Availability of New Approach Methodologies (NAMs) in the Endocrine Disruptor Screening Program (EDSP)

December 13, 2022

EPA's Office of Chemical Safety and Pollution Prevention
Office of Pesticide Programs in collaboration with
Office of Research and Development

Contents

List of Acronyms

Acronym	Description
AOP	Adverse Outcome Pathway
AR	Androgen Receptor
CERAPP	Collaborative Estrogen Receptor Activity Prediction Project
CoMPARA	Collaborative Modeling Project for Androgen Receptor Activity
EAT	Estrogen, Androgen, and Thyroid
EDSP	Endocrine Disruptor Screening Program
EDSP21	Endocrine Disruptor Screening Program for the 21st Century
EDSTAC	Endocrine Disruptor Screening and Testing Advisory Committee
EPA	U.S. Environmental Protection Agency
EPA CompTox Chemicals Dashboard	U.S. EPA Computational Toxicology Chemicals Dashboard
ER	Estrogen Receptor
ExpoCast™	Exposure Forecaster (EPA)
FFDCA	Federal Food, Drug, and Cosmetic Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FR	Federal Register
HT	High-Throughput
HTTK	High-Throughput Toxicokinetics
IATA	Integrated Approach to Testing and Assessment
IBER	Integrated Bioactivity Exposure Ratio

Acronym	Description
IVIVE	*In vitro* to *in vivo* extrapolation
NAMs	New Approach Methodologies
NIH	National Institutes of Health
OCSPP	Office of Chemical Safety and Pollution Prevention
OECD	Organization for Economic Cooperation and Development
OPP	Office of Pesticide Programs
OPPT	Office of Pollution Prevention and Toxics
ORD	Office of Research and Development (EPA)
OSCP	Office of Science Coordination and Policy (OCSPP)
OSRI	Other Scientifically Relevant Information
OW	Office of Water
QSAR	Quantitative Structure-Activity Relationship
SAP	Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel
SDWA	Safe Drinking Water Act
SeqAPASS	Sequence Alignment to Predict Across Species Susceptibility tool
ToxCast	Toxicity Forecaster (EPA)
TSCA	Toxic Substances Control Act
WoE	Weight-of-Evidence

Executive Summary

Section 408(p)(1) of the Federal, Food, Drug and Cosmetic Act (FFDCA) requires EPA to "develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or such other endocrine effects as [EPA] may designate." Pursuant to that authority, in 1998, EPA introduced the Endocrine Disrupter Screening Program (EDSP) including the use of a two-tiered *in vitro* and *in vivo* screening framework (63 FR 42852 and 63 FR 71542). EPA validated a battery of five (5) *in vitro* and six (6) *in vivo* assays (EDSP Tier 1 battery) for screening chemicals in 2008 (74 FR 54415). The purpose of Tier 1 screening is to identify chemicals that have potential biological activity ("bioactivity") in the estrogen, androgen or thyroid hormone pathways using a battery of assays. Tier 1 screening data is subjected to a weight-of-evidence (WoE) analysis where an assessment is made on the need for Tier 2 testing. The purpose of Tier 2 testing is to identify and establish a dose-response relationship for any adverse endocrine (estrogen, androgen, or thyroid) effects.

The chemical substances covered by the EDSP include approximately 1,200 pesticide active ingredients, 2,500 pesticide inert ingredients, and 6,000 drinking water contaminants, with some overlap between these lists. Because of the cost (EPA estimated industry costs in conducting a full Tier 1 battery to be approximately $1 million per chemical, largely due to the *in vivo* laboratory animal testing, 78 FR 35903) and time (up to six years) involved in conducting and reviewing the full battery of Tier 1 assays, EPA has been able to evaluate only a fraction of the thousands of chemicals that fall within the scope of the EDSP.

For more than a decade at EPA, research efforts have focused on the development and evaluation of high-throughput (HT) *in vitro* assays and *in silico* methods as new approach methodologies (NAMs), including databases and computational models, for use as alternatives to the current suite of assays in the EDSP Tier 1 battery to accelerate the pace of screening, add efficiencies, decrease costs, and reduce animal testing. NAMs refer to any technology, methodology, approach, or combination of these that can provide information on chemical hazard and risk assessment while limiting/optimizing the use of animal testing (U.S. EPA, 2018b). This effort has been supported by the Office of Research and Development (ORD) and the Office of Chemical Safety and Pollution Prevention (OCSPP) (Thomas et al., 2019; U.S. EPA, 2019; NRC, 2007) along with collaboration with the National Institute of Environmental Health Sciences (NIEHS). Test method validation is a process based on scientifically sound principles by which the reliability and relevance of a particular test, approach, method, or process are established for a specific purpose (OECD, 2005). A validated test approach can be used as an alternative for a Tier 1 assay. Other scientifically relevant information (OSRI) may also be considered in determining whether the information needs provided by a Tier 1 assay are satisfied but may not always fulfill those data needs. This White Paper announces that certain NAMs have been validated by EPA and may now be accepted by EPA as alternatives for certain EDSP Tier 1 assays, while other NAMs are useful for prioritization under the EDSP and for consideration as OSRI (U.S. EPA, 2009e) in WoE evaluations. EPA will consider the strengths, limitations, and uncertainties of the NAMs in combination with the existing assays in the EDSP tiered-framework and other potential OSRI

(U.S. EPA, 2009e), (*e.g.*, exposure data, physical-chemical properties, toxicologically relevant studies in the published literature, QSAR models and other data submitted to support chemical assessment), as part of the WoE approach (U.S. EPA, 2011b) to determine whether EDSP Tier 2 testing is necessary. This White Paper provides further details below concerning when and how the NAMs outlined may be used.

The following terms are defined below for use in this document and are consistent with EDSP materials, including the December 1998 Federal Register Notice (63 FR 71542).

- ***Priority setting*** is defined as the collection, evaluation, and analysis of relevant information, including results of HT screening, to determine the general order in which chemical substances or mixtures will be subjected to screening and testing.
- ***Screening*** is defined as the application of short-term assays to determine whether a chemical substance or mixture has the potential to interact with the endocrine system. As these are preliminary assays, a positive result does not mean that a chemical substance may have an effect in humans, fish or wildlife that is similar to the effect produced by naturally occurring hormones. Screening, identified as a "Tier 1" process, is mandatory for all pesticides (active ingredients and inerts) under FFDCA Section 408(p)(3) and discretionary for other chemicals that are regulated under SDWA, although exemptions for pesticides can be granted under FFDCA Section 408(p)(4).
- ***Testing*** is defined as a customized combination of assays and endpoints designed to determine whether a chemical substance or mixture may cause adverse effects in humans, fish, or wildlife similar to the effects produced by naturally occurring hormones (estrogens, androgens, and thyroid hormones). Tests are designed to confirm and further define the results obtained in Tier 1 screens by identifying and establishing a dose-response relationship for any adverse effects that might result from the potential endocrine (estrogen, androgen, or thyroid) bioactivity identified through the Tier 1 assays (Table 1), accepted alternatives to the Tier 1 assays, or OSRI. Where Tier 1 results, accepted alternatives, or OSRI indicate a potential for affecting estrogen, androgen and thyroid (EAT) activity, Tier 2 testing is necessary to determine that a substance may have an effect similar to that of a naturally occurring hormone.
- ***Weight of evidence (WoE)*** is the process by which the strengths and weaknesses of a collection of information is judged to render an overall conclusion that may not be evident from consideration of the individual data (76 FR 60022). More specifically, WoE is conducted as part of the evaluation of EDSP Tier 1 screening data to identify the need for Tier 2 testing (U.S. EPA, 2011b). In conducting WoE, EPA may consider and use relevant information besides Tier 1 data (*e.g.*, EDSP NAMs data, exposure data, physical-chemical properties, toxicologically relevant studies in the published literature, QSAR models and other data submitted to support chemical assessment) to determine whether any Tier 2 tests are needed.

As with any cutting-edge science, there has been a vibrant scientific discussion that reflects a diversity of viewpoints, including within EPA, of which NAMs could be considered validated and which should only be used as OSRI during WoE evaluations. EPA welcomes comments on the approach suggested in this document.

Moving forward, EPA will refine the list below as appropriate, using public comments received on this White Paper, and will continue work to validate additional NAMs as alternatives to Tier 1 screens.

NAMs Validated for Screening. EPA has determined that the following NAMs may be used as alternatives for the following four (4) EDSP Tier 1 screening assays when evaluated on a chemical-by-chemical basis. For some chemicals, these NAMs cannot yet serve as an alternative for some or all of the four screening assays (for example, pesticides that have limited solubility or high volatility have been shown to be difficult to test). EPA will evaluate the properties of each chemical and the strengths and limitations of each NAM to determine whether the NAM may serve as an alternative appropriate for screening a specific chemical. Specifically, the potential use of the Estrogen Receptor (ER) pathway model will be evaluated separately for each of the three screening assays. As part of the chemical data evaluation processes, EPA reviews the quality of data available across multiple lines of evidence. For the NAMs methods listed here, EPA will specifically consider the quality of the individual ToxCast assay data and the level of confidence in and biological relevance of the predictions.

(1) The Estrogen Receptor (ER) pathway model based on the full 18-assay ToxCast/Tox21 battery (Browne et al., 2015; Judson et al., 2015) (referred to in this document as the full ER pathway model) may be used as an alternative to performing all three (3) current EDSP Tier 1 screening assays:

- ER binding *in vitro* assay (OCSPP 890.1250; (U.S. EPA, 2009b))
- ER transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300; (U.S. EPA, 2009c))
- *In vivo* Uterotrophic assay (rat) (OCSPP 890.1600; (U.S. EPA, 2009d))

(2) The Androgen Receptor (AR) pathway model based on the full 11-assay ToxCast/Tox21 battery (Kleinstreuer et al., 2017) (referred to in this document as the full AR pathway model) may be used as an alternative for one current EDSP Tier 1 screening assay:

- AR binding *in vitro* assay (OCSPP 890.1150; (U.S. EPA, 2009a)).

As explained in more detail in this White Paper, EPA considers these two NAMs validated for screening. Both models were reviewed by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) and may now be considered as validated alternatives to the four EDSP Tier 1 assays. Thus, for example, data derived from these validated models and assays may be used in the registration and registration review processes and may be determined to satisfy the specified EDSP data needs, depending on the properties of a pesticide.

NAMs Acceptable for Priority Setting and WoE Analysis. Priority setting is important for EPA to test the chemicals posing the greatest risk first, as best as can be determined through the use of available (or easily generated) exposure and hazard data.

As EPA prioritizes which chemicals go through Tier 1 screening first, EPA will use all information available including information from data submitted to support chemical assessment and peer-reviewed sources relating to chemical exposure and potential endocrine

bioactivity. NAMs used for prioritization do not need to be validated methods. EPA will use NAMs and other tools (listed below), in addition to those described as alternatives for EDSP Tier 1 screening assays, for priority setting of chemicals and for consideration as OSRI, where appropriate, in WoE evaluations. These tools are especially useful for prioritizing thousands of chemicals for which other sources of bioactivity and exposure information do not exist.

While work continues, the following four NAMs are not yet accepted by the EPA as validated alternatives for Tier 1 screening assays. Recognizing the potential for uncertainties and limitations, these models and assays may be used for priority setting of large sets of chemicals for EDSP Tier 1 screening or for consideration as OSRI (in combination with additional information) in WoE evaluations:

(1) ER and AR pathway models using assay subsets (also referred to as 'reduced or minimal assay data sets;' see citations below)
    (a) ER agonist assay subset pathway models (Judson et al., 2017).
    (b) AR agonist and antagonist assay subset pathway models (Judson et al., 2020).

(2) *In Silico* Qualitative Structure Activity Relationship Consensus Models for ER and AR (Mansouri et al., 2020; Mansouri et al., 2016). Available in the OPERA tool.

(3) Integration of Bioactivity and Exposure (Integrated Bioactivity Exposure Ratio, IBER), which compares an estimated external dose threshold for a biological effect, based on an internal dose (*i.e.*, plasma concentration) derived from bioactivity data (*e.g.*, ER and AR pathway model outputs), with estimates of exposure. (Friedman et al., 2020; Thomas et al., 2019; Bell et al., 2018; Wambaugh et al., 2018; Sipes et al., 2017; Wetmore, 2015; U.S. EPA, 2014c; Wetmore et al., 2012; Rotroff et al., 2010).

(4) The Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool for interspecies extrapolation (Lalone et al., 2018; Ankley et al., 2016). The tool provides information that can be used to understand how broadly screening data (*e.g.*, from ER pathway model assays) or adverse outcome pathways (AOPs) may plausibly be extrapolated across species and taxa. For example, SeqAPASS could be used to extrapolate mammalian ER *in vitro* bioactivity data to predict potential susceptibility of non-mammalian species (Ankley et al., 2016).

Progress Updates and Future Directions

EPA is also providing a status update regarding various NAM tools under development including HT approaches to assess disruption of steroidogenesis and thyroid pathways and additional development of SeqAPASS for the androgen receptor and thyroid-related targets. EPA is making no conclusions about their potential utility as prioritization or screening tools under the EDSP tiered framework at the current time. However, EPA might consider their use in conjunction with other data in a WoE framework. This paper does not address alternatives to the current Tier 2 tests, and there are no plans to offer alternatives for Tier 2 tests within the next few years.

A summary of the conclusions of this document is as follows:

- The full estrogen receptor (ER) and androgen receptor (AR) pathway models have been validated, and the results from those models may be used as alternatives at this time for some Tier 1 assays (ER binding, estrogen receptor transcriptional activation (ERTA), and Uterotrophic [ER pathway model] and AR binding [AR pathway model]). For any particular chemical, the suitability of a model will be decided on a case-by-case basis considering the limitations of the models (see Section III.E.) and the properties of the chemical.
- All the New Approach Methods (NAMs)/tools discussed in this paper (including full ER and AR pathway models, reduced ER and AR pathway models, Integration of Bioactivity and Exposure (IBER), Collaborative Estrogen Receptor Activity Prediction Project (CERAPP), and Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA)) may be used directly to prioritize chemicals for screening or to inform prioritization or hazard assessment (Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)).
- In some cases (considering the limitations of the model, additional available information, and EPA's guidance on Other Scientifically Relevant Information (OSRI)), the following NAMs may be considered OSRI and used during Weight of Evidence (WoE) evaluation to make decisions: reduced ER and AR pathway models, CERAPP, CoMPARA, SeqAPASS, and IBER.
- During WoE evaluation, which precedes Tier 2 testing, results from the Tier 1 battery, appropriate NAM alternatives, and OSRI are considered to determine which, if any, Tier 2 tests should be conducted. Thus, this WoE occurs between Tier 1 screening and any Tier 2 testing.
- None of these NAMs is meant to be alternatives for the current Tier 2 tests.
- The Agency will consider all public comments received on this document as it begins to implement these new approaches.

I. Introduction

A. Historical Framework for EDSP

In 1996, Congress amended the FFDCA (section 408(p)) which requires EPA to develop a 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 such other endocrine effects as [EPA] may designate’’ (FFDCA section 408(p) (21 U.S.C. 346a(p))). When carrying out the program, EPA ‘‘shall provide for the testing of all pesticide chemicals’’ and ‘‘may provide for the testing of any other substance that may have an effect that is cumulative to an effect of a pesticide chemical if the Administrator determines that a substantial population may be exposed to such a substance’’ (21 U.S.C. 346a(p)(3)). In addition, Congress amended the Safe Drinking Water Act (SDWA) and gave EPA authority to provide for the testing of endocrine disrupting effects ‘‘of any other substance that may be found in sources of drinking water if the Administrator determines that a substantial population may be exposed to such substance’’ (SDWA Amendments of 1996, section 136 (42 U.S.C. 300j–17)).

In 1996, EPA convened the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), which was chartered under the Federal Advisory Committee Act (5 U.S.C. App.2, 9(c)), to make recommendations on how to develop the screening program mandated by Congress. The EDSTAC was comprised of members representing the commercial chemical and pesticide industries, Federal and State agencies, worker protection and labor organizations, environmental and public health groups, and academic research scientists.

EPA largely adopted the EDSTAC recommendations and proposed the basic components of the EDSP in a *Federal Register* notice issued August 11, 1998 (63 FR 42852) (FRL-6021-3). After public comments, external consultations and scientific peer review, EPA provided additional details in a second *Federal Register* notice on December 28, 1998 (63 FR 71542) (FRL-6052-9). The design of the EDSP was based on the recommendations of the EDSTAC (U.S. EPA, 1998):

- Address both potential human and ecological effects from chemical exposures
- Focus examination of effects of these chemicals on estrogen, androgen, and thyroid hormone-related processes
- Include pesticide and non-pesticide chemicals, contaminants, and mixtures (after evaluating single chemicals)
- Develop a two-tiered screening and testing strategy

The following terms are defined below for use in this document and are consistent with EDSP materials, including the December 1998 Federal Register Notice (63 FR 71542).

- ***Priority setting*** is defined as the collection, evaluation, and analysis of relevant information, including results of HT screening, to determine the general order in which chemical substances or mixtures will be subjected to screening and testing.
- ***Screening*** is defined as the application of short-term assays to determine whether a chemical substance or mixture has the potential to interact with the endocrine system.

As these are preliminary assays, a positive result does not mean that a chemical substance may have an effect in humans, fish or wildlife that is similar to the effect produced by naturally occurring hormones. Screening, identified as a "Tier 1" process, is mandatory for all pesticides (active ingredients and inerts) under FFDCA Section 408(p)(3) and discretionary for other chemicals that are regulated under SDWA, although exemptions for pesticides can be granted under FFDCA Section 408(p)(4).

- ***Testing*** is defined as a customized combination of assays and endpoints designed to determine whether a chemical substance or mixture may cause adverse effects in humans, fish, or wildlife similar to the effects produced by naturally occurring hormones (estrogens, androgens, and thyroid hormones). Tests are designed to confirm and further define the results obtained in Tier 1 screens by identifying and establishing a dose-response relationship for any adverse effects that might result from the potential endocrine (estrogen, androgen, or thyroid) bioactivity identified through the Tier 1 assays (Table 1), accepted alternatives to the Tier 1 assays, or OSRI. Where Tier 1 results, accepted alternatives, or OSRI indicate a potential for EAT activity, Tier 2 testing is necessary to make a determination that a substance may have an effect similar to that of a naturally occurring hormone.
- ***Weight of evidence (WoE)*** is the process by which the strengths and weaknesses of a collection of information is judged to render an overall conclusion that may not be evident from consideration of the individual data (76 FR 60022). More specifically, WoE is conducted as part of the evaluation of EDSP Tier 1 screening data to identify the need for Tier 2 testing (U.S. EPA, 2011b). EPA may also conduct WoE using relevant information (*e.g.*, EDSP NAMs data, exposure data, physical-chemical properties, toxicologically relevant studies in the published literature, QSAR models and other data submitted to support chemical assessment) to determine whether any Tier 2 tests are needed.

For the EDSP, bioactivity (determined as part of Tier 1 screening) indicates that a chemical has the potential to alter endocrine function. However, confirming whether the chemical alters endocrine function and whether that altered function produces an adverse outcome cannot be determined without further testing (*e.g.*, Tier 2 or other testing). It is important not to equate a determination of a chemical's bioactivity with a determination that a chemical causes endocrine disruption. The World Health International Programme on Chemical Safety (IPCS) defines an endocrine disruptor as "an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes *adverse* [emphasis added] health effects in an intact organism, or its progeny, or (sub) populations" (WHO, 2002).

At this time, the full estrogen receptor (ER) and androgen receptor (AR) pathway models have been validated and thus the results from those models may be used as alternatives for specific Tier 1 screens. EPA will consider the strengths, limitations, and uncertainties of the NAMs in combination with the existing, validated assays in the EDSP tiered-framework and other potential OSRI (U.S. EPA, 2009e), (*e.g.*, exposure data, physical-chemical properties, toxicologically relevant studies in the published literature, QSAR models and other data submitted to support chemical assessment), as part of the WoE approach (U.S. EPA, 2011b) to determine whether additional data are needed.

B. Tier 0, Priority Setting

Because of the thousands of chemicals that might fall under the purview of FFDCA 408(p) and SDWA 1457, EDSTAC also recommended EPA establish a priority setting or Tier 0 prioritization process (EDSTAC, 1998) to determine the order in which chemical substances might undergo Tier 1 screening (see definition of Tier 1 screening (63 FR 42852). EPA may use certain NAMs described in this document as part of prioritization of chemicals, prior to EDSP Tier 1 screening. Priority setting may use NAMs singly or together with other available data and tools. These approaches are especially useful for prioritizing thousands of chemicals for which other sources of bioactivity and exposure information do not exist.

C. Tier 1 and Tier 2 Guidelines

Table 1 below identifies the Tier 1 guidelines for 5 *in vitro* and 6 *in vivo* assays, and the Tier 2 guidelines, which are all *in vivo*. As stated in the 2015 Policy Statement (U.S. EPA, 2015b) "the ultimate purpose of the EDSP is to provide information for evaluation of possible endocrine effects associated with the use of a chemical and take appropriate steps to mitigate any related risks to ensure protection of public health."

Table 1. EDSP Test Guidelines

Note: Listings are grouped by Tier and type (*in vitro* vs. *in vivo*). Each listing includes: assay/test name (test species); guideline(s); whether the data are informative of potential interaction with the estrogen receptor (E), androgen receptor (A), steroidogenesis (STR), and thyroid (THY) hormonal pathways). All Tier 2 tests are *in vivo* assays.

Assay Name	E	A	STR	THY
EDSP Tier 1 Battery (*in vitro* & *in vivo*)				
***In vitro* Assays**				
OCSPP 890.1250 – Estrogen Receptor Binding (Rat Uterine Cytosol	■			
OCSPP 890.1300 – Estrogen Receptor Transcriptional Activation (Human Cell Line HeLa-9903)	■			
OCSPP 890.1150 – Androgen Receptor Binding (Rat Prostate Cytosol)		■		
OCSPP 890.1550 – Steroidogenesis (Human Cell Line – H295R)			■	
OCSPP 890.1200 – Aromatase (Human Recombinant)			■	
***In vivo* Assays**				
OCSPP 890.1600 – Uterotrophic (Rat)	■			
OCSPP 890.1400 – Hershberger (Rat)		■	■[1]	
OCSPP 890.1450 – Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats	■		■	■
OCSPP 890.1500 – Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats		■	■	■
OCSPP 890.1100 – Amphibian Metamorphosis Assay (Frog)				■
OCSPP 890.1350 – Fish Short-Term Reproduction Assay	■	■	■	

Assay Name	E	A	STR	THY
EDSP Tier 2 Tests (all *in vivo*)[2]				
OCSPP 870.3800 – Reproduction and Fertility Effects (Rat)	■	■	■	
OECD TG 443 – Extended One-generation Reproductive Toxicity Test (Rat) (EOGRT)	■	■	■	■
OCSPP 890.2200 –Medaka Extended One-generation Reproduction Test (fish) (MEOGRT)	■	■	■	
OCSPP 890.2300 – Larval Amphibian Growth and Development Assay (frog) (LAGDA)				■
OCSPP 890.2100 – Avian Two-Generation Toxicity Test in the Japanese Quail (bird) (JQTT)	■	■	■	■

[1] 5α-reductase inhibition only
[2] EPA may request additional tests for Tier 2, *e.g.*, Special EPA Test - Comparative Thyroid Assay (U.S. EPA, 2005).

D. List 1 and 2 Chemicals

In 1999, EPA convened a joint meeting of EPA's Science Advisory Board (SAB) and Federal Insecticide, Fungicide, and Rodenticide (FIFRA) Scientific Advisory Panel (SAP) to review the proposed EDSP. In their report, the SAB/SAP recommended that EPA "*focus its initial screening efforts on a smaller and more manageable universe of chemicals that emphasizes the early attention to the pesticide chemicals that Congress specifically mandated the EPA to test for possible endocrine effects*" (U.S. EPA, 1999). The joint SAB/SAP recommended that initial screening be limited to 50 to 100 chemicals (EDSTAC, 1998).

Following the advice of the SAB/SAP, EPA set an initial course for EDSP priority setting. The number of chemicals (50 to 100) was thought to be a reasonable undertaking to allow the nascent EDSP to gain experience with the test systems before proceeding through the larger universe of chemicals. The first list of chemicals was based solely on exposure potential. In September 2005, EPA published (70 FR 56449) its approach for selecting the initial list of chemicals subject to the EDSP screening, referred to as "List 1." On April 15, 2009, EPA announced the policies and procedures for initial EDSP screening and the first list of chemicals to be screened with the Tier 1 battery, for their potential to interact with the endocrine system (74 FR 17559). This is known as Tier 1 screening. EPA began issuing these test orders for List 1 on October 29, 2009. The first group of chemicals identified for screening included pesticide active ingredients and Toxic Substances Control Act (TSCA) High Production Volume (HPV) chemicals also used as pesticide inert ingredients (also termed "other ingredients").

To identify the specific chemicals that would comprise List 1, EPA applied the EDSTAC exposure recommendations (EDSTAC, 1998), focusing on four pathways of human exposure to pesticide active ingredients:

1. Consumption of food containing pesticide residues (food pathway)
2. Consumption of drinking water containing pesticide residues (water pathway)
3. Residential use of pesticide products (residential use pathway)
4. Occupational contact with pesticide–treated surfaces (occupational exposure pathway).

For HPV pesticide inert ingredients, EPA focused on:

1. Human biological samples (human biological monitoring pathway)
2. Animal tissues that have human food uses (ecological biological pathway, *e.g.,* fish tissues)
3. Drinking water (drinking water pathway)
4. Indoor air (indoor air pathway)

Based on these exposure considerations and public review, EPA began issuing Tier 1 test orders for a list of 67 pesticide chemicals (58 active ingredients and 9 HPV inert ingredients) for List 1 in 2009. Registrations for 15 of the 67 chemicals were subsequently canceled or discontinued by the pesticide registrant and are no longer in use. The registrants' responses to the test orders for the remaining 52 pesticides, including conducting Tier 1 assays and EPA review of the Tier 1 data (including OSRI), and drafting the EDSP Tier 1 WoE evaluations for List 1 chemicals, took approximately 6 years. In 2015, EPA completed WoE screening determinations for the 52 supported List 1 chemicals (50 pesticidal active ingredient and 2 inert ingredients). The Tier 1 WoE screening results for these 52 List 1 chemicals are determinations of their potential to impact endocrine function and should not be construed as meaning that EPA concluded any chemical was an endocrine disruptor. Of the 52 chemicals evaluated, there was no evidence for potential interaction with any of the endocrine pathways for 20 chemicals, and for 14 chemicals that showed potential interaction with one or more pathways, EPA already has enough information to conclude that they did not need Tier 2 EDSP testing. Based on the WoE results, EPA identified 18 (of 52) List 1 chemicals as potentially needing EDSP Tier 2 testing (U.S. EPA, 2021b) or other OSRI.

In November 2012, EPA published the document, "Endocrine Disruptor Screening Program Universe of Chemicals and General Validation Principles," which identifies approximately 10,000 substances for EDSP priority setting and screening (U.S. EPA, 2012), based on the authorities of the FFDCA and SDWA. In June 2013, EPA published a second list (List 2) of 109 chemicals prioritized for Tier 1 screening under the EDSP, which included pesticide active and inert ingredients, and contaminants that may be found in drinking water (78 FR 35922). In developing the List 2 chemicals for EDSP screening, EPA focused on continuing to address pesticides and beginning to address drinking water contaminants. Pesticides on List 2 represented those scheduled for Registration Review during fiscal years 2007 and 2008. EPA identified drinking water contaminants for List 2 from chemicals that were regulated under the national primary drinking water regulation (NPDWR) (40 CFR part 141) under the SDWA or were unregulated contaminants on the third Contaminant Candidate List (CCL 3). Based on these considerations, the final List 2 for EDSP Tier 1 screening, consisting of 107 chemicals (41 pesticide active ingredients and 86 SDWA chemicals, some of which are overlapping) was published, although test orders have not been issued for these chemicals (78 FR 35922).

Of the approximately 10,000 EDSP substances, including pesticide active ingredients and inerts (covered under FIFRA) and chemicals found in sources of drinking water (covered under SDWA), only 67 (List 1) and 107 (List 2) chemicals have been prioritized for screening and potential testing to date. Based on the current pace of the Tier 1 screening assays, it could take decades to screen all 10,000 chemicals in the EDSP domain. Therefore, EPA's EDSP is actively pursuing the application of computational toxicology and exposure estimation methods using NAMs to create a more efficient and robust screening program. Incorporating innovative

computational toxicological tools allows EPA to integrate bioactivity and exposure to prioritize and screen chemicals and is consistent with the approach originally recommended by EDSTAC in 1998 (EDSTAC, 1998). Advances in computational toxicology have brought EPA to an "evolutionary turning point" or "pivot" for EDSP priority setting, screening, and testing (U.S. EPA, 2015b). In addition to rapidly screening thousands of chemicals and overcoming throughput limitations of traditional chemical toxicity testing, computational toxicology methods are helping EPA to offer NAMs as alternatives for some EDSP Tier 1 assays.

II. Approach for Advancing the EDSP

From 1996 to 2012, EPA invested significant resources in developing and validating Tier 1 screening and Tier 2 testing guidelines for use in the EDSP. Concurrently, the need for a more comprehensive review of new, state-of-the-science technologies for toxicity testing was also recognized. EPA requested the National Research Council develop a strategy for implementation of toxicity testing. Following the 2007 publication of *Toxicity Testing in the 21st Century: A Vision and Strategy* (NRC, 2007), EPA increased its focus on development and evaluation of newer technologies to accelerate the pace of screening and testing and reduce reliance on more resource intensive animal-based toxicity testing. In 2007, for example, ORD launched the Toxicity Forecaster (ToxCast) program to prioritize and screen chemicals (Dix et al., 2007). ToxCast is a broadly based HT screening program that generates data on a variety of chemicals and their impact on important biological processes; a subset of methods and data from this effort is directly relevant to the EDSP. ToxCast (as of the August 2020 release of *invitroDB* version 3.3) has ER and AR pathway model results for over 1,800 chemicals from a broad range of sources including pesticides, industrial and consumer products, food additives, pharmaceuticals, and other chemicals of interest for the development of predictive toxicology methods (Richard et al., 2016). ToxCast provides a means of quickly and efficiently prioritizing and/or screening large numbers of chemicals for endocrine bioactivity, and can minimize the number of required, time-consuming laboratory animal-based toxicity tests.

EPA is also a leader in the Toxicology in the 21st Century (Tox21) federal agency collaboration whose purpose is to develop and evaluate NAMs as alternatives to animal-intensive toxicity studies (Thomas et al., 2018). Tox21 pools resources and expertise from EPA, the National Toxicology Program (NTP), National Center for Advancing Translational Sciences (NCATS), and the Food and Drug Administration (FDA) to apply HT screening to thousands of chemicals for potential bioactivity. Together EPA ToxCast and multi-Agency Tox21 programs have developed and evaluated NAMs and determined how chemicals are evaluated for effects on both human health and the environment. Current chemical domain includes ~2000 chemicals studied in >800 assays representing over 400 biological targets and pathways, and an even larger set of >8000 chemicals have been tested in a subset of these assays (for access to data, see EPA CompTox Chemicals Dashboard and Tox21 Toolbox. These data can also be found at NTP's Integrated Chemical Environment which has a curated version of the HTS data that includes chemical QC information and mapping of assays to mechanistic targets. The CompTox Chemicals Dashboard (Williams et al., 2017), and InvitroDB are living databases and models that are regularly updated with new information from EPA research and external stakeholders.

In fiscal year 2012, EPA began a multi-year transition to evaluate, validate, and incorporate computational toxicology methods and HT screens to serve as alternatives for current Tier 1 screening assays (U.S. EPA, 2012). EPA efforts to develop and evaluate EDSP NAMs have been jointly coordinated by ORD and OCSPP. In September 2011, EPA published the *Endocrine Disruptors Screening Program for the 21st Century Workplan* (U.S. EPA, 2012) ("EDSP21 Work Plan"). The Workplan advances continued efforts for the EDSP to reflect the current state of the science for evaluating potential effects on endocrine-mediated processes. Based on the EDSP21 Workplan, EPA began a multi-year transition to validate and more efficiently use computational toxicology methods and HT *in vitro* assays (primarily from EPA's ToxCast program). The EDSP21 Work Plan proposes to use a multi-level and integrated approach to determine whether a chemical has the potential to interact with specific endocrine signaling pathways. The near-term goal relied on computational methods to prioritize chemicals for screening. The intermediate-term goal involved replacing current validated *in vitro* screening (Tier 1) assays with validated *in vitro* HT assays. The results of this effort would also inform efforts to replace current *in vivo* Tier 1 assays. The long-term goal was to replace all current Tier 1 screening assays by incorporating advances in computational modeling and molecular biology and using robotics for conducting rapid, low-cost, non-animal assays on hundreds to thousands of chemicals. Progress on this effort was provided in the 2014 EDSP Comprehensive Management Plan.

In a *Federal Register* notice (U.S. EPA, 2015a), EPA announced its intention to "pivot" towards the use of validated HT assays and *in silico* models that can serve as alternatives for some of the current assays in the EDSP Tier 1 battery and requested public comment (U.S. EPA, 2015b). EPA requested comment on its intention to use HT assay and computational tools in the EDSP as well as on the use of 18 ER HT *in vitro* assays and ER pathway model as validated alternatives to three of the 11 Tier 1 screening assays: ER binding *in vitro* assay using rat uterine cytosol (OCSPP 890.1250), ER transcriptional activation (ERTA) *in vitro* assay (Human Cell Line HeLa-9903) (OCSPP 890.1300; OECD No. 455), and Uterotrophic *in vivo* assay in rat (OCSPP 890.1600; OECD No. 440). EPA received comments from 12 groups/individuals generally supportive of EPA's transition to the use of HT assays and computational models as alternatives to assays in the EDSP Tier 1 battery (U.S. EPA, 2015b, see comments in https://www.regulations.gov/docket/EPA-HQ-OPPT-2015-0305). EPA prepared a response to comments document that has been added to the public docket for the 2015 Federal Register Notice (U.S. EPA, 2022b, see https://www.regulations.gov/docket/EPA-HQ-OPPT-2015-0305).

Subsequent to this 2015 *Federal Register* notice, EPA has worked on the further development and evaluation of alternatives to the existing EDSP Tier 1 screening battery. Much of this work is the subject of this document. Although the science will continue to evolve and additional NAMs will be available in the future, EPA has determined that some NAMs addressed in this White Paper may now be considered as alternatives in satisfying specified Tier 1 assays (see **Table 2** below). Table 2 also highlights potential future efforts for NAM development of Tier 1 and Tier 2 alternatives. Additional NAMs may also be used for priority setting in the EDSP and may be considered as OSRI in a WoE approach as appropriate (U.S. EPA, 2011b, 2009e).

Table 2. EDSP Test Guidelines with Validated Alternatives

Each listing includes: assay/test name (test species); type of assay (in vitro or in vivo); and validated alternative status. All Tier 2 tests are *in vivo* assays. ER = estrogen receptor; AR = androgen receptor; STR = steroidogenesis; THY = thyroid.

EDSP Tier 1 Battery	Type	Tier 1 Battery Alternatives	Animals Used Per Assay [1]
OCSPP 890.1250 – Estrogen Receptor Binding (Rat Uterine Cytosol)	**In vitro**	**ER Pathway Model**	**13 (0)**
OCSPP 890.1300 – Estrogen Receptor Transcriptional Activation (Human Cell Line HeLa-9903)	**In vitro**	**ER Pathway Model**	**0**
OCSPP 890.1600 – Uterotrophic (Rat)	**In vivo**	**ER Pathway Model**	**18**
OCSPP 890.1150 – Androgen Receptor Binding (Rat Prostate Cytosol)	**In vitro**	**AR Pathway Model**	**10 (0)**
OCSPP 890.1200 – Aromatase (Human Recombinant)	In vitro	STR Model (Future)	0
OCSPP 890.1550 – Steroidogenesis (Human Cell Line – H295R)	In vitro	STR Model (Future)	0
OCSPP 890.1400 – Hershberger (Rat)	In vivo	AR/STR Model (Future)	48
OCSPP 890.1450 – Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Female Rats	In vivo	ER, STR, THY Models (Future)	45
OCSPP 890.1500 – Pubertal Development and Thyroid Function in Intact Juvenile/Peripubertal Male Rats	In vivo	AR, STR, THY Models (Future)	45
OCSPP 890.1350 – Fish Short-Term Reproduction Assay	In vivo	ER, AR, STR Models (Future)	96
OCSPP 890.1100 – Amphibian Metamorphosis Assay (Frog)	In vivo	THY Model (Future)	320
EDSP Tier 2 Tests [2]	**Type**	**Tier 2 Test Alternatives**	**Animals Used Per Assay**
OCSPP 870.3800 – Reproduction and Fertility Effects (Rat)	In vivo	ER, AR, STR, THY (Future)	224 or more
OECD TG 443 – Extended One-generation Reproductive Toxicity Test (Rat) (EOGRT)	In vivo	ER, AR, STR, THY (Future)	100
OCSPP 890.2200 –Medaka Extended One-generation Reproduction Test (fish) (MEOGRT)	In vivo	ER, AR, STR (Future)	126
OCSPP 890.2300 – Larval Amphibian Growth and Development Assay (frog) (LAGDA)	In vivo	THY (Future)	24
OCSPP 890.2100 – Avian Two-Generation Toxicity Test in the Japanese Quail (bird) (JQTT)	In vivo	ER, AR, STR, THY (Future)	480 [3]

[1] Certain Tier 1 results were obtained from data presented in Table 1 of Bishop and Willett (Bishop and Willett, 2014). Tier 2 results were calculated based on the guidelines. The EPA NAMs Workplan (U.S. EPA, 2021a) only counts "intact animals" for computing animal usage. In that case the animal count would be 0 (as shown in parentheses). However, animals are sacrificed for organ use in these screens.

[2] EPA may request additional Tier 2 tests, *e.g.*, Special EPA Test - Comparative Thyroid Assay (U.S. EPA, 2005).

[3] Estimated number including adults and hatchlings for two generations for five groups. Guideline allows some variation in number of animals used.

A. Role of the FIFRA SAP

The FIFRA SAP, a federal advisory panel mandated under FIFRA Section 25(d), provides scientific advice on methods used in assessing exposure and effects of pesticides. Independent peer review is an important aspect in reviewing methodologies used in chemical risk assessment. EPA provides public notice of when FIFRA SAP meetings will occur and requests public comment. Table 3 provides a summary of the FIFRA SAP scientific topics associated with the use of HT assay and computational tools for prioritizing and screening the universe of EDSP chemicals for estrogen, androgen, and thyroid bioactivity and steroidogenesis; SAP meeting reports are also cited. In parallel with this White Paper, EPA has separately responded to the 2017 SAP (U.S. EPA, 2017, see https://www.regulations.gov/document/EPA-HQ-OPP-2017-0214-0024) comments and questions (U.S. EPA, 2022a); those responses, where appropriate, are included in this White Paper.

Table 3. Summary of FIFRA SAP Meeting Topics: Scientific Issues Associated with the Use of NAMs for Prioritizing and Screening Chemicals in the EDSP

Key: ER = estrogen receptor, ERES = ER Expert System, QSAR = Quantitative Structure Activity Relationship, ExpoCast = Exposure Forecaster, SEEM = Systematic Empirical Evaluation of Models, HTTK = High Throughput Toxicokinetic, RTK = Reverse Toxicokinetic, IBER = Integrated Bioactivity Exposure Ranking, AR = Androgen Receptor, E= Estrogen, A = Androgen, T = Thyroid

FIFRA SAP Meeting Titles	Scientific Topics	FIFRA SAP Recommendations and EPA Responses
Scientific Issues associated with Prioritizing the Universe of EDSP Chemicals using Computational Toxicology Tools (January 8-10, 2013)	• Proposed prioritization scheme – use of 8 HT ER agonist assays • EPA Estrogen Receptor Expert System (ERES) Quantitative Structural Activity Relationship (QSAR) • Physical-chemical properties to exclude certain chemical substances from EDSP	FIFRA SAP Report (U.S. EPA, 2013, see https://www.regulations.gov/document/EPA-HQ-OPP-2012-0818-0037). Regulatory docket: EPA-HQ-OPP-2012-0818. EPA Responses: 2015 FR Notice (U.S. EPA, 2015a) and the various tools in this document.
Scientific issues associated with New HT Methods to Estimate Chemical Exposure (July 29 – August 1, 2014)	• EPA ExpoCast (Exposure Forecaster) • Systematic Empirical Evaluation of Models (SEEM) Framework • HT Toxicokinetics/Reverse Toxicokinetics (HTTK/RTK) for *in vitro* to *in vivo* extrapolation	FIFRA SAP Report (U.S. EPA, 2014c, see https://www.epa.gov/sites/default/files/2015-06/documents/072914minutes.pdf) Regulatory docket: EPA-HQ-OPP-2014-0331

FIFRA SAP Meeting Titles	Scientific Topics	FIFRA SAP Recommendations and EPA Responses
		EPA Responses: December 2014 White Paper (U.S. EPA, 2014a, see https://www.regulations.gov/document/EPA-HQ-OPP-2014-0614-0029) and IBER Section in this document
Scientific Issues associated with Integrated Endocrine Bioactivity and Exposure-Based Prioritization and Screening for the EDSP (December 2-5, 2014)	• Use of 18 HT ER pathway *in vitro* assays and ER pathway model as alternative methods to Tier 1 Estrogen Receptor Binding Assay Using Rat Uterine Cytosol (ER-RUC) (OCSPP 890.1250), Estrogen Receptor Transcriptional Activation (Human Cell Line HeLa-9903) (OCSPP 890.1300; OECD No. 455), and Uterotrophic Assay (Rat) (OCSPP 890.1600; OECD No.440) (included Consensus QSAR ER Bioactivity Model) • Preliminary work – use of HT AR pathway *in vitro* assays and AR pathway model as alternative method to Tier 1 AR binding assay • Prioritization using Integrated Bioactivity and Exposure Ratio (IBER).	FIFRA SAP Report (U.S. EPA, 2014a, see https://www.regulations.gov/document/EPA-HQ-OPP-2014-0614-0029). Regulatory docket: EPA-HQ - 2014-0614 EPA Responses: 2015 FR Notice (U.S. EPA, 2015a); EPA Response to Comments on the 2015 FR Notice (U.S. EPA, 2022b); Response to November 2017 SAP Recommendations (U.S. EPA, 2022a); and this Document (Section III)
Continuing Development of Alternative HT Screens to Determine Endocrine Disruption, Focusing on Androgen Receptor, Steroidogenesis, and Thyroid Pathways (November 28-29, 2017)	• Revised Androgen Receptor (AR) pathway model (11 assays) from the December 2014 FIFRA SAP as an alternative method for Tier 1 AR binding assay (OCSPP 890.1150) • New HT H295R steroidogenesis assay and use of Mahalanobis distance to integrate data from 11 hormones as an alternative method to the low-throughput Tier 1 H295R assay (OCSPP 890.1550/ OECD TG 456) • Draft thyroid adverse outcome pathway (AOP) framework to identify substances that can perturb thyroid function.	FIFRA SAP report (U.S. EPA, 2017, see https://www.regulations.gov/document/EPA-HQ-OPP-2017-0214-0024). Regulatory docket: EPA-HQ-2017-0214 EPA Responses: EPA White Paper ; Response to November 2017 SAP Recommendations (U.S. EPA, 2022a) and this Document

B. Performance-Based Approach to Establishing Confidence in NAMs

Historically, test methods have been validated according to principles described in the Organisation for Economic Co-operation and Development (OECD) *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (GD 34)* (OECD, 2005). Specifically, OECD GD 34 states that "new test methods undergo validation to assure that they employ sound science and meet regulatory needs" (*i.e.*, the methods are fit-for-purpose), "the validation process should be flexible and adaptable," and that performance must be "demonstrated using a series of reference chemicals" and "evaluated in relation to existing relevant toxicity data." OECD GD 34 further defines ***relevance*** of a test method to include, "the relationship between the test and the effect in the target species and whether the test method is meaningful and useful for a defined purpose, with the limitations identified." ***Reliability*** is defined in OECD GD 34 as the extent of reproducibility of results from a test within and among laboratories over time, when performed using the same standardized protocol.

OECD guidance states that the validation process should be "flexible and adaptable." The steps used in the performance-based validation of ER and AR pathway models meets many elements in OECD GD34 of scientifically supported data quality evaluation approaches including those for NAMs and toxicological test data. There have been several attempts by multiple national/international organizations (including OECD) to streamline the validation process and allow for more rapid adoption of reliable and relevant NAMs considering the context in which NAMs are expected to be used. Also, the validation of a NAM has historically included the generation of data in "ring-trials" involving multiple laboratories using a range of chemicals and controls in an expensive, multi-year process. The ability of multiple labs to use the same assay protocol and get similar results is an indication of the transferability of the method. The transferability of a method has historically been seen as a prerequisite for validation, despite the reality that some NAMs require specialized equipment, expertise, or intellectual property considerations. Newer, performance-based, validation approaches supplant the need for inter-laboratory ring trials with more flexible, fit-for-purpose approaches, including demonstration of reproducibility over time and assessments that consider expanded reference chemical sets.

To more flexibly accommodate the range of decision contexts and rapid pace of NAM development, multiple entities and individuals have proposed frameworks for building confidence and accelerating the use of NAMs (ICCVAM, 2018; Casati et al., 2017; Patlewicz et al., 2015; Patlewicz et al., 2013). EPA developed a set of criteria for evaluating the scientific reliability and relevance of NAMs within TSCA and presented these criteria in the TSCA Strategic Plan (U.S. EPA, 2018a). While many of the criteria in the TSCA Strategic Plan are fundamental to evaluating the reliability and relevance of NAMs, a generic framework that applies across EPA's myriad of statutes and regulations has not yet been developed. As described in EPA's NAMs Work Plan (U.S. EPA, 2021a), a scientific confidence framework for use across regulatory contexts broader than TSCA needs to be developed to evaluate the quality, reliability, and relevance of NAMs. Assessment criteria to facilitate regulatory use and international adoption were recommended by the International Cooperation on Alternative Test

Methods (ICATM) and described in Casati et al. (2017). In addition, the 2021 OECD guideline 497 includes "ANNEX 1: Evaluation Framework to the OECD Supporting Document on Defined Approaches for Skin Sensitisation" which provides criteria for establishing scientific confidence in NAMs, specifically defined approaches. The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) has recently created a new technical workgroup to evaluate existing validation and confidence building frameworks and to develop updated guidance and recommendations. OCSPP and ORD are actively engaged in the NAMs Work Plan and ICCVAM activities. The outcomes of these efforts will help further the development and implementation of NAMs for EDSP.

EPA used a performance-based approach to validate the ER and AR pathway models (see description in Section III of this paper) for potential ER and AR bioactivity. In this approach, the performance of the ER (or AR) pathway model is compared against the results for previously validated tests for the same endpoint (*e.g.*, ER binding, ER transactivation, AR binding). The robustness of the model is characterized using large sets of *in vitro* and *in vivo* reference chemicals with well-defined activities in the existing methods. A performance-based approach relies on analytical specifications such as required sensitivity, specificity, and reproducibility. As a result, this approach offers a flexible, fit-for-purpose evaluation process to building confidence in the acceptability of the ER and AR pathway models, which supports EPA's decision to validate those models.

Sensitivity, specificity, and balanced accuracy are examples of performance metrics that are important to performance-based validation. Sensitivity measures the proportion of positives that are correctly identified as such, while specificity measures the proportion of negatives that are correctly identified as such. Mathematically, sensitivity = the number of true positives ÷ (number of true positives + number of false negatives). Specificity = number of true negatives ÷ (number of true negatives + number of false positives). Balanced accuracy is the proportion of correct outcomes of a test method, or mathematically, it is the arithmetic mean of sensitivity and specificity. Reproducibility indicates that re-running the assay (or some models) will give you a similar answer.

As part of its risk assessment activities, EPA routinely evaluates significant amounts of data, including new methods and approaches, for use in hazard and exposure assessments and the readiness of new methods and models. The strengths and uncertainties of the ER and AR full pathway models are described in Section III, C and D. Further, additional analyses to address some of the test limitations are described in Section VIII. Future Direction.

In the 2015 Federal Register Notice (USEPA, 2015b), EPA had announced the plan to use the ER pathway model as an alternative to three (3) of the current EDSP Tier 1 battery of assays (ER binding assay, ERTA, and Uterotrophic assay). This FRN states:

> *"The approach incorporates validated high throughput assays and a computational model and, based on current research, can serve as an alternative for some of the current assays in the Endocrine Disruptor Screening Program (EDSP) Tier 1 battery."*

The 2015 FRN further describes the following steps taken to use a performance-based approach to validate the ER pathway model and to use as OSRI:

- The ‘‘ER Model’’ bioactivity scores were validated by comparing the scores to 45 reference chemicals, equivalent to a performance-based approach to validation.
- EPA also compared the ‘‘ER Model’’ results to a database of curated uterotrophic studies published in peer-reviewed literature. ER agonist bioactivity scores accurately predicted *in vivo* ER agonist activity for a large set (~150) of chemicals with uterotrophic data.
- The validation of the ‘‘ER Model’’ as an alternative screening method for three current Tier 1 assays (ER binding, ER transcriptional activation (ERTA), and uterotrophic) was peer reviewed by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) in December 2014.
- The FIFRA SAP fully endorsed the use of these alternatives for the ER binding and ERTA assays; however, there was not consensus among panel members on the use of the ‘‘ER Model’’ as an alternative for the uterotrophic assay.

In this White Paper, EPA announces its intent to use the results from the full ER pathway model as a fully validated alternative to the respective EDSP Tier 1 assays (ER binding, ERTA, and uterotrophic). When appropriate, the model, which integrates results from a battery of high throughput assays, can replace some or all of these low throughput assays (depending on the properties of the chemical at issue *e.g.*, solubility or volatility). Thus, even though the model has been fully validated, EPA will decide on a chemical-by-chemical basis whether the model can replace these assays. For example, these models can produce a false negative if used on an extremely volatile chemical, so EPA would not accept the data from the model for such substances.

With respect to the AR pathway model, the steps taken by EPA to validate the AR pathway model are consistent with those described in the 2015 FRN for the ER pathway model and include the following:

- The ‘‘AR pathway model’’ bioactivity scores were validated by comparing with a set of reference chemicals. To be included in this set, a chemical had to have consistent (active or inactive, agonist or antagonist) results across multiple literature reports. The final set of reference chemicals included 37 for AR agonism and 28 for AR antagonism (Kleinstreuer et al., 2017).
- The performance-based validation of the ‘‘AR pathway model’’ as an alternative screening method for one current Tier 1 assay (AR binding) was peer reviewed by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) in December 2014 and November 2017 (U.S. EPA, 2017, 2014a).
- In response to some of the recommendations of the FIFRA SAP in 2017, EPA modified the AR pathway model to include an uncertainty characterization (Judson et al., 2020; Watt and Judson, 2018).

In this White Paper, EPA announces its intent to use the results from the full AR pathway model as a fully validated alternative to the AR binding assay.

III. ER and AR HT Assays and Pathway Models

A. General Considerations

The prior two sections of this document addressed the historical framework for the EDSP and the approach for advancing the EDSP utilizing NAMs. This section will provide greater details on the strengths and limitations of the full ER and AR pathway models and the HT assays that the models use. The results of the full ER and AR pathway models are intended to be used as validated alternatives for a number of EDSP Tier 1 assays.

Screening thousands of chemicals to identify potential estrogen and androgen bioactivity could cost millions of dollars and take decades to complete using current low throughput toxicology methods. Researchers from around the world, including from EPA and NIEHS, have developed rapid and relatively inexpensive HT *in vitro* screening and computational toxicology approaches to serve as potential alternatives to low throughput assays. The ER pathway model combines the results from 18 HT assays from the ToxCast and Tox21 research programs (Judson et al., 2015). The AR pathway model combines the results from 11 HT screening assays from the ToxCast and Tox21 research programs (Kleinstreuer et al., 2017). The ER pathway model and AR pathway models were evaluated using a performance-based approach requiring the use of high-quality reference chemicals whose activity was determined by international test method validation efforts and systematic literature review. Both the ER pathway and AR pathway models have undergone extensive, external peer review by the FIFRA SAP (see Table 3) and are considered validated alternatives for certain Tier 1 assays.

While EPA recognizes that the full ER and AR pathway models have uncertainties and limitations (see III.), EPA will be using results from these models as alternatives for certain Tier 1 screening assays. EPA continues efforts to address limitations. For example, the ToxCast assays have limited to no metabolic capability**,** but the Phase I metabolism capability is in development (Deisenroth et al., 2020) and has been applied to over 700 ToxCast chemicals (Hopperstad et al., 2022). This metabolic uncertainty will be considered in WoE evaluations of each chemical's potential for estrogen and androgen bioactivity.

B. Estrogen Receptor Pathway Model

The use of HT assays and computational model approaches are consistent with the recommendations of the 2007 NRC report (NRC, 2007). This report recommended a fundamental shift from chemical safety decisions based on apical animal endpoints toward broader application of *in vitro* testing and predictive toxicology methods. EPA has moved to quantifying the perturbation of molecular events and cellular pathways using higher throughput, *in vitro* assays and integrating results across diverse chemical classes and biological endpoints using computational modeling.

The ER pathway model is a computational approach that integrates activity from 18 HT *in vitro* assays to characterize ER bioactivity (Judson et al., 2015). The ER pathway model is based on the series of molecular events that typically occur in a nuclear receptor-mediated response

(Judson et al., 2015). The 18 HT assays include biochemical and cell-based *in vitro* assays that evaluate perturbations of ER pathway responses at key activity sites within the cell including ER receptor binding, ER receptor dimerization, chromatin binding of the mature transcription factor, gene transcription, and changes in ER-dependent cell proliferation (Judson et al., 2015). The ER pathway model integrates the activity patterns across the 18 assays to predict whether a chemical is a potential ER agonist or antagonist and whether there is "assay interference" or cytotoxicity. Assay interference refers to activity in an assay that is likely not due to interaction of the chemical with its intended target (*e.g.,* ER) or assay endpoint. Assay interference is a phenomenon whereby assays designed to measure binding to a protein or perturbation of a given pathway may produce false signals when the target protein itself, or other pathways in the system, are altered non-specifically. For instance, a chemical could cause protein denaturation, which could give rise to a false positive signal in cell-free, radioligand competitive-binding assays.

The output of the ER pathway model provides an area under the curve (AUC) value for the potential of a chemical to cause ER agonism and/or ER antagonism. The AUC scores represent the chemical or concentration-specific probabilities that the chemical is interacting with the corresponding receptor. AUC scores are scaled to activity and range from 0 to 1 (*e.g.*, AUC (ER agonist) = 1 for 17α-ethinylestradiol). ER Pathway AUC (agonist) scores ≥0.1 are considered "active," while AUC (agonist) values between 0.01 and 0.1 may indicate weak or ambiguous potential activity and are considered inconclusive (Judson et al., 2015). The AUC score is highly correlated with the logarithm of the chemical potency against the receptor (ER or AR).

Full concentration-response data for all 18 HT ER pathway assays were collected on 1,812 chemicals (Judson et al., 2015). Of the 1,812 chemicals evaluated, 111 (6.1%) were predicted to be strongly ER active in agonist and/or antagonist mode. The ER pathway model was constructed to assess assay interference, including cytotoxicity (which is the most prominent cause of false-positive activity), and is accounted for within the ER pathway modeling scores (Watt and Judson, 2018; Judson et al., 2015). Cytotoxicity in cell-based assays may confound receptor antagonism in particular (*i.e.*, it may be difficult to distinguish the source of a decreased assay signal, resulting in higher false positive rate in those particular assays). Additional analysis on the effect of cytotoxicity on HT screening data has been reported (Judson et al., 2016), as have the results on the variability in the curve-fitting for the HT screening assays (Watt and Judson, 2018). Cytotoxicity was measured using a collection of 35 assays in the ToxCast battery that detect cytotoxicity or other forms of cell loss across several cell lines and primary cell types (Judson et al., 2016).

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) constructed a database of high-quality guideline or guideline-like rodent Uterotrophic studies (OCSPP 890.1600 and OECD 440 Test Guidelines OECD, 2018b; U.S. EPA, 2011a) (Kleinstreuer et al., 2016). The Uterotrophic bioassay is a short-term *in vivo* screen for estrogenicity and is one of the 11 EDSP Tier 1 screening assays (see Table 1). EPA EDSP Uterotrophic assay, OCSPP 890.1600, is performed on immature female rats or ovariectomized (OVX) adult female mice or rats (U.S. EPA, 2009d), and is a measure of *in vivo* estrogen receptor agonism in female mammals without an intact hypothalamic-pituitary-gonadal axis or in pre-pubertal animals. A comparison between the

results from the ER pathway model and the curated NICEATM database of *in vivo* rodent Uterotrophic bioassays was subsequently performed to evaluate the performance of the *in vitro* ER pathway model that predicts estrogenic activity. A literature review of journal articles with uterotrophic studies resulted in the identification of 442 studies using 103 chemicals, which met all six minimum criteria to be considered "guideline-like" uterotrophic, and were used to evaluate (Browne et al., 2017; Browne et al., 2015; U.S. EPA, 2014b) the performance of the ER pathway model against the Uterotrophic assay.

A performance-based approach was used to evaluate the ER pathway model for agonist activity (see Section II.A. Performance-Based Approach to Establishing Confidence in NAMs). The performance evaluation included consideration of chemicals tested previously in both *in vitro* and *in vivo assays as* reference chemicals, and results of EDSP Tier 1 screening assays for List 1 chemicals (Browne et al., 2017; Browne et al., 2015). From this analysis performance metrics were provided for *in vitro* reference chemicals (28 positives; 12 negatives), *in vivo* reference chemicals (30 positives; 13 negatives), chemicals from guideline-like uterotrophic studies (57 positives; 48 negatives), and chemicals from Tier 1 studies (0 positives; 49 negatives).

For ER Agonist activity, the ER pathway model sensitivity was 93%, 97%, 89%, and 'not meaningful' in the *in vitro* reference, *in vivo* reference, uterotrophic, and Tier 1 chemicals, respectfully. There were no positives among the Tier 1 chemicals, consequently sensitivity scores were not meaningful. Sensitivity is most important for screening as false negatives are to be avoided. Specificity was 100%, 89%, 80%, and 100% in the *in vitro* reference, *in vivo* reference, uterotrophic, and Tier 1 chemicals, respectfully. These metrics reflect the performance of the ER pathway model when inconclusive chemicals ($0.001 < AUC < 0.0501$) were excluded from the calculations. When the inconclusive chemicals were considered positive, sensitivity remained similar, but specificity decreased. To ensure a transparent scientific process, EPA has published the methodology used for the ER pathway model, the R-code for the model, and the performance-based validation (Browne et al., 2017; Browne et al., 2015; Judson et al., 2015).

In 2014, the FIFRA SAP reviewed the full 18 HT assay ER pathway model (U.S. EPA, 2014a). In the final report of the December 2014 meeting, the SAP did not recommend substituting the ER model for the Uterotrophic assay at that time (see page 14 in the final report) and suggested that EPA consider: 1) additional Uterotrophic assays on chemicals in the low/middle AUC range to test predictive capacity of the ER Model for weak ER agonists and 2) perform similar comparisons of ER model scores of weak ER agonists with other *in vivo* assays in the Tier 1 battery relevant to estrogenicity. To address the SAP's recommendations, EPA published additional scientific support for the ER pathway model including uncertainty and sensitivity analyses (Watt and Judson, 2018; Watt, 2017; Judson et al., 2016; Browne et al., 2015; Judson et al., 2015), as well as a curated Uterotrophic database which provided the systematic process for selecting *in vivo* reference chemicals (Kleinstreuer et al., 2016). EPA also considered the SAP recommendations in developing the June 2015 FR notice that introduces the use of HT assays and computational models, including the proposed use of the full ER pathway model as an alternative to three EDSP Tier 1 screening assays (U.S. EPA, 2015a).

In 2019, OECD published a case study on the use of an integrated approach to testing and assessment (IATA) for estrogen receptor active chemicals using the ER pathway model (OECD, 2019). This IATA describes an integrated testing strategy (ITS) to identify ER bioactivity primarily for Tier 1 screening without the use of animal testing. The ITS relies on a pre-defined data interpretation procedure designed to provide consistent and reliable information on whether the substance tested may act as an ER agonist. The combination of up to 16 *in vitro* HT screening assays covers multiple key events of the pathway indicative of ER activation (agonist mode). (Note that there are 16 assays that were used to assess ER agonism and two additional assays that assess ER antagonism; antagonism is not the purpose of this IATA.) Consequently, the OECD case study is supportive of the performance-based validation approach EPA has used to support the use of the ER pathway model as a EDSP Tier 1 screening tool.

C. Androgen Receptor Pathway Model

For the AR model, EPA followed a process similar to that for the ER pathway model. The AR pathway model integrates activity from 11 HT *in vitro* assays to characterize AR bioactivity. The suite of 11 *in vitro* assays evaluates perturbations of AR pathway responses at key activity sites within the cell including AR receptor binding, dimerization, transactivation, and gene transcription (Kleinstreuer et al., 2017). Unique to the AR pathway model, a pair of antagonist-mode transactivation assays were performed at two different reference agonist concentrations, providing the ability to observe a diagnostic shift in potency indicative of receptor-mediated antagonist activity. The AR pathway model integrates the activity patterns across the 11 assays to predict whether a chemical is a potential AR agonist or antagonist or whether there is “assay interference” or cytotoxicity. The output of the AR pathway model provides an AUC value for the potential of a chemical to cause AR agonism and/or AR antagonism. The AUC scores represent the chemical or concentration-specific probabilities that the chemical is interacting with the corresponding receptor. Given that AR antagonism was the biological response of greatest concern, the AUC scores were scaled to yield a value of 1 for the antagonist positive control hydroxyflutamide. AR Pathway AUC (antagonist) scores ≥0.1 are considered “active” (Kleinstreuer et al., 2017). AR antagonist activity is easily confounded by cytotoxicity; hence, results were also combined with cytotoxicity information via a confidence scoring system to contextualize the results and reduce potential false positives (addressing this source of assay interference). Further analyses binned chemicals as positive, negative, or inconclusive for AR pathway bioactivity based on the AUC scores and confidence scores. For example, a chemical with an AR pathway AUC score greater than 0.1 (approximate activity at 100 µM) and confidence score >1 was considered positive with higher AUC scores corresponding to higher potency (Kleinstreuer et al., 2018; Kleinstreuer et al., 2017).

Activity in the AR pathway model was examined across 1,855 chemicals from the ToxCast library (Kleinstreuer et al., 2017). Included in the chemical library were reference AR agonists and antagonists, as well as selective androgen receptor modulators (SARMs). Out of 1,855 chemicals tested in all 11 AR pathway assays, 1,461 were predicted to be inactive in the AR pathway model (agonist and antagonist AUC values below 0.001), 33 chemicals were predicted to be AR agonists (AUC values >0.1), and 192 chemicals were predicted to be AR

antagonists (AUC values >0.1). The remaining 174 chemicals were predicted to have weak or inconclusive AR pathway activity (AUC values of 0.001 to 0.1).

A performance-based approach was used to evaluate the AR pathway model for both agonist and antagonist activity using a selection of reference chemicals (see Section II.A. Performance-Based Approach to Establishing Confidence in NAMs). The reference chemicals with *in vitro* AR agonist or antagonist activity (or lack of activity) were compiled from international test method validation efforts and semiautomated systematic literature reviews (Kleinstreuer et al., 2017). Reference chemical concentrations that activated or inhibited AR pathway activity were identified to establish a range of potencies with reproducible reference chemical results. Performance metrics were provided for *in vitro* agonist reference chemicals (8 positives; 21 negatives) and *in vitro* antagonist reference chemicals (20 positives; 8 negatives).

For AR agonist activity, performance metrics were 100% sensitivity and 95% specificity. For AR antagonist activity, performance metrics were 94.4% sensitivity and 100% specificity. These metrics reflect the performance of the AR pathway model when inconclusive chemicals (0.001 < AUC < 0.1) were excluded from the calculations. When the inconclusive chemicals are included in the positive set the sensitivity and specificity for the agonist model is 100% and 90.5%, respectively, while the sensitivity for the antagonist mode is 94.4% and specificity remains 100% (Kleinstreuer et al., 2017).

To further assess the performance of the AR pathway model, Tier 1, List 1 AR binding data (39 chemicals) were also compared with model scores (Kleinstreuer et al., 2017). When making these comparisons and accepting the Tier 1 AR binding results as accurate (which may not be the case), the AR pathway model sensitivity was 22% and specificity was 80%. However, these differences are understandable when considering the following information. Twenty-four chemicals were “negative” in both the AR pathway model and the Tier 1 binding assay, and two were identified as “positive” in both. The model identified seven chemicals as “negative” while the Tier 1 assay identified these same chemicals as “positive”, and six were identified as “positive” in the model but not in the Tier 1 assay (Kleinstreuer et al., 2017). However, with the exception of phosmet, these seven chemicals in the AR binding assay had IC50 values (half maximal inhibitory concentration) well over 100 µM and so would be expected to be negative in the model, as the highest tested concentrations in ToxCast and Tox21 were ≤100 µM. Likewise, the discrepancy with phosmet is explained in the cited publication (Kleinstreuer et al., 2017). Six List 1 negative/model positive chemicals were also flagged as potential false positives using the antagonist confirmation assay data.

Not screening at concentrations higher than 100 µM may lead to false negatives. However, in general, EPA limits the testing to 100 µM for several reasons. First, it is very unlikely that the target sites (such as the uterus) in humans, fish, or wildlife will be exposed to concentrations higher than 100 µM. Even if concentrations of 100 µM or more could be achieved at the target sites, it is possible that such high concentrations would exceed the maximum tolerated dose (*i.e.,* evidence of toxicity other than endocrine-related would be evident), which may confound interpretation of the results. Then there are technical problems with screening at such high concentrations. The ER and AR binding assay test guidelines (OCSPP 890.1250 and 890.1150) specify avoiding screening compounds at concentrations that are insoluble.

Although the AR and ER binding assays are cell-free, other screens (like the ERTA, OCSPP 890.1300) are not and specify screening at concentrations that are not cytotoxic. Thus, it is important to recognize that screening above 100 µM in cell-free assays may yield results that are qualitative in nature. Often, either solubility or cytotoxicity issues occur at doses exceeding 100 µM. For example, of the 7 false negatives from the AR paper (Kleinstreuer et al., 2017), **all** were either tested to above the *in vivo* point of departure (6/7) (*i.e.,* chemicals were regulated based on toxic effects occurring at a lower concentration) or above the *in vitro* cytotoxicity limit (6/7). Furthermore, a common maximum target concentration enabled a rapid, high throughput screening experimental design for all chemicals screened, as large chemical libraries were screened in a blinded fashion for ToxCast and customization of the design for each chemical (*i.e.*, determining the maximum concentration based on solubility and cytotoxicity) would have been resource prohibitive.

There were 51 chemicals with AR pathway model data and ICCVAM validation data set for the AR binding assay. Twenty two were positive in both, nine were model positive/AR binding assay negative, one was model negative/AR binding assay positive, and 19 were negative in both, which yielded a model sensitivity score of 96% and a specificity score of 68% (Kleinstreuer et al., 2017). The AR pathway model provides evidence that some of the ICCVAM designations for four chemicals (17α-estradiol, 4-cumylphenol, apigenin, and bisphenol B) might need additional investigation(Kleinstreuer et al., 2017). To ensure a transparent scientific process, EPA has published the methodology for the AR pathway model, all supporting information, and the R-code for the model (Watt and Judson, 2018; Kleinstreuer et al., 2017).

The AR pathway HT assays, AR pathway model, and the performance-based validation of the AR pathway model were independently reviewed by the FIFRA SAP in December 2014 (U.S. EPA, 2014a). Following the 2014 FIFRA SAP recommendation, EPA revised the AR pathway model. EPA sought the FIFRA SAP's advice on the revised model in November 2017 SAP meeting (U.S. EPA, 2017). The November 2017 SAP recommended that EPA continue to investigate the emerging scientific issues for HT testing and use of computational models including subset models (U.S. EPA, 2017, see https://www.regulations.gov/document/EPA-HQ-OPP-2017-0214-0024), which has now been completed and published (Judson et al., 2020). The Judson et al. (2020) paper provides an updated model consisting of 14 assays, with three additional assays included to expand the biological coverage of the AR pathway and is responsive to SAP2017 recommendations. As part of the implementation of the full AR pathway, EPA will look closely at the results of both the 11 and 14 assay pathway models. EPA's detailed responses to the recommendations of the November 2017 SAP can be found elsewhere (U.S. EPA, 2022a).

D. Strengths in the ER and AR Pathway Models

The use of HT assays and computational modeling provides more rapid data generation, reduced cost to generate toxicity data, and can result in more directed and hypothesis-driven toxicity and epidemiological studies. The ER and AR pathway models primarily use *in vitro* assays that include human cells or genetic material, which makes these assays particularly relevant for assessing hazards in humans as mandated for evaluation by the FFDCA.

Prior to EPA's work on HT assay and computational models such as the ER and AR pathway models, screening chemicals for endocrine bioactivity was conducted using a combination of 11 low-throughput *in vivo* and *in vitro* tests (Tier 1 assays of Tier 1 battery; see Table 1) that are time-consuming and use many laboratory animals. Following WoE evaluation of the Tier 1 results, EPA determines whether additional testing (*e.g.*, Tier 2, targeted studies) is needed to perform a risk assessment. OCSPP has been working with the ORD to more rapidly screen chemicals and minimize the use of animal testing by expanding the set of HT tools for use in the EDSP.

The ER and AR pathway models offer many advantages compared to the current lower throughput methods in the EDSP Tier 1 battery (Kleinstreuer et al., 2018; Kleinstreuer et al., 2017; Browne et al., 2015; Judson et al., 2015; U.S. EPA, 2014a). Both models integrate multiple biochemical cell-free and cell-based HT *in vitro* assays that probe different points along an AOP using multiple technologies. The models provide the ability to prioritize and screen a large set of chemicals from diverse chemical classes, with human and ecological exposure potential, for additional *in vivo* endocrine testing at a lower cost and a fraction of the time needed to test individual chemicals using current EDSP Tier 1 methods. As part of the transition to HT and computational methods in the EDSP, EPA has screened approximately 1,800 chemicals for ER and AR-mediated activity using these models. Finally, these models provide non-animal screening alternatives for four Tier 1 assays: *in vitro* ER binding, *in vitro* ER transactivation (ERTA), *in vivo* Uterotrophic, and *in vitro* AR binding.

The ER and AR pathway-based approaches integrate multiple assays mapping to many pathway-based KEs, which provide a more holistic description of the potential of a chemical's ability to activate or inhibit signaling. The ER and AR pathway models and associated HT *in vitro* assays cover a broader range of biological processes than their Tier 1 *in vitro* counterparts (ER binding, ERTA, and AR binding assays) that focus on receptor binding and can process chemicals more rapidly than lower-throughput transactivation-type assays. These types of network models integrate concentration-response profiles (or combine data) from multiple assays assessing activity at key points in the ER or AR pathway to compensate for the individual limitations of any one assay. That is, the ER and AR pathway models combine data from multiple assays to reduce technology-specific assay interference (for example by reducing false positives). Cytotoxicity and response specificity are considered and flagged (Watt and Judson, 2018; Kleinstreuer et al., 2017; Judson et al., 2015).

In summary, ER and AR pathway models are non-animal methods developed for the rapid and cost-effective assessment of ER- or AR-mediated chemical bioactivity. They were evaluated using a well-defined set of reference substances representing a range of chemical classes and potencies. Both models include procedures for evaluating chemical cytotoxicity and accounting for processes that can cause assay interference. Performance-based validation has led to the curation and storage of legacy *in vitro* and *in vivo* toxicity studies which can be used as a global resource to compare with the HT screening data.

E. Limitations and Uncertainties in the ER and AR Pathway Models

EPA chose to focus research efforts on development and evaluation of ER and AR pathway HT assays and computational models based on classical ER- or AR-mediated-transcriptional activation processes. Neither NAM was designed to specifically identify chemicals that affect bioactivity through non-genomic signaling mechanisms (Fuentes and Silveyra, 2019; Foradori et al., 2008). There are other EDSP assays in the Tier 1 screening battery that complement the ER pathway and AR pathway models. These *in vivo* assays (*e.g.*, Male and Female Pubertal, Hershberger, Uterotrophic, and Fish Short-term Reproduction (FSTR) assays, are capable of detecting responses resulting from both classical and/or non-classical ER or AR signaling, as applicable to species, sex, and test system (see Table 1).

Almost all the *in vitro* assays that inform the ER pathway model interrogate ER-α-mediated (ESR1) activity (Judson et al., 2015), and a subset of these assays (~33%) assess activity mediated by ER-β (ESR2) as well. The Tier 1 ERTA assay (OCSPP 890.1300) detects hERα-mediated transcriptional activity, which is functionally very similar to the transcriptional activation assays used in the ER pathway model. The Tier 1 ER binding assay (OCSPP 890.1250) is performed with rat uterine cytosol, whereas the Uterotrophic assay (890.1600) is an *in vivo* assay that measures rat uterine response to chemical exposure. Studies have shown that the rat uterus estrogen receptor profile consists primarily of ER-α (ESR1) with much lower levels of ER-β (ESR2) and G protein-coupled estrogen receptors (GPERs) (Hutson et al., 2019; Blesson and Sahlin, 2012). Thus, there may be limitations in the ability of the *in vitro* assays comprising the ER model or the Tier 1 ERTA or ER binding assays to fully characterize some tissue-dependent or lifestage-dependent responses that may be operative *in vivo*, *e.g.*, differential agonist/partial agonist actions of tamoxifen (a SERM in uterus and breast tissue) (Hu et al., 2015; Judson et al., 2015). While the ER model identified all reference agonists, antagonists and SERMs as being ER-active, the ER model may not always accurately predict whether *in vivo* ER activity may be antagonistic, agonistic or both. However, while the Uterotrophic *in vivo* model does contain ER-α, ER-β and GPERs, the 890 guideline only measures responses in one tissue (*i.e.*, uterine weight) and the protocol only identifies estrogen agonism. Tier 1 is designed to assess the potential for bioactivity, and further information about dose-response is provided by Tier 2 tests. As noted previously, EPA will accept the results of the ER pathway model as an alternative for the *in vivo* Uterotrophic assay on a case-by-case basis.

Both HT *in vitro* and low-throughput *in vitro* assay results may be influenced by physical-chemical properties of the test chemical and the chosen test system. These issues help define the domain of applicability of these assays. Limitations in ER pathway HT assays underlying the ER pathway model (Watt, 2017; Browne et al., 2015; Judson et al., 2015) and AR pathway model (Kleinstreuer et al., 2017) assays are similar to those identified in other *in vitro* systems such as: limited or lack of xenobiotic metabolic capacity; technical assay interference; chemical insolubility; adsorption of chemical to testing surfaces; use of transformed cells from organs/tissues that lack biological context; and chemical volatility. HT *in vitro* and low-throughput *in vitro* assay results may be influenced by physical-chemical properties of the test chemical and the chosen test system. Recent analysis by (Mcmullen et al., 2018) on 391 substances in ToxCast “observed a strong relationship between the fraction of positive results

in ToxCast assays and compound volatility" across a range of endpoints based on the upper two quartiles of vapor pressure. Other papers have found similar effects (Fischer et al., 2017). The Armitage et al. (2014) mass balance models do not consider cross-well contamination due to volatilization, but do consider adherence to well walls, evaporative losses (via excessive portioning into the head space), and protein binding. These effects have, in part, been independently confirmed experimentally in the peer-reviewed literature (Henneberger et al., 2020; Huchthausen et al., 2020; Birch et al., 2019; Henneberger et al., 2019). Substances with an experimental or modeled Henry's Law Constant ≥ 0.1 $Pa \cdot m^3/mol$[1] (OECD, 2019) may present technical difficulties for testing. Given these uncertainties, EPA will consider whether negative or equivocal ER and AR bioactivity model results for substances could be due to physical-chemical parameters rather than a true lack of an effect.

HT assays associated with the AR and ER pathway models and low throughput *in vitro* Tier 1 assays have limited or no metabolic capability. Metabolism can occur in the *in vivo* Uterotrophic assay, and this is one reason EPA is accepting the results of the ER pathway model as an alternative for the *in vivo* Uterotrophic assay on a case-by-case basis. Because *in vitro* assays have limited or no xenobiotic metabolism, false positive results (chemical is detoxified *in vivo*) or false negative results (chemical is bioactivated *in vivo*) can occur. Both results may lead to mischaracterization of a chemical's potential ER or AR bioactivity. Research indicates that the lack of metabolic competence in the ER pathway model limits the predictive capacity to quantitatively extrapolate from *in vitro* ER activation to *in vivo* Uterotrophic response (Conley et al., 2016). More recently, Gray et al. (2020) published on the limited extrapolation capacity of the AR pathway model to predict *in vivo* responses for anti-androgens. One possible explanation is the role of toxicokinetic factors in extrapolation evaluations. For example, Casey et al. (2018) showed that when toxicokinetic factors were incorporated into the comparison, the quantitative *in vitro*-to-*in vivo* concordance was improved for ER activation (Casey et al., 2018). Information on a chemical's metabolites may be available, e.g., in the published literature or as data in support of a pesticide registration. An *in silico* approach for the prediction of estrogenic bioactivity of chemical metabolites may also be used to supplement the predictions of the ER pathway model (Pinto et al., 2016).

There are also ongoing approaches to provide metabolic competency in the HT assays including Tier 1 EDSP assays. For example, EPA researchers are using the Alginate Immobilization of Metabolic Enzymes (AIME) Platform to retrofit the ER transactivation assay with metabolic competence (Deisenroth et al., 2020) and has been applied to over 700 ToxCast chemicals (Hopperstad et al., 2022). Another method involves the transfection of modified mRNAs to introduce metabolic capacity into cells (Degroot et al., 2018).

Some logistical and capacity issues limit the use of the 18 ER HT assays and 11 AR HT assays for screening new chemicals.

- Some HT assays are no longer commercially available (e.g., three original Novascreen[2] assays in the ER pathway model). In contrast, the existing EDSP Tier

[1] This threshold is based on the indicator value presented in OECD Guidance Document 23 (OECD, 2019).

[2] Novascreen was acquired by Caliper Life Sciences, which was acquired by Perkin Elmer, which no longer offers the 3 original Novascreen assays.

1 assays use widely available cell lines and testing protocols that can be run by EPA itself, a registrant or test order recipient with an appropriate laboratory, or any competent contract laboratory.
- Some HT assays are proprietary commercial assays which may limit access to stakeholders interested in conducting the assays.
- Robotic HT equipment and expertise may not be readily available everywhere.

Although the EDSP is intended to consider humans and fish and wildlife, the ER and AR pathway models primarily use *in vitro* assays that include human cells or genetic materials. As discussed below, SeqAPASS can be used to assess the cross-species hazard potential.

EPA expects to consider the uncertainties described here as part of the implementation of next steps laid out in the Section IX. Conclusion.

F. Reduced or Minimal ER and AR Pathway Subset Models for Priority Setting

Reduced or minimal ER and AR pathway subset models strive to provide equivalent performance as the full ER and AR pathway models using smaller and more flexible combinations of assays underlying the full model set. This is particularly important as many of the assays supporting development of the full ER and AR pathway models are proprietary or are no longer offered commercially. Moreover, the reduced or minimal ER and AR pathway subset models will be more cost effective.

The following number of assay subsets are needed to provide an equivalent balanced accuracy relative to the full ER and AR pathway models for the entire set of tested chemicals:

- As few as 4 ER Agonist Model assays (Judson et al., 2017)
- As few as 5 AR Agonist Model assays (Judson et al., 2020)
- As few as 6 AR Antagonist Model assays (Judson et al., 2020).

No assay subset models were developed for the ER antagonist mode, although this was considered in the original ER pathway model publication (Judson et al., 2015).

In developing the ER agonist subset model, some of the subset models (518 models of 65,535 possible assay combinations in (Judson et al., 2017; Judson et al., 2015)) were able to achieve “acceptable” overall performance compared to the full pathway model. Here, “acceptable” means that the performance of the subset model was within the range of uncertainty provided by the bootstrapping sensitivity analysis using the method of Watt and Judson (Watt and Judson, 2018). An ER agonist subset model of 4 assays achieved a sensitivity of 98% and specificity of 92% compared to results from the full ER pathway model for the entire set of tested chemicals. Performance metrics were also provided for *in vitro* reference chemicals (28 positives; 8 negatives) and *in vivo* reference chemicals (30 positives; 11 negatives). This same

4-assay ER agonist subset model had a sensitivity of 89% and specificity of 100% with *in vitro* reference chemicals and a sensitivity of 97% and specificity of 91% with *in vivo* reference chemicals (Judson et al., 2017).

The results for the best six-assay AR antagonist subset model and five-assay AR agonist subset model are not statistically different from the full model when variability is incorporated. The best five-assay AR agonist model has a sensitivity of 89% and a specificity of 98% when compared to the full AR pathway model for all tested chemicals. A five-assay AR antagonist subset model has a sensitivity of 97% and specificity of 96% for all chemicals when compared to the results of the full AR pathway model. The best 5-assay AR agonist subset models are not statistically different from the full model once its variability is accounted for, but sensitivity was less than 90%. A six-assay AR agonist subset model provides a sensitivity of 96% and specificity of 98% when compared to the results of the full agonism AR pathway model (Judson et al., 2020).

When comparing against the *in vitro* reference compounds (37 agonist and 28 antagonist), as few as three assays for the antagonist and two assays for the agonist will achieve 100% sensitivity and specificity (Judson et al., 2020). Like the full ER and AR pathway models, uncertainties and limitations affect the use of these subset ER and AR pathway models. Unfortunately, many of the best performing subset models include assays that are no longer commercially available (although they or comparable alternatives are expected to become available in the future), and some are missing critical steps along the ER/AR activation pathways. Depending on the number of assays included, the sensitivity of the subset models was generally higher than the specificity (Judson et al., 2017).

G. Summary for ER and AR Pathway Models

Given the strengths discussed in Section III.C. as well as the uncertainties and limitations discussed in Section III.D., EPA has determined that the following NAMs may be used as alternatives for the following four (4) EDSP Tier 1 screening assays when evaluated on a chemical-by-chemical basis (each assay evaluated independently). EPA reviews the quality of available data across multiple lines of evidence. For the NAMs methods listed here, EPA will specifically consider the quality of the individual ToxCast assay data, and the level of confidence in and biological relevance of the predictions.

(1) The Estrogen Receptor (ER) pathway model based on the full 18-assay ToxCast/Tox21 battery (Browne et al., 2015; Judson et al., 2015) may be used as an alternative for three current EDSP Tier 1 screening assays:

- ER binding *in vitro* assay (OCSPP 890.1250; (U.S. EPA, 2009b)).
- ER transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300; (U.S. EPA, 2009c)).
- *In vivo* Uterotrophic assay (rat) (OCSPP 890.1600; (U.S. EPA, 2009d)).

(2) The Androgen Receptor (AR) pathway model based on the full 11-assay ToxCast/Tox21 battery (Kleinstreuer et al., 2017) may be used as an alternative for one current EDSP Tier 1 screening assay:

- AR binding *in vitro* assay (OCSPP 890.1150; (U.S. EPA, 2009a)).

The full ER and AR pathway models offer many strengths compared to the current low-throughput methods in the EDSP Tier 1 battery. The ER and AR pathway models integrate multiple biochemical cell free and cell-based HT *in vitro* assays that probe different key events using multiple technologies. The models provide the ability to prioritize and screen a large set of chemicals from diverse chemical classes at a lower cost and more quickly than using current EDSP Tier 1 methods.

EPA continues to investigate the suitability of reduced ER and AR pathway assay sets and computational models as alternatives to the EDSP Tier 1 screening assays (Judson et al., 2020; Judson et al., 2017) (See Section III).

IV. Quantitative Structure-Activity Relationship (QSAR) Models

EPA has determined that consensus [quantitative] structure-activity relationship ([Q]SAR) models are appropriate for use in priority setting of EDSP chemicals and as OSRI in combination with other relevant information. Two such QSAR models were developed by international consortia of experts led by EPA: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) to evaluate chemicals for potential ER activity (Mansouri et al., 2016) and Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA) (Mansouri et al., 2020) to evaluate potential AR activities. Expert modelers and computational toxicology scientists from 35 international groups contributed structure-based models and results to one or both projects, with methods ranging from QSARs to molecular docking in order to predict binding, agonism and antagonism activities. In simple terms, experts used the ER and AR activity from the ER and AR pathway models to train QSAR models to predict ER and AR activity (Kleinstreuer et al., 2017; Judson et al., 2015). External evaluation sets used ER and AR data from public sources other than the ToxCast data used in training the models. The specific details of the ER and AR consensus model building and literature curation protocols are outlined in the CERAPP and CoMPARA publications (Judson et al., 2020; Mansouri et al., 2016). The CERAPP QSAR model is also discussed in the EPA White Paper for the December 2014 FIFRA SAP (U.S. EPA, 2014a, see https://www.regulations.gov/document/EPA-HQ-OPP-2014-0614-0003).

Individual QSAR models developed for CERAPP and COMPARA were trained on the results for ~1,800 chemicals from the ER and AR pathway models respectively. Two classes of models were built: those that predicted potency and those that just predicted activity (active / inactive). The individual QSAR models were then evaluated using curated literature data from different sources (~7,000 results for ER and ~11,000 results for AR). To overcome the limitations of individual QSAR models, CERAPP and COMPARA combined many models together in a weighted fashion to generate consensus predictions for each chemical. Balanced accuracy is the arithmetic mean of specificity (true negative rate) and sensitivity (true positive rate). Overall balanced accuracy was 91% (CERAPP, ER agonist) and 74%-86% (CoMPARA,

AR antagonist). These consensus models were extended beyond the initial datasets by integrating them into the free and open-source web browser application OPERA (Mansouri et al., 2018) to avoid running every single model on new chemicals. This implementation was used to evaluate the entire EPA DSSTox database of ~800,000 chemicals and provided online on the EPA's CompTox chemicals dashboard. The ER and/or AR bioactivities for additional chemicals may be used to refine the individual QSAR models and associated consensus models for CERAPP and COMPARA.

There are different sources of uncertainty in the CERAPP and CoMPARA QSAR consensus models inherited from the underlying training data. Both the COMPARA and CERAPP QSAR consensus models were built using AR and ER ToxCast datasets available at the time as training data (Judson et al., 2020; Mansouri et al., 2020; Mansouri et al., 2016; Judson et al., 2015) and evaluated against well-known reference compounds. Thus, the goal is to predict *in vitro* ER or AR activity, with known limitations including differing concentration ranges *in vitro*, limited metabolic capacity, varying activity of selective AR modulators (SARMs), and other possible experimental artifacts and errors. As discussed in Section III.E, the likelihood of chemical volatilization, need for metabolism or metabolite prediction (Mansouri et al., 2020; Pinto et al., 2016), number of technical and analytical replicates, or possible differential partitioning or binding to plastic *in vitro* should be investigated when using any *in vitro* data to limit the impact of false negatives or positives. These limitations and uncertainties are applicable across all QSAR model results which were trained on the HT *in vitro* data. As such, EPA intends to consider the results of the QSAR models carefully on a case-by-case basis that factors in information from additional lines of evidence. To account for the limitations of the models and to provide maximum transparency around predictions, the OPERA implementation of CERAPP and CoMPARA provides an assessment of the applicability domain, as well as an accuracy estimate and the most similar structures for each predicted chemical from the knowledge base of the respective model; this assessment does not account for the limitations and uncertainties relevant to the *in vitro* data noted above.

Based on the current strength of the science and despite its limitations, EPA has determined that CERAPP and CoMPARA QSAR consensus models can be used for priority-setting purposes and for consideration as use as OSRI in WoE in combination with other relevant information under the EDSP.

V. Integration of Bioactivity and Exposure

In August 1998 (EDSTAC, 1998), the EDSTAC recommended that data resulting from HT screening assays "*will be combined with exposure-related information…for the purpose of setting priorities for T1S [Tier 1 Screening].*" The January 2013 FIFRA SAP (U.S. EPA, 2013) recommended that *"exposure potential could be incorporated into the ranking and prioritize chemicals for further testing*".

In July 2014, EPA brought to the SAP for review scientific issues associated with new HT methods to estimate chemical exposure for humans. Exposure Forecasting (ExpoCast) is an EPA ORD initiative to develop the necessary approaches and tools for rapidly prioritizing and

screening thousands of chemicals based on the potential for human exposure. This focus of ExpoCast is distinct from many existing traditional, lower throughput exposure tools that require considerable data to generate chemical-specific exposure predictions for screening-level assessments or full regulatory risk assessments. ExpoCast efforts have focused on providing quantitative exposure predictions while empirically assessing the uncertainty by combining multiple exposure models or chemical-specific information and calibrating the predictions with available biomonitoring data. EPA refers to this framework as the Systematic Empirical Evaluation of Models (SEEM). New methods for ExpoCast considered for prioritization of chemical screening in the EDSP were favorably reviewed by the July 2014 FIFRA SAP (U.S. EPA, 2014c, see https://www.epa.gov/sites/default/files/2015-06/documents/072914minutes.pdf).

The currently published ExpoCast model is limited to making exposure predictions for the general population due to the US NHANES data used to calibrate it. However, ORD is currently expanding application of the model to occupational populations. Once completed, the model can be used to estimate occupational exposures employing methods comparable to what was evaluated by the 2014 SAP.

In addition to predictions of exposure, the July 2014 SAP reviewed high throughput toxicokinetic (HTTK) methods for extrapolating *in vitro* doses to *in vivo* concentration (*in vitro*-to-*in vivo* extrapolation or IVIVE) for chemicals that have been run in a battery of HT endocrine screening assays (*e.g.,* ToxCast). HTTK provides a bridge between bioactivity measured in the HT screening assays and exposure by either predicting tissue concentrations from an administered dose (*i.e.,* what has been called forward toxicokinetics) or inferring administered doses that would be needed to cause tissue bioactive concentrations *in vivo* (*i.e.,* reverse toxicokinetics). The SAP agreed on the importance of applying HTTK to the HT bioactivity measurements to provide a dose context and better discriminate between chemicals for prioritization.

In December 2014, EPA asked the FIFRA SAP for advice on the use of an Integrated Bioactivity Exposure Ratio (IBER) approach to rank and prioritize chemicals for further EDSP consideration. In the IBER approach, reverse toxicokinetics is used to estimate the daily administered dose (mg/kg BW/day) necessary to produce steady-state *in vivo* blood concentrations equivalent to bioactive concentrations. These bioactive concentrations include those showing biological activity in the endocrine-related HT screening assays or the estimated potency values from the ER and AR pathway models. The putative bioactive administered doses can then be directly compared with predicted exposures (mg/kg BW/day) as portrayed in Figure 1**.** The extrapolated *in vivo* bioactive dose and the modeled exposure estimates are then used to prioritize chemicals. Considered together, these data indicate chemicals for which activity and exposure may overlap or be close enough to use in a priority setting tool for EDSP Tier 1 screening.

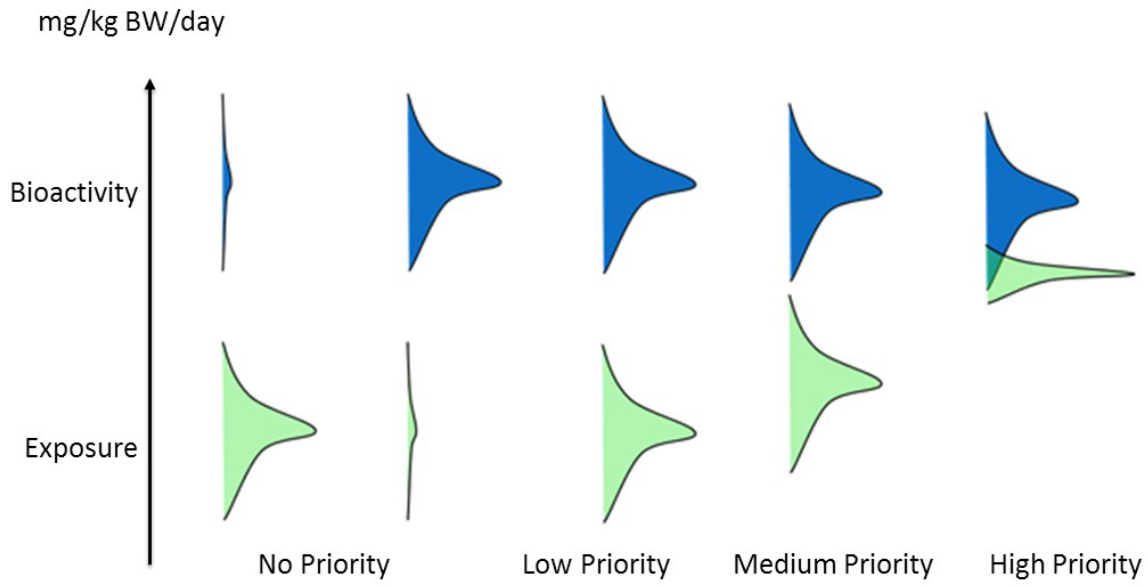

Figure 1. A graphical representation of the Integrated Bioactivity-Exposure Ratio (IBER) approach indicating no priority chemicals (those that either have no significant bioactivity or no potential exposure), low (chemicals for which extrapolated bioactive intake doses are considerably higher than potential exposures), medium, and high priority (chemicals for which bioactivity and exposure concentrations are close or overlapping).

Prioritizing environmental chemicals based solely on potential bioactivities may not be sufficiently protective for a chemical which may be less potent but also have high estimated exposures (U.S. EPA, 2014a). Though there are recognized sources of uncertainty around bioactivity, toxicokinetic, and exposure model estimates, these data may still be useful to prioritize chemicals for EDSP Tier 1 screening (U.S. EPA, 2014a). EPA considers the IBER approach appropriate for use in prioritizing chemicals with potential endocrine effects in combination with other prioritization approaches. The IBER approach is especially useful for the thousands of chemicals for which other sources of bioactivity and exposure information do not exist.

In addition to these SAP reviews, numerous articles on HT *in vitro* assays (Wambaugh et al., 2014; Rotroff et al., 2013; Kavlock et al., 2012; Rotroff et al., 2010), toxicokinetics (Wetmore et al., 2014; Wetmore et al., 2013; Rotroff et al., 2010), and exposure modeling (Wambaugh et al., 2014; Wambaugh et al., 2013; Judson et al., 2011) have appeared in peer-reviewed journals, broadly engaging the scientific community relative to potential use of these tools by EPA. The use of an *in vitro* to *in vivo* extrapolation approach that employs HT toxicokinetic data and modeling to estimate NAM-derived administered equivalent doses has been demonstrated in the literature, beyond endocrine-related bioactivity (Beames et al., 2020; Friedman et al., 2020; Wegner et al., 2020; Haggard et al., 2019; Honda et al., 2019; Pham et al., 2019; Casey et al., 2018; Wetmore, 2015). Comparison of NAM-derived administered equivalent doses to exposure to derive bioactivity exposure ratios for prioritization has also been demonstrated (Haggard et al., 2019; Thomas et al., 2019; Wetmore et al., 2015). Ongoing data collection efforts continue to increase the coverage of chemical- and species-specific *in vitro* toxicokinetic data (Black et al., 2021). At the same time, the toxicokinetic methods will also be refined over time (see discussion in Wambaugh et al., 2019; Bell et al., 2018; Wambaugh et al., 2018; Wambaugh et al., 2015). New HT *in vitro* toxicokinetic assays

and models can refine toxicokinetic assumptions (for example, replacing 100% bioavailability with chemical-specific estimates) as well as expand exposure routes (to allow, for example, the calculation of dermal and inhalation equivalent doses) (Breen et al., 2021).

EPA considers the IBER approach appropriate for use in priority setting in combination with other prioritization approaches and use as OSRI in WoE in combination with other relevant information.

VI. Interspecies Extrapolation using SeqAPASS

The EDSTAC recommended that EPA screen for endocrine disruption potential not only in humans but also in fish and wildlife. Consequently, when the EDSP was established in 1998, EPA used its discretionary authority under FFDCA to have the program focus on screening and testing for endocrine disruption in both humans and fish and wildlife (63 FR 42852 and 63 FR 71542). Currently, human and environmental risk assessments for chemicals use only a limited number of model surrogate species to generate bioactivity or toxicity test data. This creates a need to extrapolate potential bioactivity/susceptibility across multiple species and taxa. Increased demand for rapid, yet scientifically-sound, predictive approaches that maximize the use of existing data have been driven by reductions in testing resources, a global interest in reducing animal use (U.S. EPA, 2021a, 2019), and an increasing demand to evaluate chemicals in a timely and reliable manner. In response, EPA developed a fast, online screening tool, Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS), that allows researchers and regulators to use available knowledge regarding species' sensitivity to chemicals together with protein sequence information to transparently predict chemical susceptibility for hundreds of other species (Lalone et al., 2018; Ankley et al., 2016; Lalone et al., 2016; Perkins et al., 2013).

Many chemicals, including endocrine-active compounds, exert their biological effects through interactions with proteins. SeqAPASS uses available protein sequence and structural information to understand cross species conservation at the molecular level, which is used to predict chemical susceptibility (Figure 2). To examine the extent to which a protein may be conserved across species/taxa, EPA is utilizing a three-step process (Levels 1 – 3) within SeqAPASS, taking advantage of well-established, publicly accessible, and continuously expanding curated protein sequence information. The depth of the SeqAPASS evaluation depends on the extent to which the protein has been characterized along with how much information is available on a chemical-protein interaction. Therefore, the SeqAPASS tool was created with the flexibility to take advantage of existing information and generate susceptibility predictions at each level of the analysis.

The initial step (Level 1) of the process compares the entire primary amino acid sequence of the target species' protein from a known sensitive species to the protein sequences of all other species for which information is available. At the same time, ortholog candidates (*i.e.*, sequences that have diverged because of a speciation event and are therefore more likely to have similar function) are identified to set possible cut-offs for whether a species may be susceptible or not. If information is available from curated sources and from the literature regarding known functional domains (*e.g.*, ligand binding or enzymatic regions of the

sequence), the analysis proceeds to Level 2 which compares these functional domains across species. This underscores the need to have curated databases on functional domains across species and their critical amino acid residues.

The Level 3 evaluation requires knowledge of critical amino acid residues, such as those that form hydrogen bonds with ligands or are involved in the catalytic region of enzymes, which are important in the chemical-protein interaction. These amino acids can be compared across species to understand similarity and predict susceptibility (Lalone et al., 2018). Using existing data from model organisms to inform predictions reduces the need for additional resource-intensive toxicity testing. SeqAPASS minimizes the complexity of protein sequence and structural comparisons for species extrapolation, making the process more rapid and less daunting for scientists and regulators alike to help guide research and inform risk assessments.

In the context of the AOP framework (Ankley et al., 2010), results from SeqAPASS may be used to predict the extent to which signaling pathways are conserved across species. Just as KEs at relatively early points within these pathways may be predictive of adverse outcomes at the whole organism level, the extent of genomic/proteomic conservation can influence the extent to which the data can be extrapolated to other species/taxa. SeqAPASS output helps identify the scope (domain) of susceptible species across various taxa with similar (*i.e.*, well conserved) biological processes and pathways. SeqAPASS can be used to define the taxonomic domain of applicability for mammalian-based assays, identify unique species differences that may drive the development of assays to broaden taxonomic coverage or, where more specific assays should be developed, to broaden the taxonomic domain of the model and identify relevant susceptible species that might need additional in-life testing. (Lalone et al., 2018) found that SeqAPASS results for ToxCast assay targets can be obtained and proposed them as an initial line of evidence for extrapolating mammalian-based HT assay data across ecologically-relevant species (Lalone et al., 2018).

In summary, the SeqAPASS is an interspecies extrapolation tool that helps predict the extent to which data generated to evaluate endocrine bioactivity using mammalian systems (*e.g.*, ER pathway assay results and in progress, AR pathway assay results) can be extrapolated to non-mammalian species (*e.g.*, fish, amphibians, and birds). For example, ToxCast estrogenic bioactivity results were extrapolated from mammalian organisms to fish to help identify the taxonomic domain of applicability of the HT results/conclusions regarding likely active/inactive chemicals (Lalone et al., 2018; Ankley et al., 2016). As proposed in (Ankley et al., 2016), EPA would consider a pathway-based multi-step framework (in which the SeqAPASS results are part of the first step) to assemble and integrate information to conduct a comparative analysis focused on the potential for chemicals to interact with a relevant endocrine target in different species. In the 2018 update of the OECD Revised Guidance Document 150 on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption (OECD, 2018a), the SeqAPASS was noted as a tool which may assess endocrine effects on non-target species.

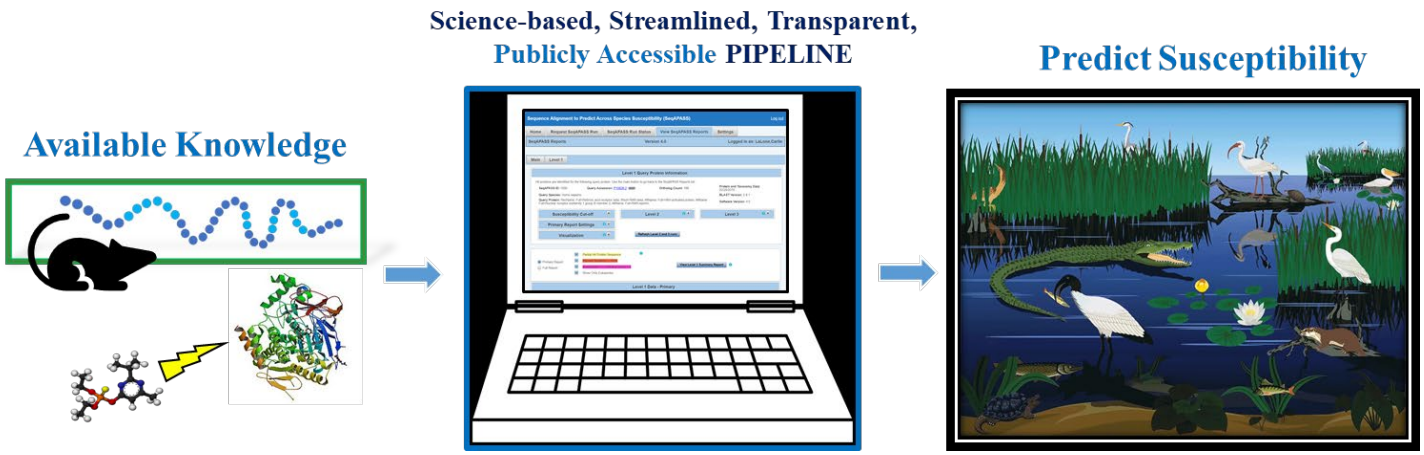

Figure 2. SeqAPASS uses available protein sequence and structural information to understand cross species conservation at the molecular level, which is used to predict chemical susceptibility. SeqAPASS is publicly accessible online on EPA's website.

EPA will be using the SeqAPASS as an interspecies extrapolation tool (Lalone et al., 2018; Ankley et al., 2016) for consideration as use as OSRI (U.S. EPA, 2009e) in WoE (U.S. EPA, 2011b) evaluations.

VII. Thyroid Adverse Outcome Pathway (AOP) Framework

EPA is investigating the biology of the thyroid system and identifying and developing HT screening assays useful for interrogating thyroid biology as a prelude to developing an integrated model that may produce results that can be considered OSRI. In November 2017, EPA requested independent, scientific advice from the FIFRA SAP on EPA's ongoing development of an approach to detect substances that can perturb thyroid function (U.S. EPA, 2017). The FIFRA SAP supported the development of EPA's initial framework (U.S. EPA, 2017, see https://www.regulations.gov/document/EPA-HQ-OPP-2017-0214-0024). Subsequent to the November 2017 FIFRA SAP, EPA published a review of the thyroid AOP network (Noyes et al., 2019). EPA's response to the SAP recommendations is provided as part of the Response to the FIFRA SAP November 2017 Report EPA (U.S. EPA, 2017, see https://www.regulations.gov/document/EPA-HQ-OPP-2017-0214-0024). EPA will consider the November 2017 FIFRA SAP recommendations as the Thyroid Pathway Framework is further developed and evaluated.

EPA has ongoing research to develop the HT screening assays for thyroid-relevant targets that were presented at the November 2017 FIFRA SAP meeting (U.S. EPA, 2017), including those involved in thyroid regulation (thyroid hormone receptor, thyroid stimulating hormone receptor, and thyrotrophin-releasing hormone receptor); thyroid synthesis (Haselman et al., 2020; Paul-Friedman et al., 2019; Wang et al., 2019); peripheral thyroid deiodination (Olker et al., 2019) and iodide recycling (Olker et al., 2021); serum thyroid transport; and markers of increased hepatic catabolism of thyroid, among others (Noyes et al., 2019). Thyroid regulation differs across life stages (*e.g.*, alterations in the requirement for iodine in pregnant females). EPA also

has ongoing research to develop models to predict thyroid-related apical outcomes based on biochemical inputs (Haselman et al., 2020; Hassan et al., 2020). EPA is also collaborating with international efforts, particularly those in the European Union, to advance NAMs for thyroid outcomes. Additional research and peer review are needed to identify the most important molecular initiating events and to ensure that NAMs will exist to provide orthogonal and confirmatory results for each of these events and augment the current assays in the tier battery that interrogate thyroid hormone systems.

VIII. Future Direction

The development of HT assays and computational models by the EDSP has been underway since the program was first introduced in 1998 (EDSTAC, 1998). To date, reviewing EDSP chemicals for potential bioactivity using lower-throughput *in vitro* and *in vivo* assays in the Tier 1 battery has proven to be complex, costly, and time-consuming. EPA has described several NAMs that it can use to aid in the screening and prioritization of chemicals for potential endocrine bioactivity and/or exposure.

EPA is continuing to develop NAMs for other Tier 1 tests (see Table 1). These include NAMs that can be combined with the AR Pathway model to predict the results of the Hershberger assay (Kleinstreuer et al., 2018). The AR pathway model was also compared to results from a curated rodent Hershberger database (Browne et al., 2018; Kleinstreuer et al., 2018). Overall agreement was 66% (19/29), with ten additional inconclusive chemicals. Most discrepancies were explained based on differences in dosimetry. Within the chemical set examined, the AR model had 100% positive predictive value for the *in vivo* Hershberger response, *i.e.*, there were no false positives, and chemicals with conclusive AR model results (agonist or antagonist) were consistently positive *in vivo*. These additional NAMs may include incorporating metabolic competence in the existing assays or an assay for 5-alpha reductase activity. Additionally, HT alternatives for addressing the thyroid AOP network are underway. EPA will continue to refine the HT assays, models, and tools to more efficiently assess a chemical's potential to interact with the estrogen, androgen, and thyroid systems.

A HT H295R cell-based assay (Karmaus et al., 2016) has been developed that uses high-performance liquid chromatography followed by tandem mass spectrometry to measure multiple components of steroid synthesis. EPA performed a comparison (sensitivity and specificity) of the low-throughput and HT H295R assays for detecting the disruption of synthesis of estradiol and testosterone and presented this analysis to the FIFRA SAP in November 2017 (Haggard et al., 2018; U.S. EPA, 2017). The November 2017 FIFRA SAP made a number of recommendations to improve the robustness and performance of the HT-H295R assay and application of the mean Mahalanobis distance for prioritization of chemicals in the EDSP universe for Tier 1 screening (U.S. EPA, 2017). Some of these concerns have been addressed, while other issues remain in progress. EPA partially responded to the 2017 FIFRA SAP comments in Haggard et al. (2019) and as part of the detailed Response to Comments document (U.S. EPA, 2022a).

EPA is continuing to investigate the use of the HT H295R cell-based assay. Furthermore, EPA is continuing research on steroidogenesis. As this is considered a work in progress, EPA is not proposing any use of the HT H295R cell-based assay at this time.

Thomas et al. (2019) laid out a blueprint to guide the strategic and operational direction for computation toxicology research at EPA. This blueprint outlines a framework for the development and application of HT and computational modeling approaches in a tiered manner for chemical risk assessment including their application in the EDSP. Furthering the development and implementation of NAMs will dramatically enhance EPA's ability to meet testing needs, reduce costs, and eliminate or greatly reduce animal-based testing. EPA will continue to investigate NAMs as alternatives to other portions of the EDSP Tier 1 screening battery and actively communicate with stakeholders on the development and evaluation of EDSP NAMs that may be used for screening and priority setting.

IX. Conclusion

A summary of the conclusions of this document is as follows:

- The full estrogen receptor (ER) and androgen receptor (AR) pathway models have been validated, and the results from those models may be used as alternatives at this time for some Tier 1 assays (ER binding, estrogen receptor transcriptional activation (ERTA), and Uterotrophic [ER pathway model] and AR binding [AR pathway model]). For any particular chemical, the suitability of a model will be decided on a case-by-case basis considering the limitations of the models (see Section III.E.) and the properties of the chemical.
- All the New Approach Methodologies (NAMs)/ tools discussed in this paper (including full ER and AR pathway models, reduced ER and AR pathway models, Integration of Bioactivity and Exposure (IBER), Collaborative Estrogen Receptor Activity Prediction Project (CERAPP), and Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA)) may be used directly to prioritize chemicals for screening or to inform prioritization or hazard assessment (Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)).
- In some cases (considering the limitations of the model, additional available information, and EPA's guidance on Other Scientifically Relevant Information (OSRI)), the following NAMs may be considered OSRI and used during Weight of Evidence (WoE) evaluation to make decisions: reduced ER and AR pathway models, CERAPP, CoMPARA, SeqAPASS, and IBER.
- During WoE evaluation, which precedes Tier 2 testing, results from the Tier 1 battery, appropriate NAM alternatives, and OSRI are considered to determine which, if any, Tier 2 tests should be conducted. Thus, this WoE occurs between Tier 1 screening and any Tier 2 testing.
- None of these NAMs is meant to be alternatives for the current Tier 2 tests.
  The Agency will consider all public comments received on this document as it begins to implement these new approaches.

This document summarizes the scientific progress for the use of NAMs in the EDSP since 2015.

Screening. Given the strengths discussed in Section III.C as well as the uncertainties and limitations discussed in Section III.D., EPA has determined that the following NAMs may be used as alternatives for the following four (4) EDSP Tier 1 screening assays when evaluated on a chemical-by-chemical basis (each assay evaluated independently for each chemical). As part of the chemical evaluation process, EPA reviews the quality of available data across multiple lines of evidence. For the NAMs methods listed here, EPA will specifically consider the quality of the individual ToxCast assay data along with the level of confidence in and biological relevance of the predictions.

(1) The Estrogen Receptor (ER) pathway model based on the full 18-assay ToxCast/Tox21 battery (Browne et al., 2015; Judson et al., 2015) may be used as an alternative for three current EDSP Tier 1 screening assays:

- ER binding *in vitro* assay (OCSPP 890.1250; (U.S. EPA, 2009b)).
- ER transcriptional activation *in vitro* assay (ERTA; OCSPP 890.1300; (U.S. EPA, 2009c)).
- *In vivo* Uterotrophic assay (rat) (OCSPP 890.1600; (U.S. EPA, 2009d)).

(2) The Androgen Receptor (AR) pathway model based on the full 11-assay ToxCast/Tox21 battery (Kleinstreuer et al., 2017) may be used as an alternative for one current EDSP Tier 1 screening assay:

- AR binding *in vitro* assay (OCSPP 890.1150; (U.S. EPA, 2009a)).

Priority Setting. EPA is also using additional NAMs for priority setting of chemicals for EDSP Tier 1 screening or for use as OSRI in WoE evaluations. Priority setting may use NAMs singly or together with other available tools to prioritize chemicals for screening. These approaches are especially useful for prioritizing thousands of chemicals for which other sources of bioactivity and exposure information do not exist.

Recognizing the potential for uncertainties and limitations, the following NAMs may be used for priority setting of large sets of chemicals for EDSP Tier 1 screening or for consideration as use as OSRI in WoE evaluations.

(1) Additional ER and AR pathway models using assay subsets

(a) ER agonist assay subset pathway models (Judson et al., 2017).

(b) Revised (14-assay) and AR agonist and antagonist assay subset pathway models (Judson et al., 2020).

EPA continues to investigate the potential suitability of reduced ER and AR pathway models as alternatives to the EDSP Tier 1 screening assays.

(2) *In Silico* Qualitative Structural Activity Relationship Consensus Models for ER and AR (Mansouri et al., 2020; Mansouri et al., 2016).

(3) IBER, which combines estimates of the external dose potentially related to bioactivity results from the ER and AR pathway models with estimates of exposure (Friedman et al., 2020; Thomas et al., 2019; Bell and Wilson, 2018; Wambaugh et al., 2018; Sipes et al., 2017; Wetmore, 2015; U.S. EPA, 2014c; Wetmore et al., 2012; Rotroff et al., 2010).

(4) The Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool is an interspecies extrapolation tool (Lalone et al., 2018; Ankley et al., 2016). The SeqAPASS tool provides information that can be used to understand how broadly HT screening data (*e.g.*, ER) or AOPs may plausibly be extrapolated across species. For example, SeqAPASS could be used to extrapolate mammalian HT ER data to non-mammalian species and potentially reduce the need to request additional animal studies. This tool could be used to help prioritize data needs for EDSP Tier 1 screening.

EPA is in the process of refining a transparent, scientifically sound, and implementable approach for using NAMs for the mandatory screening of pesticides under FFDCA 408(p). This approach will consider the strengths, limitations, and uncertainties of the NAMs described in this White Paper in combination with the existing, validated assays in the EDSP tiered-framework and other potential OSRI (U.S. EPA, 2009e), (*e.g.*, exposure data, physical-chemical properties, toxicologically relevant studies in the published literature, QSAR models and other data submitted to support chemical assessment), as part of the WoE approach (U.S. EPA, 2011b) to determine whether additional data are needed.

X. References

Ankley, GT; Bennett, RS; Erickson, RJ; Hoff, DJ; Hornung, MW; Johnson, RD; Mount, DR; Nichols, JW; Russom, CL; Schmieder, PK; Serrrano, JA; Tietge, JE; Villeneuve, DL. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment [Review]. Environ Toxicol Chem 29: 730-741. http://dx.doi.org/10.1002/etc.34

Ankley, GT; Lalone, CA; Gray, LE; Villeneuve, DL; Hornung, MW. (2016). Evaluation of the scientific underpinnings for identifying estrogenic chemicals in nonmammalian taxa using mammalian test systems. Environ Toxicol Chem 35: 2806-2816. http://dx.doi.org/10.1002/etc.3456

Armitage, JM; Wania, F; Arnot, JA. (2014). Application of mass balance models and the chemical activity concept to facilitate the use of in vitro toxicity data for risk assessment. Environ Sci Technol 48: 9770-9779. http://dx.doi.org/10.1021/es501955g

Beames, T; Moreau, M; Roberts, LA; Mansouri, K; Haider, S; Smeltz, M; Nicolas, CI; Doheny, D; Phillips, MB; Yoon, M; Becker, RA; Mcmullen, PD; Andersen, ME; Clewell, RA; Hartman, JK. (2020). The role of fit-for-purpose assays within tiered testing approaches: A case study evaluating prioritized estrogen-active compounds in an in vitro human uterotrophic assay. Toxicol Appl Pharmacol 387: 114774. http://dx.doi.org/10.1016/j.taap.2019.114774

Bell, MJ; Wilson, P. (2018). Estimated differences in economic and environmental performance of forage-based dairy herds across the UK. 7. http://dx.doi.org/10.1002/fes3.127

Bell, SM; Chang, X; Wambaugh, JF; Allen, DG; Bartels, M; Brouwer, KLR; Casey, WM; Choksi, N; Ferguson, SS; Fraczkiewicz, G; Jarabek, AM; Ke, A; Lumen, A; Lynn, SG; Paini, A; Price, PS; Ring, C; Simon, TW; Sipes, NS; Sprankle, CS; Strickland, J; Troutman, J; Wetmore, BA; Kleinstreuer, NC. (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making [Review]. Toxicol In Vitro 47: 213-227. http://dx.doi.org/10.1016/j.tiv.2017.11.016

Birch, H; Kramer, NI; Mayer, P. (2019). Time-resolved freely dissolved concentrations of semivolatile and hydrophobic test chemicals in in vitro assays-measuring high losses and crossover by headspace solid-phase microextraction. Chem Res Toxicol 32: 1780-1790. http://dx.doi.org/10.1021/acs.chemrestox.9b00133

Black, SR; Nichols, JW; Fay, KA; Matten, SR; Lynn, SG. (2021). Evaluation and comparison of in vitro intrinsic clearance rates measured using cryopreserved hepatocytes from humans, rats, and rainbow trout. Toxicology 457: 152819. http://dx.doi.org/10.1016/j.tox.2021.152819

Blesson, CS; Sahlin, L. (2012). Expression pattern and signalling pathways in neutrophil like HL-60 cells after treatment with estrogen receptor selective ligands. Mol Cell Endocrinol 361: 179-190. http://dx.doi.org/10.1016/j.mce.2012.04.006

Breen, M; Ring, CL; Kreutz, A; Goldsmith, MR; Wambaugh, JF. (2021). High-throughput PBTK models for in vitro to in vivo extrapolation [Review]. Expert Opin Drug Metab Toxicol 17: 903-921. http://dx.doi.org/10.1080/17425255.2021.1935867

Browne, P; Judson, RS; Casey, W; Kleinstreuer, N; Thomas, RS. (2017). Correction to Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model [Erratum]. Environ Sci Technol 51: 9415. http://dx.doi.org/10.1021/acs.est.7b03317

Browne, P; Judson, RS; Casey, WM; Kleinstreuer, NC; Thomas, RS. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. Environ Sci Technol 49: 8804-8814. http://dx.doi.org/10.1021/acs.est.5b02641

Browne, P; Kleinstreuer, NC; Ceger, P; Deisenroth, C; Baker, N; Markey, K; Thomas, RS; Judson, RJ; Casey, W. (2018). Development of a curated Hershberger database [Review]. Reprod Toxicol 81: 259-271. http://dx.doi.org/10.1016/j.reprotox.2018.08.016

Casati, S; Aschberger, K; Barroso, J; Casey, W; Delgado, I; Kim, TS; Kleinstreuer, N; Kojima, H; Lee, JK; Lowit, A; Park, HK; Régimbald-Krnel, MJ; Strickland, J; Whelan, M; Yang, Y; Zuang, V. (2017). Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: Position of the International Cooperation on Alternative Test Methods. Arch Toxicol 92: 611-617. http://dx.doi.org/10.1007/s00204-017-2097-4

Casey, WM; Chang, X; Allen, DG; Ceger, PC; Choksi, NY; Hsieh, J; Wetmore, BA; Ferguson, SS; Devito, MJ; Sprankle, CS; Kleinstreuer, NC. (2018). Evaluation and Optimization of Pharmacokinetic Models for in Vitro to in Vivo Extrapolation of Estrogenic Activity for Environmental Chemicals. Environ Health Perspect 126: 97001. http://dx.doi.org/10.1289/EHP1655

Conley, JM; Hannas, BR; Furr, JR; Wilson, VS; Gray, LE. (2016). A Demonstration of the Uncertainty in Predicting the Estrogenic Activity of Individual Chemicals and Mixtures From an In Vitro Estrogen Receptor Transcriptional Activation Assay (T47D-KBluc) to the In Vivo Uterotrophic Assay Using Oral Exposure. Toxicol Sci 153: 382-395. http://dx.doi.org/10.1093/toxsci/kfw134

Degroot, DE; Swank, A; Thomas, RS; Strynar, M; Lee, M, iY; Carmichael, PL; Simmons, SO. (2018). mRNA transfection retrofits cell-based assays with xenobiotic metabolism. J Pharmacol Toxicol Methods 92: 77-94. http://dx.doi.org/10.1016/j.vascn.2018.03.002

Deisenroth, C; Degroot, DE; Zurlinden, T; Eicher, A; Mccord, J; Lee, MY; Carmichael, P; Thomas, RS. (2020). The Alginate Immobilization of Metabolic Enzymes Platform Retrofits an Estrogen Receptor Transactivation Assay With Metabolic Competence. Toxicol Sci 178: 281-301. http://dx.doi.org/10.1093/toxsci/kfaa147

Dix, DJ; Houck, KA; Martin, MT; Richard, AM; Setzer, RW; Kavlock, RJ. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95: 5-12. http://dx.doi.org/10.1093/toxsci/kfl103

EDSTAC (Endocrine Disruptor Screening and Testing and Advisory Committee). (1998). Endocrine Disruptor Screening and Testing Advisory Committee, Final Report, Volume I-II. https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-and-testing-advisory-committee-edstac-final

Fischer, FC; Henneberger, L; König, M; Bittermann, K; Linden, L; Goss, KU; Escher, BI. (2017). Modeling exposure in the tox21 in vitro bioassays. Chem Res Toxicol 30: 1197-1208. http://dx.doi.org/10.1021/acs.chemrestox.7b00023

Foradori, CD; Weiser, MJ; Handa, RJ. (2008). Non-genomic Actions of Androgens [Review]. Front Neuroendocrinol 29: 169-181. http://dx.doi.org/10.1016/j.yfrne.2007.10.005

Friedman, KP; Gagne, M; Loo, L; Karamertzanis, P; Netzeva, T; Sobanski, T; Franzosa, JA; Richard, A; Lougee, RR; Gissi, A; Lee, J; Angrish, M; Dorne, J; Foster, S; Raffaele, K; Bahadori, T; Gwinn, MR; Lambert, J; Whelan, M; Rasenberg, M; Barton-Maclaren, T; Thomas, RS. (2020). Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. Toxicol Sci 173: 202-225. http://dx.doi.org/10.1093/toxsci/kfz201

Fuentes, N; Silveyra, P. (2019). Estrogen receptor signaling mechanisms [Review]. 116: 135-170. http://dx.doi.org/10.1016/bs.apcsb.2019.01.001
Gray, LE, Jr; Furr, JR; Lambright, CS; Evans, N; Hartig, PC; Cardon, MC; Wilson, VS; Hotchkiss, AK; Conley, JM. (2020). Quantification of the uncertainties in extrapolating from in vitro androgen receptor antagonism to in vivo hershberger assay endpoints and adverse reproductive development in male rats. Toxicol Sci 176: 297-311. http://dx.doi.org/10.1093/toxsci/kfaa067
Haggard, DE; Karmaus, AL; Martin, MT; Judson, RS; Setzer, RW; Friedman, KP. (2018). High-throughput H295R steroidogenesis assay: Utility as an alternative and a statistical approach to characterize effects on steroidogenesis. Toxicol Sci 162: 509-534. http://dx.doi.org/10.1093/toxsci/kfx274
Haggard, DE; Setzer, RW; Judson, RS; Paul Friedman, K. (2019). Development of a prioritization method for chemical-mediated effects on steroidogenesis using an integrated statistical analysis of high-throughput H295R data. Regul Toxicol Pharmacol 109: 104510. http://dx.doi.org/10.1016/j.yrtph.2019.104510
Haselman, JT; Olker, JH; Kosian, PA; Korte, JJ; Swintek, JA; Denny, JS; Nichols, JW; Tietge, JE; Hornung, MW; Degitz, SJ. (2020). Targeted Pathway-based In Vivo Testing Using Thyroperoxidase Inhibition to Evaluate Plasma Thyroxine as a Surrogate Metric of Metamorphic Success in Model Amphibian Xenopus laevis. Toxicol Sci 175: 236-250. http://dx.doi.org/10.1093/toxsci/kfaa036
Hassan, I; El-Masri, H; Ford, J; Brennan, A; Handa, S; Paul Friedman, K; Gilbert, ME. (2020). Extrapolating In Vitro Screening Assay Data for Thyroperoxidase Inhibition to Predict Serum Thyroid Hormones in the Rat. Toxicol Sci 173: 280-292. http://dx.doi.org/10.1093/toxsci/kfz227
Henneberger, L; Muehlenbrink, M; Koenig, M; Schlichting, R; Fischer, FC; Escher, BI. (2019). Quantification of freely dissolved effect concentrations in in vitro cell-based bioassays. Arch Toxicol 93: 2295-2305. http://dx.doi.org/10.1007/s00204-019-02498-3
Henneberger, L; Mühlenbrink, M; Heinrich, DJ; Teixeira, A; Nicol, B; Escher, BI. (2020). Experimental validation of mass balance models for in vitro cell-based bioassays. Environ Sci Technol 54: 1120-1127. http://dx.doi.org/10.1021/acs.est.9b06144
Honda, M; Robinson, M; Kannan, K. (2019). A simple method for the analysis of neonicotinoids and their metabolites in human urine. Environ Chem 16: 171-178. http://dx.doi.org/10.1071/EN18240
Hopperstad, K; Degroot, DE; Zurlinden, T; Brinkman, C; Thomas, RS; Deisenroth, C. (2022). Chemical Screening in an Estrogen Receptor Transactivation Assay With Metabolic Competence. Toxicol Sci 187: 112-126. http://dx.doi.org/10.1093/toxsci/kfac019
Hu, R; Hilakivi-Clarke, L; Clarke, R. (2015). Molecular mechanisms of tamoxifen-associated endometrial cancer (Review). 9: 1495-1501. http://dx.doi.org/10.3892/ol.2015.2962
Huchthausen, J; Mühlenbrink, M; König, M; Escher, BI; Henneberger, L. (2020). Experimental exposure assessment of ionizable organic chemicals in in vitro cell-based bioassays. Chem Res Toxicol 33: 1845-1854. http://dx.doi.org/10.1021/acs.chemrestox.0c00067
Hutson, DD; Gurrala, R; Ogola, BO; Zimmerman, MA; Mostany, R; Satou, R; Lindsey, SH. (2019). Estrogen receptor profiles across tissues from male and female Rattus norvegicus. 10: 4. http://dx.doi.org/10.1186/s13293-019-0219-9
ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety

of Chemicals and Medical Products in the United States. http://dx.doi.org/10.22427/NTP-ICCVAM-ROADMAP2018
Judson, R; Houck, K; Friedman, KP; Brown, J; Browne, P; Johnston, PA; Close, DA; Mansouri, K; Kleinstreuer, N. (2020). Selecting a minimal set of androgen receptor assays for screening chemicals. Regul Toxicol Pharmacol 117: 104764. http://dx.doi.org/10.1016/j.yrtph.2020.104764
Judson, R; Houck, K; Martin, M; Richard, AM; Knudsen, TB; Shah, I; Little, S; Wambaugh, J; Setzer, RW; Kothya, P; Phuong, J; Filer, D; Smith, D; Reif, D; Rotroff, D; Kleinstreuer, N; Sipes, N; Xia, M; Huang, R; Crofton, K; Thomas, RS. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space [Erratum]. Toxicol Sci 152: 323-339. http://dx.doi.org/10.1093/toxsci/kfw092
Judson, RS; Houck, KA; Watt, ED; Thomas, RS. (2017). On selecting a minimal set of in vitro assays to reliably determine estrogen agonist activity. Regul Toxicol Pharmacol 91: 39-49. http://dx.doi.org/10.1016/j.yrtph.2017.09.022
Judson, RS; Kavlock, RJ; Setzer, RW; Hubal, EA; Martin, MT; Knudsen, TB; Houck, KA; Thomas, RS; Wetmore, BA; Dix, DJ. (2011). Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. Chem Res Toxicol 24: 451-462. http://dx.doi.org/10.1021/tx100428e
Judson, RS; Magpantay, FM; Chickarmane, V; Haskell, C; Tania, N; Taylor, J; Xia, M; Huang, R; Rotroff, DM; Filer, DL; Houck, KA; Martin, MT; Sipes, N; Richard, AM; Mansouri, K; Setzer, RW; Knudsen, TB; Crofton, KM; Thomas, RS. (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. http://dx.doi.org/10.1093/toxsci/kfv168
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. http://dx.doi.org/10.1093/toxsci/kfw002
Kavlock, R; Chandler, K; Houck, K; Hunter, S; Judson, R; Kleinstreuer, N; Knudsen, T; Martin, M; Padilla, S; Reif, D; Richard, A; Rotroff, D; Sipes, N; Dix, D. (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25: 1287-1302. http://dx.doi.org/10.1021/tx3000939
Kleinstreuer, NC; Browne, P; Chang, X; Judson, R; Casey, W; Ceger, P; Deisenroth, C; Baker, N; Markey, K; Thomas, RS. (2018). Evaluation of androgen assay results using a curated Hershberger database. Reprod Toxicol 81: 272-280. http://dx.doi.org/10.1016/j.reprotox.2018.08.017
Kleinstreuer, NC; Ceger, P; Watt, ED; Martin, M; Houck, K; Browne, P; Thomas, RS; Casey, WM; Dix, DJ; Allen, D; Sakamuru, S; Xia, M; Huang, R; Judson, R. (2017). Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30: 946-964. http://dx.doi.org/10.1021/acs.chemrestox.6b00347
Kleinstreuer, NC; Ceger, PC; Allen, DG; Strickland, J; Chang, X; Hamm, JT; Casey, WM. (2016). A curated database of rodent uterotrophic bioactivity. Environ Health Perspect 124: 556-562. http://dx.doi.org/10.1289/ehp.1510183
Lalone, CA; Villeneuve, DL; Doering, J, onA; Blackwell, BR; Transue, TR; Simmons, CW; Swintek, J, oe; Degitz, SJ; Williams, AJ; Ankley, GT. (2018). Evidence for cross species extrapolation of mammalian-based high-throughput screening assay results. Environ Sci Technol 52: 13960-13971. http://dx.doi.org/10.1021/acs.est.8b04587

Lalone, CA; Villeneuve, DL; Lyons, D; Helgen, HW; Robinson, SL; Swintek, JA; Saari, TW; Ankley, GT. (2016). Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity. Toxicol Sci228-245. http://dx.doi.org/10.1093/toxsci/kfw119

Mansouri, K; Abdelaziz, A; Rybacka, A; Roncaglioni, A; Tropsha, A; Varnek, A; Zakharov, A; Worth, A; Richard, AM; Grulke, CM; Trisciuzzi, D; Fourches, D; Horvath, D; Benfenati, E; Muratov, E; Wedebye, EB; Grisoni, F; Mangiatordi, GF; Incisivo, GM; Hong, H; Ng, HW; Tetko, IV; Balabin, I; Kancherla, J; Shen, J; Burton, J; Nicklaus, M; Cassotti, M; Nikolov, NG; Nicolotti, O; Andersson, PL; Zang, Q; Politi, R; Beger, RD; Todeschini, R; Huang, R; Farag, S; Rosenberg, SA; Slavov, S; Hu, X; Judson, RS. (2016). CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environ Health Perspect 124: 1023-1033. http://dx.doi.org/10.1289/ehp.1510267

Mansouri, K; Grulke, CM; Judson, RS; Williams, AJ. (2018). OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminform 10: 10. http://dx.doi.org/10.1186/s13321-018-0263-1

Mansouri, K; Kleinstreuer, N; Abdelaziz, AM; Alberga, D; Alves, VM; Andersson, PL; Andrade, CH; Bai, F; Balabin, I; D, B; E, B; B, B; S, B; J, C; V, C; S, F; D, F; AT, G-S; Gramatica, P; Grisoni, F; Grulke, CM; H, H; Horvath, D; Hu, X; Huang, R; Jeliazkova, N; Li, J; Li, X; Liu, H; Manganelli, S; Mangiatordi, GF; Maran, U; Marcou, G; Martin, T; Muratov, E; Nguyen, DT; Nicolotti, O; Nikolov, NG; Norinder, U; Papa, E; Petitjean, M; Piir, G; Pogodin, P; Poroikov, V; Qiao, X; Richard, AM; Roncaglioni, A; Ruiz, P; Rupakheti, C; Sakkiah, S; Sangion, A; Schramm, KW; Selvaraj, C; Shah, I; Sild, S; Sun, L; Taboureau, O; Tang, Y; Tetko, IV; Todeschini, R; Tong, W; Trisciuzzi, D; Tropsha, A; Van Den Driessche, G; Varnek, A; Wang, Z; Wedebye, EB; Williams, AJ; Xie, H; Zakharov, AV; Zheng, Z; Judson, RS. (2020). CoMPARA: Collaborative Modeling Project for Androgen Receptor Activity. Environ Health Perspect 128: 27002. http://dx.doi.org/10.1289/EHP5580

Mcmullen, PD; Andersen, ME; Cholewa, B; Clewell, HJ; Dunnick, KM; Hartman, JK; Mansouri, K; Minto, MS; Nicolas, CI; Phillips, MB; Slattery, S; Yoon, M; Clewell, RA. (2018). Evaluating opportunities for advancing the use of alternative methods in risk assessment through the development of fit-for-purpose in vitro assays [Review]. Toxicol In Vitro 48: 310-317. http://dx.doi.org/10.1016/j.tiv.2018.01.027

Noyes, PD; Paul-Friedman, KP; Haselman, JT; Barone Jr., S; Crofton, KM; Gilbert, ME; Hornung, MW; Laws, SC; Simmons, SO; Stoker, TE; Tietge, JE; Degitz, SJ. (2019). Evaluating chemicals for thyroid disruption: Opportunities and challenges with in vitro testing and adverse outcome pathway approaches. Environ Health Perspect 127: 95001. http://dx.doi.org/10.1289/EHP5297

NRC (National Research Council). (2007). Toxicity testing in the 21st century: A vision and a strategy. In Toxicity testing in the 21st century: A vision and a strategy. Washington, DC: National Academies Press. http://dx.doi.org/10.17226/11970

OECD (Organisation for Economic Co-operation and Development). (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. In OECD SERIES ON TESTING AND ASSESSMENT. Paris, France. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2005)14&doclanguage=en

OECD (Organisation for Economic Co-operation and Development). (2018a). Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. http://dx.doi.org/10.1787/9789264304741-en
OECD (Organisation for Economic Co-operation and Development). (2018b). Uterotrophic bioassay in rodents (UT assay) (OECD TG 440) (including OECD GD 71 on the procedure to test for anti-estrogenicity). In Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Paris, France: Organisation for Economic Co-operation and Development :: OECD. http://dx.doi.org/10.1787/9789264304741-20-en
OECD (Organisation for Economic Co-operation and Development). (2019). Guidance document on aquatic toxicity testing of difficult substances and mixtures. (OECD Series on Testing and Assessment No. 23 (second Edition); JT03442844). Paris, France. http://dx.doi.org/10.1787/0ed2f88e-en
Olker, JH; Korte, JJ; Denny, JS; Hartig, PC; Cardon, MC; Knutsen, CN; Kent, PM; Christensen, JP; Degitz, SJ; Hornung, MW. (2019). Screening the ToxCast phase 1, phase 2, and e1k chemical libraries for inhibitors of iodothyronine deiodinases. Toxicol Sci 168: 430-442. http://dx.doi.org/10.1093/toxsci/kfy302
Olker, JH; Korte, JJ; Denny, JS; Haselman, JT; Hartig, PC; Cardon, MC; Hornung, MW; Degitz, SJ. (2021). In vitro screening for chemical inhibition of the iodide recycling enzyme, iodotyrosine deiodinase. Toxicol In Vitro 71: 105073. http://dx.doi.org/10.1016/j.tiv.2020.105073
Patlewicz, G; Ball, N; Boogaard, PJ; Becker, RA; Hubesch, B. (2015). Building scientific confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol 72: 117-133. http://dx.doi.org/10.1016/j.yrtph.2015.03.015
Patlewicz, G; Simon, T, ed; Goyak, K; Phillips, RD; Rowlands, JC; Seidel, SD; Becker, RA. (2013). Use and validation of HT/HC assays to support 21st century toxicity evaluations. Regul Toxicol Pharmacol 65: 259-268. http://dx.doi.org/10.1016/j.yrtph.2012.12.008
Paul-Friedman, K; Martin, M; Crofton, KM; Hsu, CW; Sakamuru, S; Zhao, J; Xia, M; Huang, R; Stavreva, DA; Soni, V; Varticovski, L; Raziuddin, R; Hager, GL; Houck, KA. (2019). Limited Chemical Structural Diversity Found to Modulate Thyroid Hormone Receptor in the Tox21 Chemical Library. Environ Health Perspect 127: 97009. http://dx.doi.org/10.1289/EHP5314
Perkins, EJ; Ankley, GT; Crofton, KM; Garcia-Reyero, N; Lalone, CA; Johnson, MS; Tietge, JE; Villeneuve, DL. (2013). Current Perspectives on the Use of Alternative Species in Human Health and Ecological Hazard Assessments [Review]. Environ Health Perspect 121: 1002-1010. http://dx.doi.org/10.1289/ehp.1306638
Pham, NNT; Park, JS; Kim, H, eeTak; Kim, HJ; Son, YA; Kang, SG, u; Lee, SG. (2019). Catalytic performance of graphene quantum dot supported manganese phthalocyanine for efficient oxygen reduction: density functional theory approach. New J Chem 43: 348-355. http://dx.doi.org/10.1039/c8nj05093g
Pinto, CL; Mansouri, K; Judson, R; Browne, P. (2016). Prediction of Estrogenic Bioactivity of Environmental Chemical Metabolites. Chem Res Toxicol 29: 1410-1427. http://dx.doi.org/10.1021/acs.chemrestox.6b00079
Richard, AM; Judson, RS; Houck, KA; Grulke, CM; Volarath, P; Thillainadarajah, I; Yang, C; Rathman, J; Martin, MT; Wambaugh, JF; Knudsen, TB; Kancherla, J; Mansouri, K; Patlewicz, G; Williams, AJ; Little, SB; Crofton, KM; Thomas, RS. (2016). ToxCast

chemical landscape: Paving the road to 21st century toxicology. Chem Res Toxicol 29: 1225-1251. http://dx.doi.org/10.1021/acs.chemrestox.6b00135

Rotroff, DM; Dix, DJ; Houck, KA; Knudsen, TB; Martin, MT; Mclaurin, KW; Reif, DM; Crofton, KM; Singh, AV; Xia, M; Huang, R; Judson, RS. (2013). Using in Vitro High Throughput Screening Assays to Identify Potential Endocrine-Disrupting Chemicals. Environ Health Perspect 121: 7-14. http://dx.doi.org/10.1289/ehp.1205065

Rotroff, DM; Wetmore, BA; Dix, DJ; Ferguson, SS; Clewell, HJ; Houck, KA; Lecluyse, EL; Andersen, ME; Judson, RS; Smith, CM; Sochaski, MA; Kavlock, RJ; Boellmann, F; Martin, MT; Reif, DM; Wambaugh, JF; Thomas, RS. (2010). Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. Toxicol Sci 117: 348-358. http://dx.doi.org/10.1093/toxsci/kfq220

Sipes, NS; Wambaugh, JF; Pearce, R; Auerbach, SS; Wetmore, BA; Hsieh, J; Shapiro, AJ; Svoboda, D; Devito, MJ; Ferguson, SS. (2017). An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library. Environ Sci Technol 51: 10786-10796. http://dx.doi.org/10.1021/acs.est.7b00650

Thomas, RS; Bahadori, T; Buckley, TJ; Cowden, J; Deisenroth, C; Dionisio, KL; Frithsen, JB; Grulke, CM; Gwinn, MR; Harrill, JA; Higuchi, M; Houck, KA; Hughes, MF; Hunter, ES; Isaacs, KK; Judson, RS; Knudsen, TB; Lambert, JC; Linnenbrink, M; Martin, TM; Newton, SR; Padilla, S; Patlewicz, G; Paul-Friedman, K; Phillips, KA; Richard, AM; Sams, R; Shafer, TJ; Setzer, RW; Shah, I; Simmons, JE; Simmons, SO; Singh, A; Sobus, JR; Strynar, M; Swank, A; Tornero-Valez, R; Ulrich, EM; Villeneuve, DL; Wambaugh, JF; Wetmore, BA; Williams, AJ. (2019). The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. Toxicol Sci 169: 317-332. http://dx.doi.org/10.1093/toxsci/kfz058

Thomas, RS; Paules, RS; Simeonov, A; Fitzpatrick, SC; Crofton, KM; Casey, WM; Mendrick, DL. (2018). The US Federal Tox21 Program: A strategic and operational plan for continued leadership. ALTEX 35: 163-168. http://dx.doi.org/10.14573/altex.1803011

U.S. EPA (U.S. Environmental Protection Agency). (1998). Endocrine disruptor screening and testing advisory committee final report [EPA Report]. Washington, DC. https://www.epa.gov/sites/production/files/2015-08/documents/exesum14.pdf

U.S. EPA. (1999). Review of the EPA's Proposed Environmental Endocrine Disruptor Screening Program: Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. https://sab.epa.gov/ords/sab/f?p=114:0:1938237319604:APPLICATION_PROCESS=REPORT_DOC:::REPORT_ID:113

U.S. EPA (U.S. Environmental Protection Agency). (2005). Guidance for thyroid assays in pregnant animals, fetuses and postnatal animals, and adult animals. Washington, DC: Office of Pesticide Programs, Health Effects Division. https://www.epa.gov/sites/production/files/2015-06/documents/thyroid_guidance_assay.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2009a). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1150: Androgen receptor binding (rat prostate cytosol) [EPA Report]. (EPA 640-C-09-003). Washington, D.C.: U.S. EPA; Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0003

U.S. EPA (U.S. Environmental Protection Agency). (2009b). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1250: Estrogen receptor binding assay using rat

uterine cytosol (ER-RUC) [EPA Report]. (EPA 740-C-09-005). Washington, D.C.: U.S. EPA; Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0005

U.S. EPA (U.S. Environmental Protection Agency). (2009c). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1300: Estrogen receptor transcriptional activation (human cell line (HeLa-9903)) [EPA Report]. (EPA 740-C-09-006). Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2009-0576-0006

U.S. EPA (U.S. Environmental Protection Agency). (2009d). Endocrine Disruptor Screening Program test guidelines: OPPTS 890.1600: Uterotrophic assay [EPA Report]. (EPA 740-C-09-0010). Washington D.C.: U.S. Environmental Protection Agency, Prevention, Pesticides, and Toxic Substances. http://www.regulations.gov/#!documentDetail;D=EPA-HQ-OPPT-2009-0576-0012

U.S. EPA (U.S. Environmental Protection Agency). (2009e). EPA's Approach for Considering Other Scientifically Relevant Information (OSRI) under the Endocrine Disrupt or Screening Program. Washington, D.C.: U.S. EPA, Office of Chemical Safety and Pollution Prevention. https://www.regulations.gov/document/EPA-HQ-OPPT-2007-1080-0032

U.S. EPA (U.S. Environmental Protection Agency). (2011a). Hershberger assay, OCSPP guideline 890.1400: Standard evaluation procedure (SEP) [EPA Report]. Washington, DC. https://www.epa.gov/sites/production/files/2015-07/documents/final_890.1600_hershberger_assay_sep_10.6.11.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2011b). Weight-of-evidence: Evaluating results of EDSP tier 1 screening to identify the need for tier 2 testing [EPA Report]. (EPA-HQ-OPPT-2010-0877-0021). Washington, DC. https://www.regulations.gov/document?D=EPA-HQ-OPPT-2010-0877-0021

U.S. EPA (U.S. Environmental Protection Agency). (2012). Endocrine disruptor screening program universe of chemicals and general validation principles [EPA Report]. https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-universe-chemicals-and-general

U.S. EPA (U.S. Environmental Protection Agency). (2013). January 2013 FIFRA SAP Meeting - Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals using Computational Toxicology Tools. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2012-0818

U.S. EPA (U.S. Environmental Protection Agency). (2014a). December 2014 FIFRA SAP Meeting - Endocrine activity and exposure-based prioritization and screening [EPA Report]. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2014-0614

U.S. EPA (U.S. Environmental Protection Agency). (2014b). Integrated bioactivity and exposure ranking: A computational approach for the prioritization and screening of chemicals in the endocrine disruptor screening program [EPA Report]. (EPA-HQ-OPP-2014-0614-0003). https://www.regulations.gov/document?D=EPA-HQ-OPP-2014-0614-0003

U.S. EPA (U.S. Environmental Protection Agency). (2014c). July/August 2014 FIFRA SAP Meeting - Scientific issues associated with new high throughput methods to estimate chemical exposure [EPA Report]. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2014-0331

U.S. EPA (U.S. Environmental Protection Agency). (2015a). Use of high throughput assays and computational tools: Endocrine Disruptor Screening Program; Notice of availability and opportunity for comment, 80 fed. reg. 118 (June 19, 2015) [EPA Report] (pp. 35350-35355). (EPA-HQ-OPPT-2015-0305). Washington, D.C.: Federal Register. https://www.federalregister.gov/articles/2015/06/19/2015-15182/use-of-high-throughput-assays-and-computational-tools-endocrine-disruptor-screening-program-notice

U.S. EPA (U.S. Environmental Protection Agency). (2015b). Use of high throughput assays and computational tools; endocrine disruptor screening program; notice of availability and opportunity for comment. Fed Reg 80: 35350-35355.

U.S. EPA (U.S. Environmental Protection Agency). (2017). November 2017 FIFRA SAP meeting - continuing development of alternative high-throughput screens to determine endocrine disruption, focusing on androgen receptor, steroidogenesis, and thyroid pathways. https://www.regulations.gov/docket?D=EPA-HQ-OPP-2017-0214

U.S. EPA (U.S. Environmental Protection Agency). (2018a). Alternative Test Methods and Strategies to Reduce Vertebrate Animal Testing. Available online at https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce

U.S. EPA (U.S. Environmental Protection Agency). (2018b). Strategic plan to promote the development and implementation of alternative test methods within the TSCA program (pp. 1-38). (EPA-740-R1-8004). https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2019). Directive to Prioritize Efforts to Reduce Animal Testing [EPA Report]. https://www.epa.gov/sites/production/files/2019-09/documents/image2019-09-09-231249.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2021a). EPA new approach methods work plan: Reducing use of animals in chemical testing (pp. 1-21). (EPA 600/X-21/209). Washington, DC. https://www.epa.gov/chemical-research/epa-new-approach-methods-work-plan-reducing-use-animals-chemical-testing

U.S. EPA (U.S. Environmental Protection Agency). (2021b). EPA's Endocrine Disruptor Screening Program Has Made Limited Progress in Assessing Pesticides. (Report No. 21-E-0186). https://www.epa.gov/system/files/documents/2021-07/_epaoig_20210728-21-e-0186.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2022a). The EPA's Responses to Panel Members at the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific2Advisory Panel (SAP) Meeting from November 28-30, 2017.

U.S. EPA (U.S. Environmental Protection Agency). (2022b). Response to Public Comments on the Public Review Draft: "Use of High Throughput Assays and Computational Tools; Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment".

Wambaugh, JF; Hughes, MF; Ring, CL; MacMillan, DK; Ford, J; Fennell, TR; Black, SR; Snyder, RW; Sipes, NS; Wetmore, BA; Westerhout, J; Setzer, RW; Pearce, RG; Simmons, JE; Thomas, RS. (2018). Evaluating in vitro-in vivo extrapolation of toxicokinetics. Toxicol Sci 163: 152-169. http://dx.doi.org/10.1093/toxsci/kfy020

Wambaugh, JF; Setzer, RW; Reif, DM; Gangwal, S; Mitchell-Blackwood, J; Arnot, JA; Joliet, O; Frame, A; Rabinowitz, J; Knudsen, TB; Judson, RS; Egeghy, P; Vallero, D; Cohen Hubal, EA. (2013). High-throughput models for exposure-based chemical prioritization

in the ExpoCast project. Environ Sci Technol 47: 8479-8488. http://dx.doi.org/10.1021/es400482g
Wambaugh, JF; Wang, A; Dionisio, KL; Frame, A; Egeghy, P; Judson, R; Setzer, RW. (2014). High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals. Environ Sci Technol 48: 12760-12767. http://dx.doi.org/10.1021/es503583j
Wambaugh, JF; Wetmore, BA; Pearce, R; Strope, C; Goldsmith, R; Sluka, JP; Sedykh, A; Tropsha, A; Bosgra, S; Shah, I; Judson, R; Thomas, RS; Setzer, RW. (2015). Toxicokinetic Triage for Environmental Chemicals. Toxicol Sci 147: 55-67. http://dx.doi.org/10.1093/toxsci/kfv118
Wambaugh, JF; Wetmore, BA; Ring, CL; Nicolas, C; Pearce, RG; Honda, GS; Dinallo, R; Angus, D; Gilbert, J; Sierra, T; Badrinarayanan, A; Snodgrass, B; Brockman, A; Strock, C; Setzer, RW; Thomas, RS. (2019). Assessing toxicokinetic uncertainty and variability in risk prioritization. Toxicol Sci 172: 235-251. http://dx.doi.org/10.1093/toxsci/kfz205
Wang, J; Hallinger, DR; Murr, AS; Buckalew, AR; Lougee, RR; Richard, AM; Laws, SC; Stoker, TE. (2019). High-throughput screening and chemotype-enrichment analysis of ToxCast phase II chemicals evaluated for human sodium-iodide symporter (NIS) inhibition. Environ Int 126: 377-386. http://dx.doi.org/10.1016/j.envint.2019.02.024
Watt, E. (2017). toxboot: bootstrap methods for ToxCast high throughput screening data. Retrieved from https://github.com/ericwatt/toxboot
Watt, ED; Judson, RS. (2018). Uncertainty quantification in ToxCast high throughput screening. PLoS ONE 13: e0196963. http://dx.doi.org/10.1371/journal.pone.0196963
Wegner, SH; Pinto , C; Ring, CL; Wambaugh, JF. (2020). High-throughput screening tools facilitate calculation of a combined exposure-bioactivity index for chemicals with endocrine activity. Environ Int 137: 105470-105470. http://dx.doi.org/10.1016/j.envint.2020.105470
Wetmore, BA. (2015). Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment [Review]. Toxicology 332: 94-101. http://dx.doi.org/10.1016/j.tox.2014.05.012
Wetmore, BA; Allen, B; Clewell, HJ, III; Parker, T; Wambaugh, JF; Almond, LM; Sochaski, MA; Thomas, RS. (2014). Incorporating population variability and susceptible subpopulations into dosimetry for high-throughput toxicity testing. Toxicol Sci 142: 210-224. http://dx.doi.org/10.1093/toxsci/kfu169
Wetmore, BA; Wambaugh, JF; Allen, B; Ferguson, SS; Sochaski, MA; Setzer, RW; Houck, KA; Strope, CL; Cantwell, K; Judson, RS; LeCluyse, E; Clewell, HJ; Thomas, RS; Andersen, ME. (2015). Incorporating high-throughput exposure predictions with dosimetry-adjusted in vitro bioactivity to inform chemical toxicity testing. Toxicol Sci 148: 121-136. http://dx.doi.org/10.1093/toxsci/kfv171
Wetmore, BA; Wambaugh, JF; Ferguson, SS; Li, L; Clewell, HJ; Judson, RS; Freeman, K; Bao, W; Sochaski, MA; Chu, TM; Black, MB; Healy, E; Allen, B; Andersen, ME; Wolfinger, RD; Thomas, RS. (2013). Relative impact of incorporating pharmacokinetics on predicting in vivo hazard and mode of action from high-throughput in vitro toxicity assays. Toxicol Sci 132: 327-346. http://dx.doi.org/10.1093/toxsci/kft012
Wetmore, BA; Wambaugh, JF; Ferguson, SS; Sochaski, MA; Rotroff, DM; Freeman, K; Clewell, HJ; Dix, DJ; Andersen, ME; Houck, KA; Allen, B; Judson, RS; Singh, R; Kavlock, RJ; Richard, AM; Thomas, RS. (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125: 157-174. http://dx.doi.org/10.1093/toxsci/kfr254

WHO (World Health Organization). (2002). GLOBAL ASSESSMENT SSESSMENT OF THE STATE-OF-THE-SCIENCE OF ENDOCRINE NDOCRINE DISRUPTORS ISRUPTOR. (WHO/PCS/EDC/02.2). World Health Organisation. https://apps.who.int/iris/handle/10665/67357

Williams, AJ; Grulke, CM; Edwards, J; McEachran, AD; Mansouri, K; Baker, NC; Patlewicz, G; Shah, I; Wambaugh, JF; Judson, RS; Richard, AM. (2017). The CompTox Chemistry Dashboard: A community data resource for environmental chemistry. J Cheminform 9: 61. http://dx.doi.org/10.1186/s13321-017-0247-6