligators. *Arch Environ Contam Toxicol* 36:447–455.
- Habibi HR, Nelson ER, Allan ER. 2012. New insights into thyroid hormone function and modulation of reproduction in goldfish. *Gen Comp Endocrinol* 175:19–26.
- Hannas BR, Wang YH, Thomson S, Kwon G, Li H, LeBlanc GA. 2011. Regulation and dysregulation of vitellogenin mRNA accumulation in daphnids (*Daphnia magna*). *Aquat Toxicol* 101:351–357.
- Holbech H, Kinnberg K, Petersen GI, Jackson P, Hylland K, Norrgren L, Bjerregaard P. 2006. Detection of endocrine disruptors: Evaluation of a Fish Sexual Development Test (FSDT). *Comp Biochem Physiol C Toxicol Pharmacol* 144:57–66.
- Interagency Coordinating Committee on the Validation of Alternative Methods and National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. 2003. ICCVAM evaluation of in vitro test methods for detecting potential endocrine disruptors: Estrogen receptor and androgen receptor binding and transcriptional activation assays. Research Triangle Park (NC): National Toxicology Program. National Institutes of Health (NIH) 03-4503.
- Jaspers VLB. 2015. Selecting the right bird model in experimental studies on endocrine disrupting chemicals. *Front Environ Sci* DOI: 10.3389/fenvs.2015.00035
- Jensen KM, Korte JJ, Kahl MD, Pasha MS, Ankley GT. 2001. Aspects of basic reproductive biology and endocrinology in the fathead minnow (*Pimephales promelas*). *Comp Biochem Physiol C Toxicol Pharmacol* 128:127–141.
- Jensen KM, Makynen EA, Kahl MD, Ankley GT. 2006. Effects of the feedlot contaminant 17 alpha-trenbolone on reproductive endocrinology of the fathead minnow. *Environ Sci Technol* 40:3112–3117.
- Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia M, Huang R, Rotroff DM, Filer DL, et al. 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high throughput screening assays for the estrogen receptor. *Toxicol Sci* 148:137–154.
- Karmaus AL, Toole CM, Filer DL, Lewis KC, Martin MT. 2016. High-throughput screening of chemical effects on steroidogenesis using H295R human adrenocortical carcinoma cells. *Toxicol Sci* 150:323–332. DOI: 10.1093/toxsci/kfw002
- Knacker T, Boettcher M, Frische T, Ruffli H, Stolzenberg HC, Teigeler M, Zok S, Braunbeck T, Schafers C. 2010. Environmental effect assessment for sexual endocrine-disrupting chemicals: Fish testing strategy. *Integr Environ Assess Manag* 6:653–662.
- Knapen D, Vergauwen L, Villeneuve DL, Ankley GT. 2015. The potential AOP networks for reproductive and developmental assay development. *Reprod Toxicol* 56:52–55.
- Kugathas S, Sumpter JP. 2011. Synthetic glucocorticoids in the environment: First results on their potential impacts on fish. *Environ Sci Technol* 45:2377–2383.
- LaLone CA, Villeneuve DL, Burgoon LD, Russom CL, Helgen HW, Berninger JP, Tietge JE, Severson MN, Cavallin JE, Ankley GT. 2013. Molecular target sequence similarity as a basis for species extrapolation to assess the risk of chemicals with known modes of action. *Aquat Toxicol* 144-145:141–154.
- LaLone CA, Villeneuve DL, Olmstead AW, Medlock EK, Kahl MD, Jensen KM, Makynen EA, Blanksma CA, Cavallin JE, Thomas LM, et al. 2012. Effects of a glucocorticoid receptor agonist, dexamethasone, on fathead minnow reproduction, growth, and development. *Environ Toxicol Chem* 31:611–622.
- Liu G, Moon TW, Metcalfe CD, Lee LE, Trudeau VL. 2005. A teleost in vitro reporter gene assay to screen for agonists of the peroxisome proliferator-activated receptors. *Environ Toxicol Chem* 24:2260–2266.
- Lopez de Alda MJ, Gil A, Paz E, Barcelo D. 2002. Occurrence and analysis of estrogens and progestogens in river sediments by liquid chromatography-electrospray-mass spectrometry. *Analyst* 127:1299–1304.
- Lubzens E, Young G, Bobe J, Cerdà J. 2010. Oogenesis in teleosts: How fish eggs are formed. *Gen Comp Endocrinol* 165:367–389.
- Maack G, Segner H. 2003. Morphological development of the gonads in zebrafish. *J Fish Biol* 62:895–906.
- Marty MS, Blankinship A, Chambers J, Constantine L, Kloas W, Kumar A, Lagadic L, Meador J, Pickford D, Schwarz T, Verslycke T. 2017. Population-relevant endpoints in the evaluation of endocrine-active substances (EAS) for ecotoxicological hazard and risk assessment. *Integr Environ Assess Manag* 13:317–330.
- Mensink BP, Kralt H, Vethaak AD, Hallers-Tjabbes CCT, Koeman JH, van Hattum B, Boon JP. 2002. Imposex induction in laboratory reared juvenile *Baccharum undatum* by tributyltin (TBT). *Environ Toxicol Pharmacol* 11:49–65.
- Mihaich EM, Schäfers C, Dreier DA, Hecker M, Ortego L, Kawashima Y, Dang Z-C, Solomon K. 2017. Challenges in assigning endocrine specific modes of action: Recommendations for researchers and regulators. *Integr Environ Assess Manag* 13:280–292.
- Milla S, Depiereux S, Kestemont P. 2011. The effects of estrogenic and androgenic endocrine disruptors on the immune system of fish: A review. *Ecotoxicology* 20:305–319.
- Mommsen TP, Vijayan MM, Moon TW. 1999. Cortisol in teleosts: Dynamics, mechanisms of action, and metabolic regulation. *Rev Fish Biol Fish* 9:211–268.
- Morthorst JE, Brande-Lavridsen N, Korsgaard B, Bjerregaard P. 2014. 17 beta-Estradiol causes abnormal development in embryos of the viviparous eelpout. *Environ Sci Technol* 48:14668–14676.
- Morthorst JE, Holbech H, Bjerregaard P. 2010. Trenbolone causes irreversible masculinization of zebrafish at environmentally relevant concentrations. *Aquat Toxicol* 98:336–343.
- Nakayama K, Inoue Y, Ikeda N, Hashizume N, Murakami H, Ishibashi T, Ikeda H, Isobe T, Kitamura S, Suzuki G. 2014. Uptake and biological effects of synthetic glucocorticoids in common carp (*Cyprinus carpio*). *Marine Poll Bull* 85:370–375.
- Nesan D, Vijayan MM. 2013. Role of glucocorticoid in developmental programming: Evidence from zebrafish. *Gen Comp Endocrinol* 181:35–44.
- [OECD] Organisation for Economic Cooperation and Development. 2000. No. 23. Guidance document on aquatic toxicity testing of difficult substances and mixtures. Paris (FR): OECD. 55 p. OECD ENV/JM/MONO(2000)06.
- [OECD] Organisation for Economic Cooperation and Development. 2010a. Workshop report on OECD countries activities regarding testing, assessment and management of endocrine disruptors. Paris (FR): OECD.
- [OECD] Organisation for Economic Cooperation and Development. 2010b. No. 123. Guidance document on the diagnosis of endocrine-related histopathology in fish gonads. Paris (FR): OECD. OECD ENV/JM/MONO(2012)34.
- [OECD] Organisation for Economic Co-operation and Development. 2012a. No. 150. Guidance document on standardized test guidelines for evaluating chemicals for endocrine disruption. Paris (FR): OECD. OECD ENV/JM/MONO(2010)14.
- [OECD] Organisation for Economic Co-operation and Development. 2012b. No. 142. Validation report (Phase 2) for the Fish Sexual Development Test for the detection of endocrine active substances. Paris (FR): OECD. OECD ENV/JM/MONO(2011)23/REV1.
- Örn S, Yamani S, Norrgren L. 2006. Comparison of vitellogenin induction, sex ratio, and gonad morphology between zebrafish and Japanese medaka after exposure to 17 α -ethinylestradiol and 17 β -trenbolone. *Arch Environ Contam Toxicol* 51:237–243.
- Parrott JL, Bjerregaard P, Brugger KE, Gray LE, Iguchi T, Kadlec SM, Weltje L, Wheeler JR. 2017. Uncertainties in biological responses that influence hazard or risk approaches to the regulation of endocrine active substances. *Integr Environ Assess Manag* 13:293–301.
- Paul-Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, Simmons SO. 2016. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast Phase I and II chemical libraries. *Toxicol Sci* 151:160–180.
- Rousseau ME, Sant KE, Borden LR, Franks DG, Hahn ME, Timme-Laragy AR. 2015. Regulation of Ahr signaling by Nrf2 during development: Effects of

- Nrf2a deficiency on PCB126 embryotoxicity in zebrafish (*Danio rerio*). *Aquat Toxicol* 167:157–171.
- Sabo-Attwood T, Blum JL, Kroll KJ, Patel V, Birkholz D, Szabo NJ, Fisher SZ, McKenna R, Campbell-Thompson M, Denslow ND. 2007. Distinct expression and activity profiles of largemouth bass (*Micropterus salmoides*) estrogen receptors in response to estradiol and nonylphenol. *J Mol Endocrinol* 39:223–237.
- Säffholm M, Norder A, Fick J, Berg C. 2012. Disrupted oogenesis in the frog *Xenopus tropicalis* after exposure to environmental progestin concentrations. *Biol Reprod* 86:126.
- Shapaugh AW, McFadden LG, Zorrilla LM, Geter DR, Stuchal LD, Sunger N, Borgert CJ. 2015. Analysis of EPA's endocrine screening battery and recommendations for further review. *Regul Toxicol Pharmacol* 72: 552–561.
- Scholz S, Renner P, Belanger SE, Busquet F, Davi R, Demeneix BA, Denny JS, Leonard M, McMaster ME, Villeneuve DL, Embry MR. 2013. Alternatives to in vivo tests to detect endocrine disrupting chemicals (EDCs) in fish and amphibians – screening for estrogen, androgen, and thyroid hormone disruption. *Crit Rev Toxicol* 43:45–72.
- Schreck CB. 2010. Stress and fish reproduction: The roles of allostasis and hormesis. *Gen Comp Endocrinol* 165:549–556.
- Schug TT, Abagyan R, Blumberg B, Collins TJ, Crews D, DeFur PL, Dickerson SM, Edwards TM, Gore AC, Guillette LJ, et al. 2013. Designing endocrine disruption out of the next generation of chemicals. *Green Chem* 15:181–198.
- Scott AP. 2013. Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical review of the evidence that steroids have biological effects. *Steroids* 78:268–281.
- Seemann F, Knigge T, Olivier S, Monsinjon T. 2015. Exogenous 17 β -oestradiol (E2) modifies thymus growth and regionalization in European sea bass *Dicentrarchus labrax*. *J Fish Biol* 86:1186–1198.
- Segner H. 2011. Moving beyond a descriptive aquatic toxicology: The value of biological process and trait information. *Aquat Toxicol* 105S:50–55.
- Sone K, Hinago M, Itamoto M, Katsu Y, Watanabe H, Urushitani H, Tooi O, Guillette LJ, Iguchi T. 2005. Effects of an androgenic growth promoter 17 beta-trenbolone on masculinization of mosquitofish (*Gambusia affinis affinis*). *Gen Comp Endocrinol* 143:151–160.
- Takahashi H. 1977. Juvenile hermaphroditism in the zebrafish, *Brachydanio rerio*. *Bull Fac Fish Hokkaido Univ* 28:57–65.
- Tell L, Shukla A, Munson L, Thosar S, Kass P, Stanton R, Needham M, Lasley B. 1999. A comparison of the effects of slow release, injectable levonorgestrel and depot medroxyprogesterone acetate on egg production in Japanese quail (*Coturnix coturnix japonica*). *J Avian Med Surg* 1:23–31.
- Thomas P, Dressing G, Pang Y, Berg H, Tubbs C, Benninghoff A, Doughty K. 2006. Progestin, estrogen and androgen G-protein coupled receptors in fish gonads. *Steroids* 71:310–316.
- [USEPA] US Environmental Protection Agency. 1998. Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report. [cited 2015 September 29]. <https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-and-testing-advisory-committee-edstac-final>
- [USEPA] US Environmental Protection Agency. 2003. Health effects division (HED) interim guidance document #G2003.02. Rodent carcinogenicity studies: Dose selection and evaluation. Washington (DC): Office of Pesticide Programs, Health Effects Division.
- [USEPA] US Environmental Protection Agency. 2013. SAP review of EDSP tier 1 screening: Assay and battery performance 2013. EPA-HQ-OPP-2013-0075.
- [USEPA] US Environmental Protection Agency. 2016. Endocrine Disruptor Screening Program (EDSP) Tier 1 Assessments. [Internet]. [cited 2016 November 21]. <https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-tier-1-assessments>
- Weyts FAA, Cohen N, Flik G, Verburg-van Kemenade BML. 1999. Interactions between the immune system and the hypothalamo-pituitary-interrenal axis in fish. *Fish Shellfish Immunol* 9:1–20.
- Wheeler JR, Panter GH, Weltje L, Thorpe KL. 2013. Test concentration setting for fish in vivo endocrine screening assays. *Chemosphere* 92:1067–1076.
- Wolf JC, Baumgartner WA, Blazer VS, Camus AC, Engelhardt JA, Fournie JW, Frasca Jr S, Groman DB, Kent ML, Khoo LH, et al. 2015. Nonlesions, misdiagnoses, missed diagnoses, and other interpretive challenges in fish histopathology studies: A guide for investigators, authors, reviewers, and readers. *Toxicol Pathol* 43:297–325.
- Zeilinger J, Steger-Hartmann T, Maser E, Goller S, Vonk R, Länge R. 2009. Effects of synthetic gestagens on fish reproduction. *Environ Toxicol Chem* 28:2663–2670.
- Zhang F, Bartels MJ, Brodeur JC, Woodburn KB. 2004. Quantitative measurement of fathead minnow vitellogenin by liquid chromatography combined with tandem mass spectrometry using a signature peptide of vitellogenin. *Environ Toxicol Chem* 23:1408–1415.
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49–54.
- Zoller AL, Kersh GJ. 2006. Estrogen induces thymic atrophy by eliminating early thymic progenitors and inhibiting proliferation of β -selected thymocytes. *J Immunol* 176:7371–7378.