UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF CHEMICAL SAFETY AND
POLLUTION PREVENTION

MEMORANDUM

DATE: May 1, 2013

SUBJECT: Transmittal of the Meeting Minutes of the FIFRA SAP Meeting Held January 29-31, 2013 on the Scientific Issues Associated with "Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools"

TO: Steven Knott, Acting Director
Office of Science and Coordination Policy

FROM: Sharlene Matten, Ph.D.
Designated Federal Official
FIFRA SAP Staff
Office of Science Coordination and Policy

THRU: Laura Bailey, M.S.
Executive Secretary, FIFRA SAP
Office of Science Coordination and Policy

Please find attached the meeting report of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) open meeting held in Arlington, Virginia on January 29-31, 2013. This report addresses a set of scientific issues associated with "Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools". An electronic copy of this report is available at the FIFRA SAP website http://www.epa.gov/scipoly/sap .

Attachment

cc:
Jim Jones
Louise Wise
Stephen Bradbury
Vicki Dellarco
William Jordan
Margie Fehrenbach
Donald Brady
Susan Lewis
Jack Housenger
Richard Keigwin, Jr.
Keith Matthews
Robert McNally
Mark Harman
Lois Rossi
Jess Rowland
Oscar Morales
Enesta Jones
Mary Manibusan
Ray Kent
Pat Schmieder
Rick Kolanczyk
Richard Judson
Michael Horning
Susan Laws
Glenn Paulson
Dale Kemery
Molly Hooven
OPP Regulatory Docket

FIFRA SAP Members
Daniel Schlenk, Ph.D. (FIFRA SAP Chair)
Marion Ehrich, Ph.D., DABT, ATS
Stephen Klaine, Ph.D.
Jim McManaman, Ph.D.
Prakash Nagarkatti
Martha Sandy, Ph.D.

FQPA Science Review Board Members
Thomas P. Burris, Ph.D., F.A.H.A.,
Janice E. Chambers, Ph.D., D.A.B.T., A.T.S.
Mark T. D. Cronin, Ph.D.
Nancy Derrick Denslow, Ph.D.
Miriam N. Jacobs, Ph.D.
Michael J. Keiser, Ph.D.
M.E. (Bette) Meek, Ph.D.
Edward J. Perkins, Ph.D.
Stephen H. Safe, Ph.D.
Terry Schultz, Ph.D.
Katrina M. Waters, Ph.D.

SAP Minutes No. 2013-01

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools

January 29-31, 2013
FIFRA Scientific Advisory Panel Meeting
Held at
One Potomac Yard
Arlington, Virginia

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). The meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of the meeting minutes does not represent information approved by the Agency. The meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of these meeting minutes do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the Environmental Protection Agency (EPA), Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Sharlene R. Matten, Ph.D., SAP Designated Federal Official, via e-mail at matten.sharlene@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented in public comment. This document addresses the information provided and presented by EPA within the structure of the charge.

TABLE OF CONTENTS

SAP Minutes No. 2013-01

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools

January 29-31, 2013
FIFRA Scientific Advisory Panel Meeting
Held at
One Potomac Yard
Arlington, Virginia

Daniel Schlenk, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel

Date: 4/27/13

Sharlene R. Matten, Ph.D.
Designated Federal Official
FIFRA Scientific Advisory Panel
Staff

Date: 4/30/13

Panel Member List for the Meeting of the Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) to Consider and Review

Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools

January 29-31, 2013

EPA-HQ-OPP-2012-0818 **OPP Docket Tel: 703-305-5805**

FIFRA SAP Chair

Daniel Schlenk, Ph.D.
Professor of Aquatic Ecotoxicology & Environmental Toxicology
Department of Environmental Sciences
University of California, Riverside
Riverside, CA

Designated Federal Official

Sharlene R. Matten, Ph.D.
U.S. Environmental Protection Agency
Office of Science Coordination & Policy
FIFRA Scientific Advisory Panel
EPA East Building, MC 7201M
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Tel: 202-564-8450, E-mail: matten.sharlene@epa.gov

FIFRA Scientific Advisory Panel Members

Marion F. Ehrich, Ph.D., D.A.B.T., A.T.S.
Professor of Pharmacology and Toxicology
Department of Biomedical Science and Pathobiology
Virginia-Maryland Regional College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, VA

Stephen J. Klaine, Ph.D.
Professor and Director
Clemson University Institute of Environmental Toxicology
Department of Biological Sciences
Pendleton, SC

James L. McManaman, Ph.D.
Professor and Chief
Section of Basic Reproductive Sciences
Departments of Obstetrics and Gynecology & Physiology and Biophysics
University of Colorado-Denver
Aurora, CO

Prakash Nagarkatti, Ph.D.
Vice President for Research
Carolina Distinguished Professor
University of South Carolina
Columbia, SC

Martha S. Sandy, Ph.D.
Senior Toxicologist and Chief
Cancer Toxicology and Epidemiology Section
Reproductive & Cancer Hazard Assessment Branch
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency
Oakland, CA

FQPA Science Review Board Members

Thomas P. Burris, Ph.D., F.A.H.A., F.A.A.A.S.
Professor
Departments of Molecular Therapeutics & Metabolism and Aging
Director, Center for Diabetes and Metabolic Diseases
The Scripps Research Institute
Jupiter, FL

Janice E. Chambers, Ph.D., D.A.B.T., A.T.S.
William L. Giles Distinguished Professor
Director, Center for Environmental Health Sciences
College of Veterinary Medicine
Mississippi State University
Mississippi State, MS

Mark T. D. Cronin, Ph.D.
Professor of Predictive Toxicology
School of Pharmacy and Biomolecular Sciences
Liverpool John Moores University
Liverpool, England, United Kingdom

Nancy Derrick Denslow, Ph.D.
Professor
Department of Physiological Sciences
University of Florida
Gainesville, FL

Miriam N. Jacobs, Ph.D.
Scientific Officer
European Food Safety Authority (EFSA)
Scientific Committee Unit
Parma, Italy

Michael J. Keiser, Ph.D.
President & Chief Operating Officer
SeaChange Pharmaceuticals, Inc.
San Francisco, CA

M.E. (Bette) Meek, Ph.D.
Associate Director
Chemical Risk Assessment
McLaughlin Centre for Population Health Risk Assessment
University of Ottawa
Ottawa, Ontario, Canada

Edward J. Perkins, Ph.D.
Senior Research Scientist (ST)
Environmental Networks and Genetic Toxicology
Environmental Laboratory
U.S. Army Engineer Research and Development Center
U.S. Army Corps Engineers
Vicksburg, MS

Stephen H. Safe, Ph.D.
Distinguished Professor
Department of Veterinary Physiology and Pharmacology
Texas A&M University
College Station, TX

Terry Schultz, Ph.D.
Professor Emeritus
Department of Comparative Medicine
College of Veterinary Medicine
University of Tennessee
Knoxville, TN

Katrina M. Waters, Ph.D.
Senior Staff Scientist
Technical Group Manager/Associate Division Director
Computational Sciences & Mathematics Division
Pacific Northwest National Laboratory
Richland, WA

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed its report of the SAP meeting regarding scientific issues associated with **"Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools."** Advance notice of the SAP meeting was published in the *Federal Register* on **November 16, 2012 (Vol. 77, No. 222)**. The review was conducted in an open Panel meeting on **January 29-31, 2013** at One Potomac Yard, Arlington, Virginia. Materials for this meeting are available in the Office of Pesticide Programs (OPP) public docket or via www.regulations.gov, **Docket No. EPA-HQ-OPP-2012-0818**. The FIFRA SAP Chair, Daniel Schlenk, Ph.D., chaired the meeting. Sharlene Matten, Ph.D., served as the Designated Federal Official. Louise Wise, J.D., Deputy Associate Administrator, Office of Chemical Safety and Pollution Prevention and Steven Bradbury, Ph.D., Director, Office of Pesticide Programs, provided opening remarks at the meeting. An overview of the EDSP21 work plan and prioritization scheme was provided by Mary Manibusan, Director, Exposure Assessment Coordination and Policy Division, Office of Science Coordination and Policy. Technical presentations were provided by the following individuals:

- Ray Kent, Ph.D., Health Effects Division, Office of Pesticide Programs, EPA;
- Patricia Schmieder, Ph.D. and Rick Kolanczyk, Mid-Continent Ecology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, EPA;
- Richard Judson, Ph.D., National Center for Computational Toxicology, Office of Research and Development, EPA; and,
- Michael Hornung, Ph.D. Mid-Continent Ecology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, EPA.

Vicki Dellarco, Ph.D., Science Advisor, Office of Pesticide Programs, presented a summary of the issues under discussion to conclude the EPA's presentations.

Background

Under 1996 Federal Food Drug and Cosmetic Act 408(p) and Safe Drinking Water Act, section 1457, the U.S. EPA is required to screen all pesticide chemicals (active and inert ingredients) and only those drinking water contaminants to which a 'substantial population" is exposed for the potential to interact with the endocrine system. The combination of both pesticide and drinking water chemicals amounts to a universe of over 10,000 chemicals. The Agency must therefore prioritize which chemicals are listed for screening and testing in multiple biologically complex and resource intensive assays. The Agency is developing the use of structure-activity relationships (SARs) and high-throughput (HTP) assay data as cost effective prioritization tools that can help focus the generation of new data on chemicals that are more likely to have the potential to interact with the estrogen, androgen or thyroid pathways.

In 2009, the EPA developed the Estrogen Receptor (ER) Expert System (ES) to help prioritize candidate chemicals specific to an estrogen receptor (ER) mediated pathway for testing in the Tier 1

Endocrine Disruptor Screening Program. The ERES was initially developed for pesticide, food use inert ingredients and antimicrobial pesticides and predicts ER binding affinity based on data derived from two types of *in vitro* assays: ER binding and ER transcriptional assays. The FIFRA SAP reviewed the initial ERES approach including the transparency of the approach, biological endpoint definition, mechanistic interpretation, and model's applicability domain (http://www.epa.gov/scipoly/sap/meetings/2009/august/082509minutes.pdf). The EPA incorporated the 2009 SAP's recommendations and expanded the ERES training set through the use of chemical read across techniques (*i.e.*, using data from a tested chemical for a particular property or effect to a similar untested chemical) and *in vitro* HTP assay data. While this SAP is focused on the estrogenic toxicity pathway, the recommendations may be more broadly applied for additional toxicity pathways (*e.g.*, androgen and thyroid). In addition, this SAP review will establish a solid foundation upon which the Agency seeks to build upon as we advance the use of these computational toxicology tools beyond EDSP chemical prioritization and into Tier 1 data replacement.

The EPA asked for advice on the following specific issues:

- Rationale and transparency of the overall prioritization approach;
- Use of physico-chemical properties to exclude chemicals from EDSP screening and testing;
- Expansion of the ERES training set through the use of read across techniques to cover a larger chemical universe;
- Transparency of the HTP techniques (*e.g.,* how the data are processed and interpreted and the extent of reliability and consistency within an assay and among HTP assays);
- Performance of the *in vitro* HTP assays against a set of reference chemicals;
- Concordance of results of the ERES training set and HTP assays;
- Use of a category based approach and strategic testing to cover chemicals with unknown binding potential in the ERES;
- How to address active metabolites in computer-based simulations and *in vitro* testing; and,
- Use of the Adverse Outcome Pathway (AOP) and category-based testing strategies for other endocrine pathways.

PUBLIC COMMENTERS

Oral statements were presented by:

1) Scott Slaughter, Center for Regulatory Effectiveness
2) Kristie Sullivan, MPH, Physicians Committee for Responsible Medicine
3) Catherine Willett, PhD, The Humane Society of the United States
4) Richard A. Becker, PhD, DABT; American Chemical Council
5) Ted Simon, PhD, DABT; Ted Simon, LLC on behalf of the American Chemistry Council
6) Terry F. Quill, Quill Law Group, LLC on behalf of the American Chemistry Council
7) Ellen Mihaich, PhD, DABT; Environmental and Regulatory Resources on behalf of the Endocrine Policy Forum
8) Christopher J. Borgert, PhD, Applied Pharmacology and Toxicology, Inc. on behalf of the Endocrine Policy Forum
9) Lisa S. Ortego, PhD, DABT; Bayer CropScience on behalf of Crop Life America
10) Sue Marty, PhD, DABT; Dow Chemical Company
11) Katy O. Goyak, PhD, ExxonMobil Biomedical Sciences, Inc.
12) David R. Geter, PhD, Bayer CropScience

Written statements were provided by:

1) Clare Thorpe, Ph.D., CropLife America and Administrator of the Endocrine Policy Forum, on behalf of the Endocrine Policy Forum
2) Richard A. Becker, PhD, DABT, American Chemical Council and Clare Thorpe, Ph.D. CropLife America on behalf of the American Chemistry Council and CropLife America
3) Scott Slaughter, Center for Regulatory Effectiveness
4) Kristie Sullivan, MPH, Physicians Committee for Responsible Medicine and Catherine Willett, PhD, The Humane Society of the United States on behalf of the Physicians Committee for Responsible Medicine and The Humane Society of the United States
5) Diana Slater, private citizen

SELECTED ABBREVIATIONS

ADME	Absorption, Distribution, Metabolism, and Excretion
AhR	Aryl hydrocarbon Receptor
AR	Androgen Receptor
AOP	Adverse Outcome Pathway
CAR	Constitutive Androstane Receptor
CYP	Cytochrome P450
DSL	Domestic Substances List (Canada)
ED	Endocrine Disruptor
EDSP	Endocrine Disruptor Screening Program (USEPA)
EFSA	European Food Safety Authority
ER	Estrogen Receptor [Note there are two forms of the ER, alpha and beta, which are encoded by different genes. While they share about 95% amino acid homology in the DNA binding domain, they only share about 53% homology when comparing the E/F domains which includes the ligand binding domain and ligand-dependent transactivation domain.]
ER A	A site on the estrogen receptor
ER B	B site on the estrogen receptor
ER C	C site on the estrogen receptor
ER-alpha	Estrogen receptor alpha is a specific estrogen receptor isoform that binds to estrogen and estrogen-like mimics for transactivation. This isoform is distinct from estrogen receptor beta that is encoded by a different gene and that also binds estrogen and estrogen-like mimics for transactivation. While both isoforms use estrogen as an agonist, they have different abilities to use estrogen-like mimics as agonists.
ER-alpha AOP	Adverse outcome pathway initiated by a ligand binding to the nuclear estrogen receptor alpha leading to downstream

	events that cause an adverse effect in an organism.
ERE	Estrogen Response Element
ES	Expert System
ESv1	Expert System version 1
ESv2	Expert System version 2
ESv3	Expert System version 3
EC	European Commission
EU	European Union
FQPA	Food Quality Protection Act
GA	Gene Activation, which in this context, refers to the liver slice vitellogenin gene activation and expression assay in male rainbow trout
HTP	High through-put
Log Kow	Octanol-water partition coefficient
MCF-7	Human breast cancer cell line
MIE	Molecular Initiating Event
MO	Mixed Organics
MP	Mixed Phenols
NF	Non-food
NVS	Novascreen Biosciences Corporation
OECD	Organization for Economic Cooperation and Development
P450	Cytochrome P450
PXR	Pregnane X Receptor (also termed SXR: Steroid and Xenobiotic Receptor)
QSAR	Quantitative Structure-Activity Relationship
RBA	Relative Binding Affinity
rtER	Rainbow Trout Estrogen Receptor

SDWA	Safe Drinking Water Act
SERM	Selective Estrogen Receptor Modulators
SMARTS	Smiles Arbitrary Target Specification
SMILES	Simplified Molecular Input Line Entry Specification
TD	Thyroid Disruptor
TDC	Thyroid Disrupting Compound
TG	OECD Test Guidelines
THR	Thyroid Hormone Receptor
Tox21	A chemical library that consists of HTP data generated by the National Institutes of Health (NIH), National Chemical Genomics Center (NCGC) as part of the Tox21 interagency partnership (initiated in 2008) between the EPA, the National Institutes of Health, and the U.S. Food and Drug Administration (added in 2010). See also pp. 49-50 in the white paper.
ToxCast	A chemical library that consists of HTP data generated by EPA and a series of contract laboratories. See also pp. 49-50 in the white paper.
TR	Thyroid Receptor
TrSet	Expert System Training Set
UnkBP	Unknown Binding Potential
Vtg	Vitellogenin

Question 1. Overall Prioritization Approach (Section 2.2; also see Section 9)

1.1 Please comment on appropriateness of the overall conceptual approach for prioritizing the EDSP chemical inventory (note: subsequent charge questions address issues associated with specific components of the prioritization approach). In your comments, please address:

The extent to which EPA's description of the process transparently captures and describes the key technical steps of the prioritization scheme and whether there are other scientific considerations the Agency *should incorporate into its EDSP prioritization scheme; the robustness of the scientific support for the overall approach and, in particular, the logical sequence of filters, and whether changes to the approach could set priorities more efficiently and effectively.*

Summary Response

The prioritization approach includes the consideration of: 1) physico-chemical properties, 2) structure- activity relationships and chemical interpolation/extrapolation within the context of the chemical category approach, 3) exposure information (for some chemicals) 4) high through-put (HTP) *in vitro* assays and computational methods and 5) the application of the AOP concept. The Panel considered the prioritization approach to be adequate to identify discrete, individual, low volatility, water-soluble chemicals with the potential to directly interact with the estrogen receptor alpha (ER-alpha) and activate ER- mediated transcription in the nucleus leading to downstream estrogenic effects. As such, the scheme is based on sound organizing principles and is appropriate for prioritizing the EDSP chemical inventory to identify those chemicals that may initiate these molecular events. Many of the initial screening techniques can be conducted relatively quickly and efficiently using non-animal approaches, such as *in vitro* assays and computational predictions, thereby reducing unnecessary animal testing.

In general, the Panel believes the key steps in the prioritization scheme were well organized and clearly described in the white paper, although there were several aspects that needed clarification or further development: 1) comparison and integration of different prioritization methods, 2) transparency in selection criteria for physico-chemical cut-off thresholds, 3) integration of relative environmental exposure factors (*e.g.*, chemical production in the U.S.) and relative potency as filters in earliest stages of prioritization, and 4) data analysis and integration steps.

The Panel believes that the inclusion of HTP data could significantly contribute to prioritization, but was divided on whether the EPA should screen all chemicals using HTP assays to identify potential ER active chemicals prior to ES analysis or whether the EPA should use HTP data as a means to expand effects-based categorization in the existing ES by combining data from HTP ER binding and estrogen response element (ERE)-dependent gene expression assays and rainbow trout assays. If the prioritization process were being designed today, initial steps would likely include validated HTP data, (however the HTP assays are not yet validated). The Panel discussed specific issues regarding expansion of the ES and the HTP assays in charge questions (CQs) 3-7.

The Panel was concerned that ~5,000 compounds in the chemical inventory were not available for testing and recommended that the Agency focus on obtaining samples of these compounds, if they are available in sufficient purity, since they are essential for prioritization for the estrogen disrupting pathway as well as other endocrine disrupting pathways (*i.e.*, thyroid, and androgen).

Although outside the specific Charge, the Panel spent considerable time during the meeting discussing the importance of integrating other MIEs and AOPs into the overall prioritization scheme. In the future, the Panel recommended that alternative MIEs and AOPs be considered in addition to the ER-alpha AOP as part of the larger overall prioritization process.

The Panel commented that EPA did not provide sufficient context of how pathways other than the ER-alpha AOP are being considered, whether potentially discriminating surrogates of exposure are considered and the envisaged objective of prioritization in the context of the entirety of the EDSP. The Panel could not fully answer some of the charge questions without further provision of the overall context. The Panel provided a sample illustration of a roadmap, Figure 1, for consideration by EPA of the type of diagram that might better contextualize their strategy as a basis both for their early consideration of aspects which might net considerable gains in efficiency, and for soliciting more effective and efficient input from the Panel. The Panel added that development of contextualized roadmaps facilitates both planning and communication.

Question 2. Physicochemical Properties (Section 2.2.2)

2.1 Please comment on the proposed exclusion criteria for each physicochemical property as appropriate for identifying chemicals that are untestable and/or unable to elicit systemic toxicity. Please include a discussion of the extent to which varying the range of exclusion criterion can influence confidence in identifying chemicals that are untestable and/or unable to elicit systemic effects (i.e., those substances that are unlikely to trigger the molecular initiating event). Please include in your comments: Whether there are additional chemical characteristics that should be considered (e.g., volatility), Whether there are additional data sources for each parameter identified (e.g., chemaxon for pKa values, EpiSuite for half-life values, etc.), and how to consider multiple data sources in a weight of evidence approach. Whether the exclusion criteria cut offs and the confidence that application of the exclusion criteria will reasonably identify those substances that are untestable and/or unable to trigger the molecular initiating event of the ER pathway. Also please discuss options for evaluating cases where uncertainty in a chemical's specific parameter range overlaps a p-chem filter cut off. Please consider in your comments the following exclusion criteria: pKa values <2 or >11.5; hydrolysis half-life of <40 days and charged species of 99% or greater at pH 7.

Summary Response

The Panel agreed that EPA's proposed exclusion criteria based on physico-chemical properties are logical, well informed and a positive addition to the EDSP screening and testing scheme. This comprehensive approach to using physico-chemical properties as filters to exclude chemicals from EDSP screening and testing is founded on strong scientific principles and the recommendations of the 1998 Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC). One caution noted by the Panel is that the cut-off values are not absolute or well justified, in some cases and use of technical experts to establish cut-off values is important. The Panel suggested *in vitro* testing of selected representative substances to confirm the cut-off values that are the basis for the exclusion criteria. The purpose of this testing would be to identify compounds on either side of the proposed cut-off to provide further evidence for the exclusion criterion (or criteria for multiple properties). This would, in effect, provide experimental evidence to support the cut-off value in the way that has been applied with log Kow. Use of physico-chemical filters to screen the EDSP chemical inventory is a novel addition to chemical risk assessment that could be widely applied to screen other large chemical inventories. The Panel agreed that other physico-chemical property filters (*e.g.*, volatility and reactivity) should be developed beyond those already available. The Panel also agreed that methods for estimating other physico-chemical properties (*e.g.*, pKa) are robust, readily available and good examples of the state-of-the-art.

Question 3. ER Expert System Development (Section 3)

3.1 Please comment on the approach used to build ESv2 through chemical testing and effect-based category and read across methods to provide a scientifically defensible approach to predict ER binding potential for a larger number of chemical groups, particularly with regard to defining new groups from within the Mixed Organics and Mixed Phenols. Please indicate any considerations unique to ESv2 that indicate the approach used to develop ESv1 (SAP, 2009a) needs to be modified.

Summary Response

The Panel congratulated EPA on putting together a well-designed, thoughtful and consistent approach in ES development. The Panel agreed that the approach to expand ESv1 to ESv2 to include non-food (NF) use inert substances is robust and scientifically defensible. The Panel considered the expansion of the chemical domain of ESv1 to ESv2 (to include NF inerts) to be based on sound science and organized in a logical fashion. The Panel noted that special rules and new structural sub-classes were needed in addition to the general rules to provide better coverage, and address the new unique structures (*e.g.*, fragrances) that the Agency intends to cover in the future.

The rationale for each chemical category and the structural boundaries needs to be clearly stated as a protocol, algorithm, or step-wise procedure. Additional groups or sub-groups within those currently proposed are likely to be identified and added to new versions of the ES as part of the iterative process. These will provide an opportunity to assess the node's predictive performance retrospectively, and quantitative performance testing should be conducted in the future. The Panel recommended random

testing of node category members and comparison of chemicals from within the ESv2 or ESv3 nodes to preliminary HTP data, where available.

Additional performance testing could include: analysis of "blind" out-group chemicals not explicitly included within the model, but within the model domain; cross-fold analysis for sensitivity, specificity, and positive and negative predictive value, and parallel analysis and testing of new chemicals.

Whilst chemical inventory coverage is currently presented in terms of percentages, use of 2-D or 3-D descriptors and plots would improve predictive accuracy of the chemical universe and the relative predictive accuracy of some chemical categories/nodes over others.

A weight of evidence approach and concordance assessment between different estrogenic AOP assays would increase the level of confidence in a prediction. Current and emerging computational approaches to assist in this include naïve Bayesian classifiers and statistical similarity methods. They could be incorporated as a follow-up screen for chemicals that ESv2 and ESv3 assign as unknown.

The Panel noted that the current software platform for the ES is developed from an initially proprietary system, that is now freely available, but also considered that there are also other chemoinfomatics approaches and analytical platforms that are not proprietary, may offer greater flexibility, and that may be suitable for the Agency to use.

The Panel further considered the gene transactivation and appropriate use of an ERE-Luc construct in the *in vitro* model, as different ER cell and promoter contexts can explain the dramatic functional differences among ER agonists, partial agonists and antagonists.

3.2 Building from the ESv1 training set, please comment on the level of scientific confidence that a chemical is unlikely to initiate the ER AOP, if the in vitro assay TrSet data shows no activity. Please comment on the extent to which the level of confidence may vary by chemical category.

Summary Response

In summary, the Panel agreed with the August 2009 SAP's evaluation of the ESv1 TrSet data. The ESv1 training set was reliable and robust for the applicability domain of food use inert ingredients and antimicrobial pesticides. Both *in vitro* tests, rainbow trout rtER binding and liver slice gene activation assays, are sensitive assays that when used in tandem increase the confidence in the predictions of ER ligand binding and activation of the ER-alpha binding AOP. Information derived for the ER-alpha may not necessarily apply to ER-beta, a different ER isotype encoded by a different gene, since the E/F domain of ER-beta that includes the ligand binding domain and the ligand-dependent transactivation domain shares only 53% identity with ER-alpha. Thus, the ESv1 TrSet data is specific for ER-alpha and is not universal for estrogen receptors. Human estrogen receptors alpha and beta greatly differ in their target genes, transcriptional potency and cofactor-binding capacity, and are differentially expressed in various tissues.

The Panel considered that the lack of data for some chemical categories leads to uncertainty, and that the choice of test chemical, relevant MIE, and assay chosen will affect the level of certainty. Further, the

size of the chemical category will affect the level of confidence and a chemical class might not fit into the classical ER-alpha binding AOP, but may still be relevant.

Thus far, the category approach does not provide equal emphasis on chemicals that bind to the B site in the ER alpha ligand- binding domain. The Panel also made the following suggestions to address the need for more data to assess whether or not an assay is a good prognosticator: that sources of collated ER modulator data from the literature be reviewed, and that chemicals acting as selective estrogen receptor modulators (SERMs) be used to explore DNA modified/knock out versions of the assays to determine prognostic quality.

Question 4. HTP ER Binding and ER Transactivation Data (Sections 4-6)

4.1 Given the importance of a well defined endpoint to determining if/how endpoint data are appropriately used, the panel is asked to comment on: Whether the assays are sufficiently described so that others can reconstruct the assay conditions and data analysis and to what extent additional information would be useful? Whether EPA has described sufficiently the important experimental conditions that affect the assay measurements, and whether EPA has sufficiently discussed their potential impact on assay results and interpretation of the results? Whether the chemical library (e.g., purity and analysis), chemical exposure (e.g., solvents used, chemical dilutions), and plates and plate layouts provides information necessary for data interpretation? The adequacy of test concentrations (maximum and minimum concentration tested) and cutoffs, and whether the concentration cutoffs affected some assays results more than others, especially in the context of false negatives.

Summary Response

The Panel agreed that the ER-alpha MIE, defined as the binding of a ligand to the ligand-binding domain of the estrogen receptor alpha followed by transactivation of the receptor, was well established in the literature. In addition, gene activation, both as used in the high throughput assays and in the liver slice assay, was a good addition to the model. For this specific MIE, data derived from the *in vitro* assays were acceptable and consistent with the mechanism identified for ER-alpha. The Panel agreed that in general the assays were well described including the detection system used for each, the endpoint tested, the potential for cytotoxicity and other factors that may interfere with the assay and the acceptance criteria.

There was considerable discussion on the choice of *in vitro* assays by the Agency, as being redundant to some extent and measuring the same molecular steps, binding to ER-alpha and transactivation of ER-alpha. The Panel suggested to the Agency that they consider other HTP assays that measure other molecular steps in the process as well. Many of these other assays are already being used by the Agency in the Tox21 and ToxCast programs.

Some panel members commented that additional information should be provided about the cellular background of the high throughput assays, as cellular background influences how ERs function and whether test chemicals can be adequately metabolized. Cell culturing media should also be well described and sometimes components in the media can bind chemicals making them less bioavailable.

The Panel cautioned against using impure chemicals that may not be fully characterized, since impurities can interfere with assay results in both a positive or negative way. Several examples of these types of interferences were provided. The Panel suggested the Agency adhere to performance characteristics described in the OECD *in vitro* test guidelines (TG) 455 and 457.

The Panel appreciated that the Agency is using some of the best characterized high throughput assays; however, there was considerable discussion about the proprietary nature of the assays and that they would not be testable by a third party. Scrambling of samples and keeping sample identity blinded to the testing laboratory are good methods to get some degree of assurance that the assays are performed correctly, but this could become an issue if results are not universally accepted.

In terms of the adequacy of the test concentrations, the Panel noted the use of only one test concentration at 25 µM in many of the HTP assays discussed in the white paper was considered insufficient, resulting in misclassification of low affinity ligands as ER-alpha negative. They suggested using full dose response curves. The rtER assays were considered a gold standard for this AOP and HTP assays should at least test doses that are relevant to those assays. The Panel also believes that the threshold calculation was too conservative and may increase the rate of false negatives. False negatives should be minimized.

4.2 With respect to data interpretation, please comment on the approach for defining an active chemical (i.e., to initiate the ER AOP) and an inactive chemical (i.e., unlikely to initiate the ER AOP) compound and whether the method of data interpretation is adequately described and if the rationale for the approach is sufficiently presented. Please include in your comments: The adequacy of the approach to generate the data plots; how and what parameters are calculated, etc. How the background data were used to establish the control level; whether the process for identifying assay interference is adequately described; adequacy of data normalization, outlier identification, curve fitting, background subtractions; and, all other data processing and calculation techniques, and appropriateness of the statistical analyses.

Summary Response

The Panel agreed that the Agency has a plan for identifying data interference; however, the plan simply eliminates data from consideration. They urged the Agency to develop methods to reduce data interference and determine whether these interferences affect false negatives.

Overall, the methods used by the Agency for data normalization, outlier identification, curve fitting, background subtractions, and statistical analyses meet currently accepted standards.

The Panel agreed that an RBA cut off value of 0.00001% is adequate and the rationale of the approach is sufficient to support the expansion of the ESv2 decision tree to ESv3. The Panel unanimously agreed that the approach was clear and concise. It would be better if there was background subtraction of the mean value for the dimethyl sulfoxide (DMSO) vehicle control from each sample, as noted in the OECD validated performance based ER-alpha stably transfected transactivation *in vitro* test guidelines (TG 455, 457).

4.3 Would you recommend other considerations or approaches to analyzing the HTP data?

Summary Response

The Panel agreed that the approaches used to evaluate HTP data are transparent, mechanistically probable, allow for a clear statement of the weight-of-evidence and allow for hypothesis-based, targeted testing. Panel members suggested additional chemical software packages that could help in data analysis and hierarchical cluster analysis software to help aggregate the results by chemical class. In addition, panelists identified agglomerative techniques that can pull in multiple types of analyses, including weighting schemes. The Panel suggested that the addition of a validation set would increase the confidence that the analysis was done correctly. The Panel also suggested that standard protocols for data analysis be developed, when the actual data are released to the public, for better transparency of the analytical process.

Question 5. Performance Evaluation of the HTP ER Binding and Transcriptional Activation Assays Against a Set of Reference Chemicals (Section 6)

5.1 Please comment on the selection of reference chemicals and whether they are sufficient to assess the performance of each HTP assay for ER agonists. How well do the reference chemicals represent the range of potencies needed to establish the reliability and relevance of these assays for use in EDSP chemical prioritization for the inert ingredients, fragrances, and SDWA chemicals?

Summary Response

The Panel discussed HTP assay performance assessment against the selected reference chemicals and recommended clarification of the classification performance of the assays to predict activity of the AOP in the documents and presentations. The Agency's evaluation appears to be mostly a statistical approach, which did not sufficiently describe the effects of thresholds, limit of detection or sensitivity of each HTP assay to detect the range of estrogen receptor activity, for the referenced chemicals.

The reference chemicals chosen represent a broad range of ER activity, classified from inactive to weakly active to strongly active. The reference chemicals were derived for *in vitro* OECD performance based test guideline for estrogen receptor stably transfected transfection assays (TG 457) and are based on validated data generated manually in the development of the test guideline. Although this is a good starting point, the Panel recommended the HTP ES use a very extensive and clearly explained reference chemical list with an equal balance of ER activity classes. This list can be obtained from the literature and the HTP assay data, and is necessary for accurate estimates of sensitivity and specificity, so that both false negatives can be minimized, and weakly active chemicals can be captured during the prioritization screening process.

The Panel also noted that the HTP assay performance was evaluated only on a binary scale of "active" or "inactive", which is more relevant for statistical analysis than for describing potency. It would be useful to evaluate assay performance and reproducibility against the positive control, 17 β-estradiol. Assays that do not reproducibly produce high quality data for estradiol should be removed and

assay sensitivity should then be evaluated separately for each assay, using weakly active to moderately active standards to establish the dynamic range of the assay.

5.2 Please comment on whether the Agency's *evaluation of the performance of the HTP ER binding and activation assays has considered and accurately assessed all relevant aspects of the assays. For example, have signal-to-noise ratios, background subtraction and interferences been adequately characterized when describing the strengths and limitations of the assay?*

Summary Response

The Panel indicated that while much careful analysis had been conducted, the results and explanations for poor data were not always adequate. Some clarifications were provided during the meeting.

5.3 Based on analyses in sections 5 and 6, please comment on the comparative performance of the 8 HTP assays for detecting ER reference agonists (e.g., do some assays perform differently, and if so to what extent?) Was sufficiently detailed information provided explaining the likely reasons for assay differences, when they are observed (e.g., cutoffs, assay interferences, background differences)?

Summary Response

The Panel observed that the EPA has provided extremely useful information on the comparative performance of the eight HTP assays for detecting ER-alpha reference agonists, but that it was preliminary and far from complete. The Agency noted this fact in the white paper.

The Agency's evaluation of differences between the eight HTP assays was based on the assumption that they provided redundant measures of the same biological activity, but the Panel stated that these different assays may represent complementary, but distinct components of the estrogenic MIE relevant to the ER alpha AOP. The Panel then described a variety of different ways in which an integrative and systems biology network approach to assessing assay performance could be undertaken. The Panel underlined the importance of understanding the cellular context of the different assays, when interpreting the results, and conducting comparative and combinatorial analyses. When assays have completely similar results, there may be a cost savings in removing the redundant assays.

Question 6. Analysis of ER Expert Training Set and HTP *in vitro* Assay Data (Section 7)

6.1 Please comment on the Agency's *assessment of how HTP in vitro assays performed with respect to expectations of performance for chemicals within chemical categories when compared with the in vitro training set data used to build the ESv1/2.*

Summary Response

The Panel noted that the Agency's assessment of HTP assay performance was clear and able to provide explanations for most instances where assays were different from ES TrSet. In general, the

largest cause of discordance between the two data sets was due to the high potency ER alpha ligand/activator bias of the HTP assay dataset vs. the rainbow trout assay datasets that were optimized for low potency chemicals (as noted in CQ 4.1.d). The Panel stated that it was unreasonable to expect that chemicals within categories should all behave similarly in the context of TrSet and HTP data. Comparison using the concentration-range mismatched datasets led to misclassification of chemicals in chemical classes where both low and high potency chemicals exist. This reinforces previous observations of the Panel that it is critical to compare assays across an equivalent concentration range.

Some panel members noted that the Agency could incorporate existing HTP data into the current scheme for identification of highly potent ER-alpha activators. Shifting the concentration maximum in the TrSet to match those in the HTP data set and comparing results could be used to validate HTP data. The ESv1/2/3(all versions) could then be used to prioritize remaining chemicals for further testing. It was noted that care must be taken in using mixed data (*e.g.*, TrSet and HTP set) since the Agency is looking at the ES as a QSAR and mixing endpoints is a violation of OECD principles. Data can be mixed if one considers the ES to be a series of structural alerts based on relative activity for both ER-binding and gene expression.

Given the capabilities of HTP assays in general, and access to fully validated HTP assays in the future, a portion of the Panel suggested that the Agency focus on screening as many compounds in the EDSP "universe" that are amenable to HTP testing. These panel members stated that having actual ER binding data (HTP data) rather than just predicted binding data (ES predictions) empirical information should be pursued.

6.2 Please comment on EPA's *approach to comparing the ES TrSet data with the HTP data and the interpretations regarding discordant results between the two assay approaches. Please recommend other considerations or approaches to conducting this comparative analysis?*

Summary Response

The Agency's approach seems reasonable given the inherent mismatch of the HTP assay vs. ES TrSet assay sensitivities due to differences in the experimental design (*i.e.*, use of different concentration ranges). The Panel recommends identifying those chemicals that were "inactive" in the ES TrSet data and "active" in the HTP assay data and were predominantly the result of assay interference (*e.g.*, fluorescence, cytotoxicity or flash sites). The Panel believes that the assays have different cellular contexts; the Agency should not expect complete agreement within receptor assays, transcriptional activation assays or between either, even if similar concentration ranges are examined. A more efficient analysis may be achieved using the larger HTP assay data set (*e.g.*, all 21 ER-alpha relevant HTP assays) to compare to TrSet to enable identification of assays/or combinations of assays that capture TrSet results.

6.3 Given that there may be varying occurrences of false negatives and false positives based on different sources of training set data, please include in your comments options for addressing these uncertainties for different categories in the context of building ESv3?

Summary Response

The Panel recommended examining a number of strategies for addressing uncertainty in different categories including exploring a range of statistical treatments, and categorization strategies in addition to use of screening data designed to detect a target threshold of activity (*e.g.*, low potency chemicals, chemical and performance standards). Data compatibility may also be enhanced through use of open data concepts to provide transparency of the raw and analyzed data sets to interested stakeholders.

A number of approaches were identified by the Panel to potentially improve comparisons and prioritization efforts. HTP assays could effectively be used to categorize chemicals in a tiered manner or through the use of a decision tree. Exposure potential could be incorporated into the ranking and prioritize chemicals for further testing at lower or higher concentration ranges. The use of additional HTP assays targeting different effects related to ER-alpha AOP key events would add more context and connectivity to the AOP, and add more utility to the ES. Since HTP screening of about 5,000 chemicals is relatively inexpensive and rapid, the HTP assays and effects based screening could be readily extended by testing all chemicals in the space that can be obtained or synthesized (*e.g.*, the 4,957 inerts and SDWA chemicals). Use of simple, robust HTP assays could potentially be as predictive as ESv3.

Question 7. Category-Based Approach and Strategic Testing to Expand the ES applicability domain with additional *in vitro* data (Section 8)

7.1 Please comment on the adequacy and efficiency of the category based approach to select chemicals for testing to expand the training sets in terms of covering the ~1700 chemicals not covered in the domain of ESv2.

Summary Response

The Panel noted that it is important to assess the tested or screened domain in relation to the possible structural space of the category and that uncertainty associated with classifying the universe of chemicals into smaller groups as a basis to select "representative" chemicals for a training set should be explicitly recognized. Characterization based on structure and log Kow is necessarily subjective, identifying groupings that may be unrelated to ER-alpha effects (*e.g.*, low affinity binders). Similarly, selection of representative chemicals from the upper or lower bounds of a class, *e.g.*, characterized on the basis of log Kow, may result in mixed responses or missing the effects of intermediate chemicals.

The Panel recommended that the Agency better articulate the strategy and rationale in defining the categories and selecting derivatives to assess them *a priori* before any additional development of categories and derivatives. These include criteria for both definition (extending beyond structural to a range of relevant physico-chemical properties) and appropriate bounding based on incorporation of a broader range of computational tools and HTP testing assays, for which *in vitro* assays could be verified to have sufficient sensitivity to measure the RBA <0.00001% of 17β-estradiol activity threshold.

Since category-based strategies may be inadequate for compounds remaining outside of clearly defined groups or HTP data, the Panel also suggested that EPA consider a broader range of options for

selecting and prioritizing derivatives for assessing new categories, especially as molecular complexity increases.

One option may be the incorporation of supporting "second pass" chemical informatics and systems pharmacology computational approaches, particularly for recalcitrant and diverse unknown binding potential (UnkBP) compounds. This is consistent with the 2009 SAP recommendations to consider the use of Simplified Molecular Input Line Entry Specification (SMILES) or SMILES Arbitrary Target Specification (SMARTS) strings or structural-similarity approaches.

Opportunistic HTP prescreening was also suggested as a potential option for consideration. Immediate expansion of HTP screening to address the entire chemical space prior to ES analysis could complement the categories training data set approach, unless precluded by constraints such as time and resources.

7.2 Based on the HTP assay performance (analysis of reference chemicals and comparative analysis with the ES TrSet), please comment on the adequacy of the HTP data for advancing the ER expert system's *rules to cover the additional groups of chemicals in ESv3.*

Summary Response

The HTP assay data may be useful to cover additional groups of chemicals in ESv3, when both conducted in a full dose-response format and validated**.** Currently, HTP assay sensitivity for detecting weakly active reference chemicals remains to be fully evaluated; as a result, it does not advance the ER-alpha ES predictions for weakly active chemicals. The HTP data should also be weighted, taking into account that where HTP data are included with results from the TrSet assays, statistical power grows but each observation becomes less certain. As a result, quantification of the retrospective predictive sensitivity and specificity of a particular HTP assay would be considered explicitly by applying a corresponding scalar weighting or confidence score to any new ES rule set that is derived in full or part from HTP data. The Panel suggested that broader integration of data from better targeted, purpose-designed, prognostic HTP testing would likely significantly enhance the ERES.

The Panel also suggested the following exploratory techniques to integrate ESv3 with HTP data (where sufficiently available): either as a *parallel input* alongside the ES, where both contribute to a single overarching prioritization score; or as an *overlay* to the ES itself. These approaches are not necessarily mutually exclusive. Thus, ES prediction and HTP screening could be run as *parallel inputs*, and their results combined into a single higher-confidence decision metric. Each model, be it a "physical" single HTP assay, or a "virtual" model such as the ESv3 or another computational method, should be assigned its own weighting or confidence score.

Currently, the ES's rule sets (*i.e.*, the "diamonds" in the ES schematic diagram) operate primarily over physical properties or structural patterns. However, the Panel's suggestion of immediately expanding at least some subset of HTP testing to the full chemical inventory would, if feasible, generate a new property that could also be exploited to build new ES rules: namely, the chemical's observed "HTP-fingerprint". Since HTP data may be available for only a subset of chemicals, HTP-fingerprints could be applied solely in a separate ES *overlay* or as a parallel "HTP-ESv3". This would preserve the

capability of ESv3 to make ER-binding predictions on chemicals for which no HTP data are yet present. Thus, rather than the proposed HTP-ESv3 rule sets being comprised of "contains a particular chemical group", the rule sets in HTP-ESv3 might be extended to "contains a particular chemical group AND is active in particular HTP assays". As a result, the Panel indicated that the HTP-ESv3 would extend chemical activity prediction beyond a traditional QSAR.

7.3 To the extent there are differences between the ES TrSet and HTP data in detecting the ability of chemicals to initiate the ER AOP, especially in the context of minimizing "false negatives" for low potency compounds, please comment on the strengths and limitations of combining data from the different assays to generate training set data for building the structure based rules within the ESv3. What are the strengths and limitations of using training set data from assays that have the same or similar degree of sensitivity?

Summary Response

The Panel cautioned against simply combining the HTP receptor binding and transcription activation data with the ES TrSet data. While each set of assays may be equally sensitive to an ER-alpha activating chemical, the current data can only be combined if the different concentration ranges tested by HTP assays (25 μM for initial screening and 100 μM for transcriptional activation assays) vs. the ES TrSet is taken into account. Potential options include re-evaluation and possible elimination of filtering thresholds as a basis to minimize false negatives for the HTP data, or augmentation of the HTP ER-alpha binding data with at least a single higher concentration for all negative chemicals. Alternatively, chemicals that are negative based on current HTP data could be tested in a second round of HTP assays at higher concentration range, similar to those administered in the trout liver assays.

Another approach is to use "reliable" HTP hits, where available, to push chemicals to the top of the "test using TrSet assays list." For example, if a compound was positive in any two of the HTP assays then it would have a high priority for testing via the full ES TrSet rainbow trout ER-alpha binding and Vtg gene expression liver slice assays (assays considered "gold standards"). A weakness is that this may increase the false positive rate for the initial prioritization step, but given the key programmatic goal of minimizing "false negatives" (white paper, p. 120, line 1847), it seems acceptable. Depending on testing resources and encountered false positive rates, the weakness might also be mitigated by adopting instead a "consensus-based" method, wherein a weighted average across all assay positives contributes to a single aggregate score by which further testing is prioritized. Considering the careful work that has been done to build the ESv1 and ESv2, it would seem sensible that ESv3 TrSet compounds be subject to confirmatory rtER and liver slice assay for all positives or borderline cases.

More assays covering other key events in the ER-alpha AOP would likely increase confidence that low affinity binders were not excluded. The inclusion of other assays or increased dose range of existing assays would help address low affinity binding chemicals. Additionally, assay specific weighting of data could also be used to inform prioritization and address potency issues.

Question 8. ***In Vitro*** **Testing and Computer Based Simulations: Addressing Active Metabolites (Note: This question is looking toward the future and there is no proposal presented in the white paper. Discussion of the issue of active metabolites in examples are presented in Section 7.4.3)**

8.1 Please provide any initial thoughts on how varying empirical and/or computational techniques could be employed to account for metabolites that may trigger the ER MIE.

Summary Response

The Panel acknowledged that a chemical directly eliciting a molecular event leading to downstream adverse effects (*e.g.*, ER-alpha binding triggering downstream toxic effects) is more easily defined than those triggered by active metabolites. Thus, if the Agency wants to increase the confidence that it has identified the complete universe of chemicals (including daughter products) requiring prioritization for further testing and evaluation, it will need to improve the capabilities of *in silico* metabolic simulations in addition to using HTP assays incorporating metabolism. The Panel noted that new capabilities in the next generation of metabolic simulators such as the ability to identify the most likely/probable metabolite(s) based on the structure of the parent compound and not just a list of possible metabolites as generated by current simulators (*e.g.*, the OECD Toolbox) could assist the EDSP, especially with respect to the ER-alpha binding ES. The Panel also suggested that the Agency develop a strategy for using predicted stable metabolites. Two possible strategies were mentioned: 1) include the predicted metabolites within the categories and 2) implement a cyclical process where metabolites would be predicted and then run through the ES. Each strategy has its advantages; however, based on current ES, the second strategy may be more practical. The Panel also offered some suggestions how potential HTP assays may assist with issues of metabolism. Finally, the Panel alerted the Agency to other metabolic applicability approaches being examined outside the Agency.

Question 9. Use of the AOP and Category-Based Testing Strategies for Other Endocrine Pathways (Section 9)

9.1 Please comment on whether the principles and concepts used to develop the ER expert system are generally applicable for any category-based prioritization system for other molecular initiating events (e.g., other AOPs for perturbing estrogen, androgen receptor, and thyroid hormone systems)?

Summary Response

The Panel supported the use of the proposed ER prioritization approach discussed during this SAP as a roadmap for other MIEs related to endocrine disruption. Specifically, the development process used for the ER-alpha prioritization model; the iterative steps of model-building, experimental testing, and model extension/refinement, is an excellent one. The Panel emphasized that one challenge for moving forward to other MIE/AOP based ESs is the choice of the MIE/AOP(s). To this end, the Panel noted that when the Agency develops additional ESs to model a specific MIE/AOP, the Agency needs to advance a transparent scientific rationale for why the particular MIE/AOP was selected as the basis for the pre-

screening and prioritization scheme. A second challenge stated by the Panel is the identification of true-positive chemicals for system development and validation.

The Panel felt that attention to issues which reduce uncertainty during the development and justification of any new ES will aid the Agency in their efforts to garner support for any new system. The Panel noted that issues which reduce uncertainty include: 1) transparency, mechanistic plausibility/probability, 2) weight-of-evidence, and, 3) ability to do hypothesis-based testing, especially with rapid and inexpensive methods.

The Panel stated that there were several important steps in development and acceptance of an AOP: 1) separate events that control toxic potency from events that report symptoms, 2) attain acceptance of the key events included in the AOP and ES development, 3) attain agreement on the method(s) used to assess these key events, and, 4) attain agreement on the *in vivo* data used as the standard for verifying methods/data for key events. Overall, there needs to be a transparent mechanistic understanding of the AOP.

Critical to the weight-of-evidence is assessing the: 1) key events that comprise the AOP, 2) relevance of each key event to the final adverse effect in question, 3) number of methods used to assess each key event, 4) relevance of each method to a given key event, and 5) reliability of the data measured or otherwise determined for each method.

9.2 What lessons have been learned from the development of the ER focused prioritization model that will facilitate more efficient and effective development of an effect-based prioritization model for the androgen pathway.

Summary Response

The Panel assumed the primary AR-pathway leading to endocrine disruption will be similar to the ER-alpha pathway, in that the seminal event will be receptor binding. If this assumption is true, the prioritization scheme developed for the ER-alpha pathway is directly applicable to the androgen (AR) pathway. The Panel noted that the "AOP for AR-mediated Fish Growth Impairment" proposed by Volz and co-workers (2011) may be a good starting point and alerted the Agency to other AR-related applicability being examined outside the Agency.

9.3 What will be the challenges in developing a prioritization using in vitro methods and computer based simulations for the thyroid hormone system?

Summary Response

The Panel noted that the thyroid hormone system is more complex (*i.e.*, multiple mechanisms and pathways) than either the ER- or AR-pathways. In addition, there is a general perception of having less data to construct an ES. Given the inherent complexity of the thyroid hormone system and lack of data, it will be a great challenge to substantiate selection of a single AOP to use in developing an ES to prioritize testing for alterations of the thyroid hormone system.

DETAILED RESPONSES TO CHARGE QUESTIONS

Question 1: Overall Conceptual Approach for Prioritization

1.1 Please comment on appropriateness of the overall conceptual approach for prioritizing the EDSP chemical inventory (note: subsequent charge questions address issues associated with specific components of the prioritization approach). In your comments, please address: The extent to which EPA's description of the process transparently captures and describes the key technical steps of the prioritization scheme and whether there are other scientific considerations the Agency *should incorporate into its EDSP prioritization scheme; The robustness of the scientific support for the overall approach and, in particular, the logical sequence of filters; and, Whether changes to the approach could set priorities more efficiently and effectively.*

Panel Response

General comments

The Panel offered a few general comments before responding to this charge question. The Panel commended the EPA for their efforts to develop a logical prioritization scheme for screening a massive number of chemicals. The Panel was pleased that EPA had incorporated most of the advice from the August 2009 SAP into the proposed scheme. The Panel thanked EPA for the opportunity to comment on scientific issues concerning the overall prioritization process for the EDSP chemical inventory while this effort is still under development and where recommendations from the Panel might be readily considered.

The Panel understands the fact that EPA is faced with the challenge of implementing the legislative mandate to test approximately 10,000 EDSP chemicals and the level of effort and resources that would be needed to do a thorough job of screening all of these chemicals. With this in mind, the Panel agreed that a prioritization process based on best scientific principles would make an enormous screening process manageable.

Key technical steps of the proposed prioritization scheme

The Panel indicated that the EPA EDSP chemical prioritization approach presented for review represents a limited, but critical, portion of the Estrogen Receptor (ER)-dependent Adverse Outcome Pathway (AOP). AOPs are conceptual frameworks used to organize mechanistic and/or predictive relationships between initial chemical-biological interactions (*i.e.*, molecular initiating event or MIE) that perturb cellular function sufficiently to elicit disruptions at higher levels of organization, culminating in an adverse phenotypic outcome relevant to risk assessment (*i.e.*, disease progression or organ dysfunction in humans; impacts on wildlife populations or critical ecosystem functions (Ankley *et al*. 2010). Here, the EPA EDSP chemical prioritization approach specifically focused on the MIE where a chemical binds to the alpha estrogen receptor (ER-alpha), thereby causing it to translocate into the cell nucleus and activate gene transcription. The Panel considered the prioritization approach to be adequate to identify discrete, individual, low volatility, water-soluble chemicals with the potential to directly

interact with ER-alpha and activate ER-mediated transcription (key MIEs of the ER-alpha AOP). As such, the scheme is based on sound organizing principles and is appropriate for prioritizing the EDSP chemical inventory to identify those chemicals that may initiate these molecular events. Many of the initial screening techniques can be conducted relatively quickly and efficiently using non-animal approaches, such as *in vitro* assays and computational predictions, thereby reducing unnecessary animal testing. The prioritization approach includes the consideration of: 1) physico-chemical properties, 2) structure- activity relationships and chemical interpolation/extrapolation within the context of the chemical category approach ES, 3) exposure information (for some chemicals), 4) high through-put (HTP) *in vitro* assays and computational methods, and 5) the application of the AOP concept.

The ER-alpha mediated fish reproductive impairment AOP outlined in Figure 3.1.1 (white paper) demonstrates the typical responses that might be observed at various levels of biological organization and the types of assays (*in vitro* or *in vivo*) from which observations along the pathway can be made using the rainbow trout as a model system. The MIE is triggered by a chemical binding to the rainbow trout ER-alpha (rtER), which is measured with *in vitro* binding assays and verified at a higher level of biological organization using activation of vitellogenin (Vtg) gene expression in liver slices from male rainbow trout. The EDSP has used these two measures of the MIE of the ER-alpha AOP to characterize model endocrine disrupting chemicals in order to define chemical groups within the knowledge-based (ES) to predict which chemicals have the potential to bind to the ER-alpha and initiate the AOP. This is a scientifically sound and transparent means of doing targeted testing that provides data that increases the weight-of-evidence supporting the ES and, in turn, increases confidence in the chemical categories and predictions based on them. This work, taken *in toto,* has moved the principles and concepts discussed in the August 2009 SAP (report found in Appendix C) to the proof-of-concept stages.

The Panel made several additional statements in support of the overall prioritization scheme.

1) The physico-chemical filters exclude thousands of chemicals (according to the white paper, approximately 5,000 of the 9,000 SDWA chemicals, inerts and fragrances) from Tier 1 testing because they are likely to be untestable since they are, for example, polymers, highly charged, and/or corrosive, when utilizing *in vitro* methods and systemically inactive (see discussion physico-chemical filters in the response to CQ 2). The application of the ER-alpha AOP leading to endocrine disruption is the logical organizing principle fundamental to the prioritization scheme. This is because the ER-alpha AOP relates the biological plausibility and adverse response to a meaningful MIE that is relevant to a particular toxicological mechanism and end point. The ER-alpha AOP will allow the assessment of internal consistency of the data as additional data are accumulated as a result of subsequent levels of screening. It also provides transparency and mechanistic justification for the key events that are either measured using *in vitro* assays for ER-alpha binding and gene activation, or predicted from *in silico* knowledge-based ESs.

2) The use of multiple types of assays for each key event compensates for some of the weaknesses of particular assays and increases the overall certainty as to whether or not a compound may be an endocrine disruptor and should lead to a higher priority for that chemical to be in Tier 1 testing.

3) The ES has been thoughtfully developed and well-communicated in the context of the organizing construct of ER-alpha AOP as a basis to "anchor" predictions or observations of MIEs and other key events to adverse outcomes traditionally considered in regulatory risk assessment. In essence, depending upon the proximity of these key events to the adverse outcome of interest, it also provides a basis for assessing confidence or uncertainty in the outcome of various approaches to prioritize chemicals for testing. The ES is able to tolerate false positives in order to reduce the number of chemicals prioritized for Tier 1 testing, but should minimize missing chemicals through false negatives. Although the ES may be able to tolerate false positives, a false negative could have a significant impact due to a delay in testing a true positive and the ability to act on the results. While the prioritization methodology presented at this SAP focused on the ER-alpha AOP, the overall concept and logic behind it should provide a generic approach that could be adapted to other endocrine disruptor pathways (see discussion in CQ 9) and other related estrogenic MIEs.

Transparency of the approach

In general, the Panel believes the key steps in the prioritization scheme were well organized and clearly described in the white paper, although there were several aspects that needed clarification and/or further development.

1) *Comparing different methods.* The Panel stated that more transparency is needed to describe how the results from each of the different components in the prioritization scheme are compared. Each method has different goals, strengths, and limitations.

2) *Transparency of criteria for selection of the 0.00001% cut point for active vs. inactive chemicals for initial prioritization*. The basis for selection of this value in prioritizing substances for additional testing is neither well specified nor justified and should reasonably include consideration of factors other than the limit of detection of the assay. These factors include a targeted degree of discrimination (i.e., what is considered a reasonable proportion of substances to capture for testing) and what the cut point represents in the context of potential for adverse effect. The overall approach was unclear on the acceptable rates of false negatives and false positives for correct identification of chemicals that activate the ER MIE in the initial phases of the prioritization, *i.e.*, what is the required degree of discrimination for judging the utility of the assays and the ES. Since the ES is designed to prioritize chemicals for testing, its ER-binding cut-off value for active vs. inactive is set low (0.00001%). This assures that having a false negative via the ER-binding pathways, especially with binding at the receptor A- and/or B-site is extremely remote. Concomitant to having an extremely low probability of a false negative is the higher probability of having a false positive. Therefore, the percent of chemicals screened as positive and testing as positive in Tier 1 assays may vary with relative binding affinity as well as with the weight of experimental evidence for the particular structural alert sequence triggered within the ES.

3) *Physico-chemical filters.* The Panel recommended that the rationale for the cut-off thresholds selected for the filters should be provided in more detail than is currently present in the white paper. While generally there is no right or wrong answer in relation to cut-off thresholds, there

needs to be transparency in regard to criteria for the selection of specific thresholds. This can relate to identifying a reasonable proportion of substances considered to be highest priorities, *i.e.*, a target threshold. Since the physico-chemical filters are designed to eliminate only the chemicals that are extremely unlikely to be testable because of deficient bioavailability (*e.g.*, excess volatility, inadequate water solubility), cut-off values for any physico-chemical filter should be set so they separate out only the most obvious outliers.

4) *Consideration of relative environmental exposure factors and relative potency as filters to prioritize chemicals.* The white paper should be written with more clarity and transparency with respect to how environmental exposure factors (contributing to bioavailability in the broadest sense) are considered in the prioritization process. The overall conceptual approach for prioritization could be improved by including relative environmental exposure factors and relative potency as early stage filters in the prioritization scheme. Consideration of simple exposure surrogates such as use, or which of these chemicals are measurable, or are likely to be measureable in environmental matrices (such as drinking water) or food sources are important factors that could be used to exclude chemicals from testing or allow them to be placed in lower priority categories. For example, if information about chemical production or use of certain listed chemicals in the U.S. provides no evidence that these chemicals are entering the environment or have the potential to enter the food chain, then the theoretical endocrine disruption would be a moot point from a practical standpoint. If production or other processes change in the future such that excluded chemicals become likely to enter the environment, then the compound's priority rating could be increased and screening could be conducted.

   The Panel recommended that the Agency have a "roadmap" on how various components of exposure and hazard will be integrated to prioritize chemicals for screening and assessment. A significant opportunity to meaningfully reduce the numbers of chemicals that need to be tested may be missed because simple elements of exposure are not being addressed up front in the prioritization process. The fact that a significant portion of the relevant substances with unknown binding potential fall into one use category indicates that environmental exposure factors would be beneficial to the prioritization process (see discussion in CQ 7).

   To a certain extent the water stability and other physico-chemical filters do address the potential for environmental exposure in terms of environmental stability (*e.g.*, the rate of hydrolysis in the environment), but not explicitly.

5) *Data analysis, integration steps, and ranking.* The Panel indicated that data analysis and integration steps are in flux and require greater explanation and analysis. For example, how will the HTP assay data be integrated into the existing ES, which is based on empirical data from rainbow trout? Mixing and matching data is inconsistent with the OECD validation principles for QSAR models. Many of the initial screening techniques can be conducted relatively quickly and efficiently, for example computationally (on the basis of read across and structure-activity relationships), and therefore will not invest in animal lives unnecessarily; such initial screening information will give insights into whether the ER-alpha pathway is likely to be activated and, if

so, whether the concentrations at which the compound is active, are of concern. The current prioritization scheme appears to classify chemicals into broad categories such as high, low or non-binding/activating chemicals, although this is not explained in depth in the white paper.

Ranking chemicals based on potencies across all assays/activities, similar to a weight of evidence approach, rather than just individual assays, would also improve the prioritization approach. Such an approach could be achieved by giving a numerical ranking to potency in each of the assays tested and then sum the values across assays for that chemical. This could be used as a cumulative score for potency in ranking chemicals.

Scientific support for the overall prioritization approach

Expert System. The Panel identified issues related to expanding the tested chemical space to cover more structurally diverse chemical inventories that have yet to be resolved. The Panel encouraged the Agency to move forward quickly to resolve these issues for the ER-alpha AOP so these schemes can be used for other MIE/AOP-based ESs. More specifically, the panel members indicated that the ES version 2 (ESv2) is expected to have high accuracy across the chemical space on which it has been trained, but lower accuracy on "novel" chemistry in the inventories to be covered in ES version 3 (ESv3). The use of HTP data from ToxCast and Tox21 programs to inform novel chemicals spaces followed by strategic experimental testing is a viable approach, but the results presented to the SAP were preliminary and raised many issues. See additional discussion of the knowledge-based ES in the Panel's responses to CQs 3-7.

HTP assay data. The test concentration range used in the HTP assays had a lower maximum concentration than the concentration range used with the rainbow trout assays. As a result, the identification of ER-alpha active chemicals by HTP and rainbow trout assays are not directly comparable without adjusting for concentration ranges. The difference in concentration ranges means that the HTP assays (as currently conducted) will result in too many false negatives and be less predictive of the ER-alpha activation potential of chemicals in comparison to the rainbow trout assays. The lack of comparable concentration ranges used in the rainbow trout assays and the HTP assays reduces the scientific support for the overall approach.

Availability of chemicals. The Panel was concerned that ~5,000 compounds in the library were not available for testing. The Panel recommended that the Agency put significant priority on obtaining samples of these compounds if they are available in sufficient purity since they are essential for this initial prioritization effort as well as extension of this strategy to other endocrine disrupting chemicals. See additional discussion on HTP assays in the Panel's responses to CQs 4- 7.

Higher priority for HTP assays

The Panel believes that the inclusion of HTP data could significantly contribute to prioritization, but was divided on whether the EPA should screen all chemicals using HTP assays to identify potential ER active chemicals prior to ES analysis or whether the EPA should use HTP data as a means to expand effects-based categorization in the existing ES by combining data from HTP ER binding and estrogen response element (ERE)-dependent gene expression assays and rainbow trout assays. If the prioritization

process were being designed today, the process could be based on validated HTP data. While much investment has been made into the current ES, which has significant value, the Agency needs to reconcile the computation model (*i.e.*, ES) of whether or not a chemical can bind to ER-alpha and activate ER-alpha-dependent transcription vis-à-vis empirical HTP testing of ER-alpha binding and ER-alpha-dependent gene activation. Since it is feasible to have access to actual ER binding data rather than just predicted binding, obtaining this empirical information should be pursued.

Some panel members stressed that an effective HTP approach would perhaps employ both biochemical (*e.g.*, ER-alpha binding and gene activation assays) and a limited number of cell-based assays that would allow identification of active compounds with a low probability of false negatives.

Importance of considering other possible ligand-bound ER MIEs that can lead to downstream effects

The Panel appreciates the rationale behind EPA's decision to focus its initial prioritization approach on the ER-alpha AOP, but stressed there are other functionally significant ligand-bound ER interactions that can lead to downstream estrogenic effects. In addition to addressing the specific Charge, the Panel spent considerable time during the meeting discussing the importance of integrating other MIEs and AOPs into the overall prioritization scheme. Given the focus of the meeting on one specific, well-known MIE for one AOP, the Panel agreed that the ER-alpha AOP is a good initial focal point to examine the prioritization approach. In the future, the Panel recommends that alternative MIEs and AOPs be considered in addition to the ER-alpha AOP as part of the larger overall prioritization process.

Importance of a roadmap

The Panel commented that EPA did not provide sufficient context of how pathways other than the ER-alpha AOP are being considered, whether potentially discriminating surrogates of exposure are considered and the envisaged objective of prioritization in the context of the entirety of the EDSP. As a result, the Panel could not fully answer some of the charge questions without further provision of the overall context.

The Panel recommended that a roadmap be constructed that would illustrate all major assessment steps used to evaluate the endocrine disruption potential of a chemical in the proposed EDSP universe and how these steps fit with envisaged yet to be developed components (*e.g.*, prioritization via other pathways). The Panel provided a sample illustration of a roadmap, **Figure 1**, for consideration by EPA of the type of diagram that might better contextualize their strategy as a basis both for their early consideration of aspects which might net considerable gains in efficiency, and for soliciting more effective and efficient input from the Panel. The development of a contextualized roadmap facilitates both planning and communication.

As shown in **Figure 1** (see also description in the white paper, pp. 2-7; EPA overview presentation), each chemical (of the approximately 10,000) in the EDSP universe is separated into one of two endocrine disruptor assessment pathways: 1) pesticide active ingredients associated with a registration review timeline (FIFRA pathway) and 2) chemicals subject to the Safe Drinking Water Act (SDWA), pesticide inerts and fragrances not associated with a registration review timeline (SDWA pathway).

Pesticides in the FIFRA pathway are prioritized according to the registration review schedule. Along this pathway, there might be consideration of read across and chemical categorization methods before Tier 1 testing. Chemicals in the SDWA pathway (focus of this SAP meeting) are filtered through a series of environmental exposure and physico-chemical exclusion filters; remaining chemicals are prioritized using a combination of HTP assays and *in silico* methods before Tier 1 testing. Following Tier 1 testing, a weight-of-evidence approach is used to determine whether a chemical should be further assessed using targeted Tier 2 tests. The final assessment of the endocrine disruptor potential of a chemical occurs after the completion of the Tier 2 tests. The Panel also recommended adding a filtering step using relevant exposure environmental factors early in the prioritization process for chemicals in the SDWA pathway and clearly delineating factors that relate to potential for limited exposure rather than hazard such as half-life in the environment.

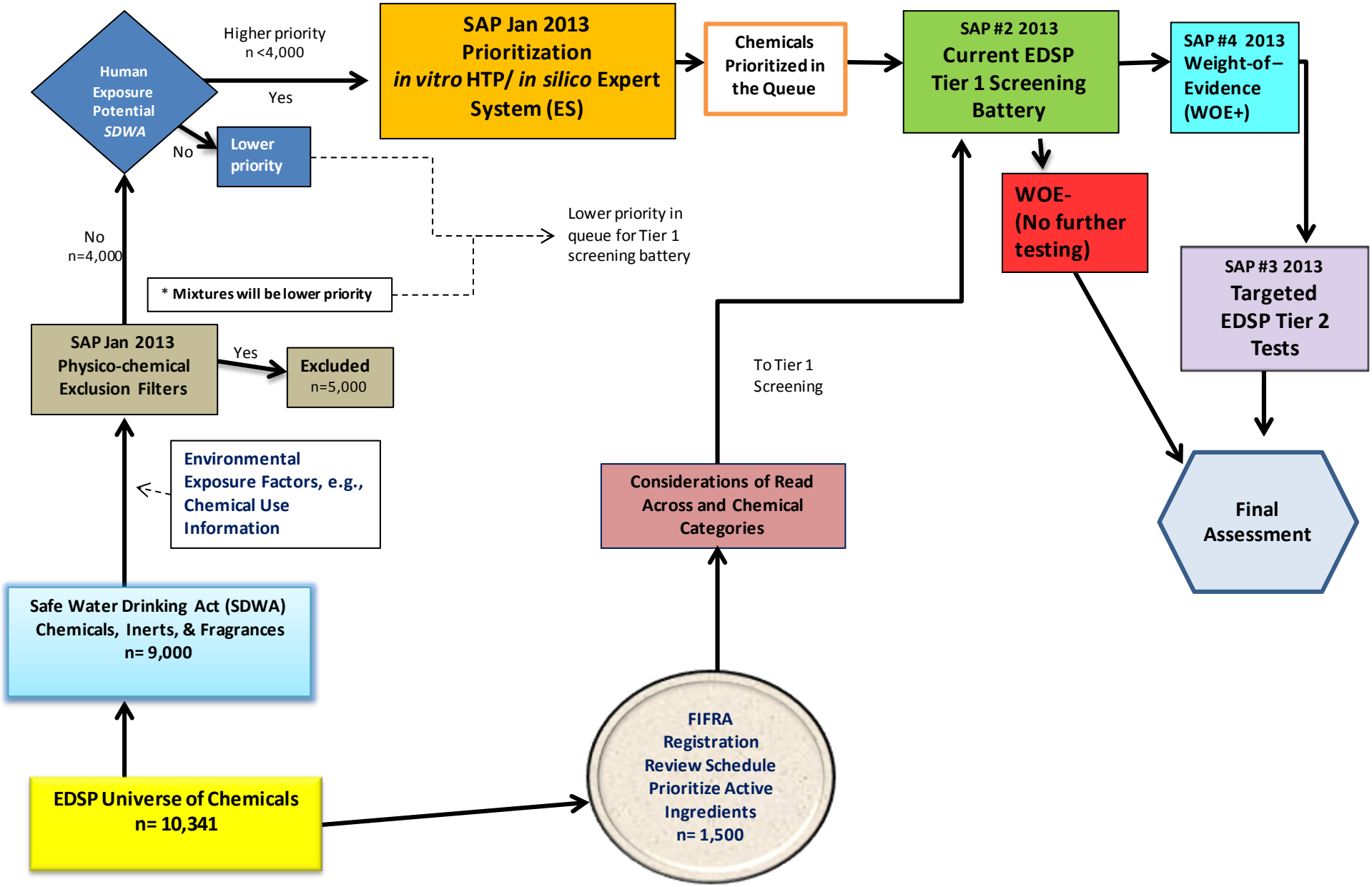

Figure 1. Sample Illustration of EDSP Roadmap. [Adapted from EPA presentation, slides 16 and 25.]

Not every chemical necessarily needs to undergo EDSP screening and testing. Decisions of whether to screen a chemical for endocrine interaction are made in part based on information about a chemical's inherent properties (e.g., pKa values, molecular weight, reactivity/stability, corrosivity, known functional groups, and charged species). EPA may use this information to identify untestable chemicals or those that are unlikely to pose systemic effects (e.g., endocrine perturbation). Application of filters based on physiochemical properties enables EPA to reduce the chemical universe that would need further evaluation (as illustrated in the white paper, from ~10,000 chemicals to ~5000 or less).

2.1 Please comment on the proposed exclusion criteria for each physicochemical property as appropriate for identifying chemicals that are untestable and/or unable to elicit systemic toxicity. Please include a discussion of the extent to which varying the range of exclusion criterion can influence confidence in identifying chemicals that are untestable and/or unable to elicit systemic effects (i.e., those substances that are unlikely to trigger the molecular initiating event).Please include in your comments:

- *2.1.a. Whether there are additional chemical characteristics that should be considered (e.g., volatility),*
- *2.1.b. Whether there are additional data sources for each parameter identified (e.g., chemaxon for pKa values, EpiSuite for half-life values, etc.), and how to consider multiple data sources in a weight of evidence approach.*
- *2.1.c. Whether the exclusion criteria cut offs and the confidence that application of the exclusion criteria will reasonably identify those substances that are untestable and/or unable to trigger the molecular initiating event of the ER pathway. Also please discuss options for evaluating cases where uncertainty in a chemical's specific parameter range overlaps a p-chem filter cut off. Please consider in your comments the following exclusion criteria: pKa values <2 or >11.5; hydrolysis half-life of <40 days and charged species of 99% or greater at pH 7.*

Panel Response

Note: The Panel answered this question with reference to all of Section 2.2 (not just Section 2.2.2) in the white paper. There are many subparts to Question 2.1 and each is answered sequentially, *i.e.*, Question 2.1.a., 2.1.b, and 2.1.c. There is no Question 2.2.

As the Agency examines the EDSP universe of chemicals considered for Tier 1 testing, it will be important to remove substances from consideration which, for reasons of physico-chemical properties, fall outside the domain of reliable testing. The Panel agreed that EPA's proposed exclusion criteria based on physico-chemical properties are logical, well informed and a positive addition to the EDSP screening and testing scheme. This comprehensive approach to using physico-chemical properties as

filters to exclude chemicals from EDSP screening and testing is founded on strong scientific principles and the recommendations of the EPA advisory committee, Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) (EDSTAC 1998). One caution noted by the Panel is that the cut-off values are not absolute and use of technical experts to establish cut-off values is important. The Panel suggested *in vitro* testing of selected representative substances to confirm the cut-off values that are the basis for the exclusion criteria. The purpose of this testing would be to identify compounds on either side of the proposed cut-off to provide further evidence for the exclusion criterion (or criteria for multiple properties). This would, in effect, provide experimental evidence to support the cut-off value in the way that has been applied with log Kow. Use of physico-chemical filters to screen the EDSP chemical inventory is a novel addition to chemical risk assessment that could be widely applied to screen other large chemical inventories. The Panel agreed that other physico-chemical property filters should be developed beyond those already available. The Panel also agreed that methods for estimating other physico-chemical properties (*e.g.*, pKa) are robust, readily available and good examples of the state-of-the-art.

The EPA proposed that chemicals which are too acidic or basic, too reactive or unstable in water, charged species, or are too corrosive would be excluded from any further testing because they fall outside the domain of reliable testing. EPA's analysis summarized in **Table 1** (reproduced from Table 2.2.2.1, white paper, p. 13) demonstrates how such exclusion criteria could be a very powerful and rational means of reducing the number of EDSP chemicals that would need to go through the Tier 1 screening battery of tests. As noted in the white paper (Section 2.2.2, p. 7; Table 3.2c.1 pp.43-44), the number of chemicals would be reduced from ~10,000 to ~5000 or less. The Panel noted that while each physico-chemical property can be used as a unique filter, multiple physico-chemical filters may yield overlapping results that can be compared or correlated. For example, the four physico-chemical properties used as exclusion criteria (**Table 1**) yield overlapping results for some chemicals. Greater certainty that a chemical should be excluded from screening and testing may be associated with compounds outside of more than one of the exclusion criteria.

Table 1. Summary of EPA's Analysis of Physico-Chemical Properties Exclusion Criteria (reproduced from Table 2.2.2.1, white paper)

Criterion	Parameters	Data Sources	Boundaries	# Chemicals Excluded
Too acidic or basic	pK_a of acids or conjugate acids of bases	Chemaxon's "Chemicalize" predictive software	pK_a < 2 or pK_a of conjugate acid >11.5	Approx. 700 Excluded
Too reactive or unstable in water	Hydrolysis half-life	Epi Suite's HYDROWIN predictive module	Eligible for exclusion if hydrolysis half-life < 40 days	Approx. 100 Excluded
Charged at pH 7	pK_a based net negative or positive charge at pH7	Chemaxon's "Chemicalize" predictive software	Lowest acid pK_a < 5 or highest conjugate acid pK_a >9	Approx. 700 Excluded
Too corrosive	Member of a chemical class with reactive functional groups	DOT8 list; MSDS indicates corrosive to skin, eyes	Solid, water insoluble chemicals not eligible	Approx. 200 Excluded

The Panel stated that there are several general reasons why certain compounds with inherent physico-chemical properties that are outside particular ranges will not be active as endocrine disruptors or be suitable for *in vitro* testing. These reasons include:

1) *Some compounds are not biological availability to cause effects in whole organisms, e.g., not susceptible to pass transport or absorption due to physico-chemical properties.* Compounds that are not susceptible to passive transport across a concentration gradient or absorption due adverse physico-chemical properties will not be bioavailable. These include molecular size (compounds that are too large are unable to migrate through membranes) or bulky, *i.e.*, the molecular dimensions in terms of breadth and width are too great). Compounds lacking in appropriate aqueous or lipid solubility will be unable to be absorbed.

   Chemicals may be inactive because they are rapidly metabolized or excreted. A pharmaceutical substance with an Absorption, Distribution, Metabolism, and Excretion (ADME profile that is not conducive to good bioavailability is unlikely to be active following oral dose. Some of the properties relating to poor absorption (and hence bioavailability) are captured by the exclusion criteria proposed*, e.g.*, molecular size and ionization potential. These factors have been used in the pharmaceutical industry particularly following oral dosing to determine the bioavailability of a drug.

For example, the Lipinski "rules of five" (Lipinski *et al*., 1997) are used to screen for compounds that are not likely to have significant activity following oral exposure. The Lipinski rules[1] identify compounds with poor solubility following oral exposure based on molecular weight, log Kow and counts of hydrogen bond donors and acceptors. Compounds with poor solubility are not likely to have significant biological activity following oral exposure. Rules of thumb for other uptake parameters are also available and could be developed further (Madden 2010a, b).

2) *Some compounds will be degraded rapidly in the environment and not be available as an exposure source.* Inclusion of criteria such as half-life is indicative of a chemical's potential to cause exposures in the environment. Chemicals with very low half-life values (the "cut-off" value would have to be determined) would be put in a lower priority for testing or perhaps, excluded from testing. This is one component of consideration of simple surrogates of limited exposure, which can likely be further developed to include additional parameters such as scale and type of use, and availability (see CQ 1. Comments on integration of exposure potential into early stages of prioritization process).

3) *Each HTP in vitro assay has its own physico-chemical "cut-offs", e.g., solubility, which define the applicability domain.* All experimental assays have limits, in terms of physico-chemical properties, in which they can be used reliably. This is as relevant to *in vitro* assays as it is for *in vivo* assays. With regard to *in vitro* assays a good example of a physico-chemical property associated with a cut-off is solubility, *i.e.*, poorly soluble compounds may not be able to be tested in some *in vitro* assays. Therefore, reliable data will not be obtained for compounds are outside of the solubility limits or cut-off of a particular assay. However, such cut-offs must be treated with caution, simply because a compound may not be testable *in vitro*, it may still elicit an *in vivo* response

Exclusion of polymers, mixtures, and undefined substances

The Agency proposed that mixtures, polymers and undefined substances are excluded or placed in the lower-priority category based on the following criteria (see p. 7, white paper):

> *• Mixtures of either defined or undefined chemicals. Mixtures of defined discrete chemical structures will be screened after the single chemicals have been addressed;*
>
> *• Polymers with an average molecular weight >1,000 Daltons based on its lack of ability to interact and bind to the ER, initiate the ER pathway and initiate systemic effects are excluded; and,*
>
> *• Undefined substances (i.e., not having a defined chemical structure) are excluded.*

[1] Note: Lipinski's rules do not hold for many steroid receptor ligands due to the inherent hydrophobic nature of this ligand class. There are also other exceptions.

1) Exclusion of polymers. The Panel agreed that the exclusion of polymers from EDSP testing is justified for the most part. Such compounds are typically high molecular weight and consequently are likely to have limited bioavailability. The assumption that they will not trigger the ER-alpha MIE is reasonable. However, the Panel also noted that proteins and peptides, which are polymers, are able to influence estrogenic actions. Thus, the recommendation was made to more rigorously define the term "polymer" for the purposes of categorization. The exclusion of polymers should be reconsidered when prioritizing for other MIEs.

2) Exclusion of mixtures. The Panel agreed that mixtures could be placed as a lower priority in the EDSP until single defined chemicals are prioritized. The Panel agreed that doing so after single substances have been considered is a pragmatic solution. Mixture interactions can be unpredictable due to the occurrence of additivity, antagonism, and synergism. The activity of single chemicals is likely to be simpler than mixtures and can be used to estimate additive effects based on toxic equivalent doses that may occur in mixtures. Therefore, the scheme to focus on prioritizing single chemicals first is the most efficient. The Panel noted that some mixtures may contain a single chemical that might be of concern with the rest being inert. The Agency could consider testing these formulations earlier in the process, *e.g.*, before testing mixtures of multiple chemicals of concern.

3) Undefined substances (e.g., natural products or those from a botanical or other origin). The Panel agreed with EPA that undefined substances should not be excluded based on the proposed filtering criteria and should be considered as a lower priority category.

4) General definitions. In general, clear and better definitions are needed for many terms used in the white paper, *e.g.*, polymers (as stated above), charged species and any reference to specific chemical domains.

5) Specific parameter ranges. The Panel commented that the term 'specific parameter range" was interpreted to mean (in the context of the White Paper and this report) the spread of a predicted physico-chemical parameter in the universe of chemicals being considered, which would cross the cut-off value.

2.1.a Please include in your comments: Whether there are additional chemical characteristics that should be considered (e.g., volatility).

The Panel indicated that volatility is a logical extension of physico-chemical characteristics included in the proposed exclusion criteria. *In vitro* testing of highly volatile substances often gives results that are less active than hypothesized because concentrations that need to elicit the measured response must compensate for abiotic loss. *In vitro* tests may need to be specially designed to accommodate volatile chemicals. Additionally, Henry's Law constant and/or vapor pressure may be good physico-chemical surrogates for volatility.

The Panel provided other physico-chemical characteristics that could be used as exclusion filters (some of these may be a subset of the proposed exclusion criteria):

1) *Poor in vivo bioavailability*. Indicators of rapid metabolism or clearance *in vivo* might be used to characterize poor bioavailability. The Panel noted that some of these properties were covered by the proposed exclusion criteria. See also earlier discussion on the Lipinski rules.

2) *Stability with particular reference to the applicability of HTP and in vitro assays*. There are other measures of stability (or lack of it), for instance, oxidation in air if this is a significant route of exposure.

3) *(Quantitative) In Vitro-In Vivo Extrapolation (QIVIVE).* There is considerable interest in extrapolating the results from *in vitro* tests to *in vivo* situations, *i.e.*, QIVIVE modeling. These models in themselves may not provide cut-off criteria, but their assessment of other relevant factors related to exposure might be useful. For example, this approach is attempting to identify and model the properties of chemicals that affect *in vitro* concentration such as volatility (mentioned above) and sorption to cells (Blaauboer 2010). This information, *i.e.*, the determination of the factors that affect *in vitro* testing which could be post-rationalized into cut-offs, may provide some further input and insight into the development of physico-chemical filters.

4) Electro(nucleo)philic reactivity, especially soft electrophilic reactivity and redox cyclers. Other chemical characteristics that could be used in excluding chemicals from screening and testing include electro(nucleo)philic reactivity, especially soft electrophilic reactivity as related to covalent protein binding. A large number of chemical sub-structures and fragments have been identified which could be used to screen for highly reactive chemicals (Enoch *et al*. 2011; Schwöbel *et al.* 2011). Similarly, chemicals that act as redox cyclers may be excluded from screening.

2.1.b. Whether there are additional data sources for each parameter identified (e.g., chemaxon for pKa values, EpiSuite for half-life values, etc.), and how to consider multiple data sources in a weight of evidence approach.

The Panel indicated that there were additional data sources for each parameter identified (*e.g.*, Chemaxon for pka values, EpiSuite for half-life values, *etc*.). There are other relevant criteria for corrosivity (Section 2.2.3.4, white paper, p. 12-13). Highly corrosive chemicals are unlikely to be testable *in vitro* (for practical reasons) or active *in vivo*. The list of structural alerts for corrosivity is comprehensive (and may overlap with alerts for ionization and possibly electrophilic reactivity). These alerts could be defined in detail to ensure consistent use. For instance, they can be defined "on paper" or in an electronic format as SMARTS strings or similar methodologies such as the definition of alerts in the OECD QSAR Toolbox (www.qsartoolbox.org).

There is increasing interest in the use of using multiple data sources in a weight of evidence approach (Ellison *et al.* 2010). There are also many possible sources of predicted values for the descriptors considered. However, there should be some rationale to the selection of predictive methods. For instance, in such a scheme it may be preferable to have freely available methods, those that are transparent, well documented with definable and defined applicability domains, and are demonstrably robust *etc*. Further methods to calculate pKa, degradation, *etc.*are available (see the recent review by Dearden *et al.* (2013). Some potentially useful prediction methods are summarized in **Table 2.**

Table 2. A summary of methods to predict physico-chemical properties (Adapted from Dearden *et al*. 2013).

Software	Vapor pressure	Water solubility	log K_{ow}	log D_{ow}	Flash point	Flammability limits	Self-ignition temperature	Dissociation constant	Viscosity	Henry's law constant
Absolv	✓		✓							✓
ACD/PhysChem Suite	✓	✓	✓	✓	✓			✓		
ADMET Predictor		✓	✓	✓				✓		
ADMEWorks Predictor			✓							
ChemAxon			✓	✓				✓		
ChemOffice			✓							✓
ChemProp	✓	✓	✓						✓	✓
ChemSilico		✓	✓	✓						
CHETAH						✓				
ClogP			✓							
Episuite	✓	✓	✓							✓
KlogP			✓							
MOE			✓							
Molecular Discovery		✓						✓		
Molecular Modeling Pro	✓	✓	✓						✓	
Molinspiration			✓							
MOLPRO		✓	✓	✓				✓		
OCHEM			✓					✓		

Table 2, continued **Software**	Vapor pressure	Water solubility	log K_{ow}	log D_{ow}	Flash point	Flammability limits	Self-ignition temperature	Dissociation constant	Viscosity	Henry's law constant
PALLAS			✓	✓				✓		
PhysProps	✓									
Pipeline Pilot		✓	✓	✓				✓		
PredictionBase		✓								
PREDICTPlus	✓								✓	
ProChemist		✓	✓	✓				✓		
ProPred	✓	✓	✓		✓		✓	✓	✓	✓
Schrödinger		✓	✓					✓		✓
SPARC	✓	✓	✓	✓				✓		✓
StarDrop[a]		✓	✓	✓						
TerraQSAR-LOGP			✓							
T.E.S.T.	✓	✓			✓				✓	
TOPKAT			✓							
VCCLAB		✓	✓	✓				✓		

[a]StarDrop offers log D at pH 7.4 only.

[1]Availability of software: P = purchase, C = available to consortium members, FD = free to download, and FO = free online.

MOLPRO: P, www.chemdbsoft.com;
OCHEM: FO, www.ochem.eu;
PALLAS: P, www.compudrug.com;
PhysProps: P, www.gpengineeringsoft.com;
Pipeline Pilot: P, www.accelrys.com;
PredictionBase: P, www.idbs.com;
PREDICTPlus: P, www.mwsoftware.com/dragon;
ProChemist: P, www.pro.chemist.online.fr;
ProPred: C, www.capec.kt.dtu.dk;
Schrödinger: P, www.schrodinger.com;
SPARC: FO, ibmlc2.chem.uga.edu/sparc/
SPARC: FO, ibmlc2.chem.uga.edu/sparc/
StarDrop: P, www.optibrium.com;
TerraQSAR-LOGP: P, www.terrabase-inc.com;
T.E.S.T.: melting point and vapor pressure prediction not available in version 4.0, but will be available in v. 4.1; FD, www.epa.gov/nrmrl/std/cppb/qsar/#TEST;
TOPKAT: P, www.accelrys.com;
VCCLAB: FO, www.vcclab.org

Multiple predictive algorithms and weight of evidence. There are a number of issues to consider when different predictive algorithms are used to calculate values. A strategy should be adopted to use these data – simply calculating more values does not necessarily increase accuracy or the quality of the output. Multiple data sources often use different algorithms based on the same experimental data. Since the models are often fitting exercises, the errors in their predictions vary with chemical class, *etc*. Using the worst case scenario to set the cut offs reduces the probability of excluding an active chemical. However, at the same time, it will increase the probability of evaluating a chemical that could be excluded (see below). The Panel agreed that a weight of evidence approach is appropriate should there be multiple predictive algorithms used to estimate chemical-specific values. This may be based on the relative performances of the models and may, for instance, be dictated by the sensitivity and specificity of individual models.

2.1.c. Whether the exclusion criteria cut offs and the confidence that application of the exclusion criteria will reasonably identify those substances that are untestable and/or unable to trigger the molecular initiating event of the ER pathway. Also please discuss options for evaluating cases where uncertainty in a chemical's *specific parameter range overlaps a p-chem filter cut off. Please consider in your comments the following exclusion criteria: pKa values <2 or >11.5, hydrolysis half-life of <40 days and charged species of 99% or greater at pH 7.*

1) *pKa values <2 or >11.5.* The Panel agreed that compounds with pKa values of <2 and >11.5 are highly unlikely to trigger the MIE of the ER-mediated AOP. There is no evidence that such compounds have been associated with this effect. This is supported by the assumption that such compounds are unlikely to be bioavailable (due to ionization). Lastly, highly acidic and basic compounds are usually intrinsically corrosive and are unsuitable for testing *in vitro* or HTP.

   The Panel stated it was possible to use the physico-chemical properties exclusion criteria twice; first to eliminate the obvious un-testable substances and second, after the compound has been placed into an appropriate chemical category. For example, all chemicals with pKa values < 2 or > 11.5 could be eliminated early in the scheme and then, if the chemical of interest was a phenol, a refined set of pKa-based criteria could be used to further prioritize this chemical for testing.

2) *Hydrolysis half life of <40 days.* The Panel agreed that highly degraded compounds are unlikely to cause an *in vivo* effect following exposure through environmental waters. This is due to a rapid decrease in concentration, or even elimination. In the white paper, EPA provided a half-life of 40 days as the filter. The Panel indicated that additional justification is needed to support this cut-off value. For example, the greater than 40 day half-life seems long, especially if this could be from multiple exposures to lipophilic compounds. This is despite EPA's presentation suggesting that there was little difference between a half-life of less than 1 day and that of greater than 40 days. The Panel believes that public perception would be to consider the difference to be greater than it is. The greater than 40 days half-life cut-off may (in certain circumstances) be appropriate for single point sources, but this cut-off point should be explored

further with regard to sensitive life stages and chronic (repeated) exposure. This discussion underscores the need for an overall roadmap, which contextualizes the priority setting objective in relation to the entire EDSP.

3) *Charged species of 99% or greater at pH 7.* The Panel agreed that compounds exhibiting a net charge of 99% or greater at pH 7 are highly unlikely to be bioavailable *in vivo.* However, there should be justification that pH 7 is a suitable pH to consider in environmental waters or *in vitro* systems. The Panel indicated that strong acids are un-ionized at the pH of the stomach, or in other circumstances problems may arise if a compound naturally forms a salt. However, the Panel agreed that 99% ionization for the compound and test system considered is appropriate. The Panel suggested the possibility that the criteria for pKa could be combined with the charge, as there is considerable overlap (as demonstrated in Table 2.2.2.1, white paper, p. 13).The Panel indicated that the definition of a charged species needs further clarification to include special situations. For instance, the definition of charged species with metal ions will be different from (simple) organic molecules.

2.1.d Other considerations

The Panel agreed that the purpose of the physico-chemical exclusion criteria (cut-offs) was to remove chemicals for which biological testing was not possible (outside the domain of reliable testing) or that were not available systemically in the whole organism (aspects of exposure such as bioavailability). Most of the Panel saw the exclusion criteria as independent of AOPs; however, there were panel members who noted there were exceptions, *e.g.*, charged species with metal ions (noted above), which need to be addressed on an AOP-specific basis. The ES, on the other hand, is ER-alpha AOP specific. Each ES would use a set of physico-chemical properties and filters specific to that AOP.

The Panel agreed that expert opinion should be used to review the physico-chemical property estimations that will be used as cut-offs. The process of using "cut-offs" is an iterative process. As the Agency gets more experience with a particular cut-off point and understands the extent of uncertainty in physico-chemical property estimations then they can revise the cut-off values or even develop different cut-off values for different classes of chemicals. Expert opinion can help with these decisions. Expert opinion is an important part of the weight-of-evidence approach to provide further confidence in excluding a chemical from testing, *e.g.*, if a chemical is excluded by more than one criterion. Expert opinion should be used to evaluate chemicals using the specific criteria for exclusion within the given time constraints. There are ways to streamline the filtering process without the use of experts for chemicals that are not near any cutoff points.

Appropriate statistical analysis, *e.g.*, an assessment of the performance of the cut-offs with regard to eliminating non-testable compounds without false negatives, could be used to evaluate the relative performance of the cut-off values. This could also assist in more clearly defining the domain of applicability of each filter and the extent of uncertainty.

The ER ES was originally built based on a training set (TrSet) of chemicals using in vitro assays specifically optimized to measure a well defined endpoint as indicated by ER binding and gene activation and to cover the domain of applicability (in the case of ESv1 food use pesticidal inert ingredients and antimicrobial active ingredients). This ES was the subject of a 2009 FIFRA SAP review. Since that review additional work has been done and additional TrSet data have been generated to cover pesticide non-food use inert ingredients (i.e., ESv2). In addition, an analysis has determined that the ESv2 can cover ~70 percent of the EDSP Chemical Universe that includes SDWA chemicals and fragrances.

Panel Response

Background

The ERES Version 1 (ESv1) is an effects-based ES that predicts Relative Binding Affinity (RBA) to the ER-alpha for food use inert ingredients and antimicrobial pesticides. The ESv1 was built using data from rainbow trout ER (rtER) binding and liver slice gene activation (GA) assays, as was the subsequent ESv2. The ES is a transparent, flexible and expandable logic rule-based decision tree that encodes mechanistic understanding with respect to both the chemical and biological aspects of ER-binding, and ER-alpha agonist stimulation of vitellogenin production in the liver slices.

A positive hit or "Active" classification is any statistically significant signal above background at any concentration up to solubility or cytotoxicity, whichever is reached first. Based on a training set of representative chemicals, ER agonist or binding ligand is defined as a chemical having an RBA compared with the endogenous ER-alpha ligand (17-β estradiol) of greater than 0.00001%, and a chemical that does not bind to ER-alpha as having an RBA of less than 0.000001%. Therefore, the ER-alpha ES is very sensitive to chemicals with low potencies (*e.g.*, those which some of the HTP ER assays could not detect), and provides mechanistic information that cannot always be deduced from the higher tier *in vivo* tests.

The applicable chemical domain in ESv1 was expanded to include non-food use inert substances that remained unclassifiable by ESv1 (223 discrete substances), *i.e.*, SDWA chemicals and fragrances - primarily mixed organics and mixed phenols. Selected chemicals from these groups were then tested, and the results modified the ESv1 training rule set data to extend chemical coverage, *i.e.*, ESv2. While the approach is sound, as with any model, the predictivity of the ES is only as good as the input data.

3.1. Please comment on the approach used to build ESv2 through chemical testing and effect-based category and read across methods to provide a scientifically defensible approach to predict ER binding potential for a larger number of chemical groups, particularly with regard to defining new groups from within the Mixed Organics and Mixed Phenols. Please indicate any considerations unique to ESv2 that indicate the approach used to develop ESv1 (SAP, 2009a) needs to be modified.

Panel Response

The Panel congratulated EPA on putting together a well-designed, thoughtful and consistent approach in ES development. The Panel agreed that the approach to expand ESv1 to ESv2 to include non-food use inert substances is robust and scientifically defensible. The ESv1 was expanded through rational analysis of chemical groups and directed testing to predict ER-alpha binding for a larger group of chemicals in the EDSP universe. While the Agency has made a good start at defining yes/no decision nodes within the ES (v1 and v2), the Panel suggested making some yes/no decisions more important than others, *e.g.*, applying weights or confidence scores to the final decision nodes (*i.e.*, chemical categories) within the ES decision tree.

Level of confidence – domain of ESv1 expanded to ESv2

The chemical domain of ESv1 was expanded to cover a larger universe (1,423 discrete substances) of non-food use (NF) inert ingredients. The ESv1 successfully classified 85% of the NF inerts list. There were 215 discrete substances not within the ESv1 model domain that were classified as chemicals with Unknown Binding Potential (UnkBP). The NF UnkBP chemicals were segregated into a series of structurally similar groups (*e.g.*, mixed phenols, mixed organics). Subsequently, chemicals were selected within the different structural groups of UnkBP and tested to expand the ERES TrSet to better cover NF use chemicals in the ESv2. The Panel considered the expansion of the chemical domain of ESv1 to ESv2 (to include NF inerts) to be based on sound science and organized in a logical fashion. The Panel noted that special rules were needed in addition to the general rules, *e.g.*, special rules are not listed in Table 3.2a.1 (white paper).

The Panel had several comments on the coverage of the expanded chemical domain in ESv2.

1) The 14 new structural groups added during expansion of ESv1 to ESv2 do not completely cover all the NF Mixed Phenols and Mixed Organics discussed during the August 2009 SAP meeting (see Appendix C, white paper).

2) New structural groups were coded into ESv2 based on strategic testing that expanded the ES TrSet, but these structural analogs (Fig. 3.2a.10, white paper) could also be divided into additional sub-classes for better characterization of the NF inerts and to better address the scope of the larger EDSP inventory.

3) Many chemicals in the fragrance list have unique structures, but these are not covered in the ESv2. The white paper did not have a description of how these will be covered in future iterations of the ES. These chemicals will be subjected to more targeted *in vitro* testing, but this had not yet been done.

4) The rationale for each chemical category and the structural boundaries needs to be clearly stated as a protocol, algorithm, or step-wise procedure. For example, categories might be organized first by largest common substructure, important functional groups, or some standard metric of chemical structural similarity, *etc.*, as the first step. Then in the second step, an expert could manually break "ties" between competing categories for any chemical that would have been assigned to multiple categories in the first step. Whereas a procedure like this may already be in place, it would be useful to delineate it in the white paper explicitly. For ESv2 and before, the chemical category organization seems to derive from manual expert curation, with descriptive rationale for each category illustrated by detailed multi-page examples in the white paper. However, a different expert may organize some categories differently, so establishing a standard step-wise procedure (even one incorporating a human-supervised final selection) would improve transparency going forward. Much of this work depends upon the testing still to be done.

5) The Panel commented that additional groups or sub-groups within those currently proposed are likely to be identified and added to new versions of the ES as part of the iterative process. Furthermore, as new training set data (from the continued testing of representative chemicals in TrSet assays for each of the structural groups or subgroups) become available, it will be important to identify performance statistics that can be used to assess the predictive performance of each new and existing node. As the currently untested chemicals within a node (chemical category) are eventually run in TrSet assays, there is an opportunity to assess the node's predictive performance retrospectively. Since this will take some time, the Panel advised EPA to assign node-by-node weightings as soon as is feasible. The Panel offered two approaches to set preliminary node-by-node weightings.

Random testing of node category members. Chemicals that have been grouped into "covered" categories, but have not yet been tested, might be selected at random for TrSet confirmatory testing, as one would do for internal QSAR model validation, using multivariate data analysis tools. The disadvantage of a random-confirmation approach is that it could easily miss problematic chemicals; however, the advantage is that it is an unbiased way to test the nodes.

Compare chemicals from within the ESv2 or ESv3 nodes to preliminary HTP data, where available (preferred). Cases where the ES-derived predictions diverge from HTP results could be prioritized for earlier TrSet testing to resolve the discrepancies. This would serve as a check on both the predictive accuracy of the HTP assays and on the predictive power of the ES. Charge question #7.2 expands on this concept, extending it from the confirmatory testing described here to a prioritization decision-making metric.

More broadly, the Panel recommended that ESv2 and subsequent versions should be subject to quantitative performance testing. Approaches would include: (1) analysis of "blind" out-group chemicals not explicitly included within the model but within the model domain; (2) cross-fold analysis for sensitivity, specificity, and positive and negative predictive value, via automated data subdivision and test ES model rebuilding; and (3) parallel analysis and testing of new chemicals. Such analyses

would be consistent with earlier SAP recommendations (in this report) and quantify the model's dependence on known chemistry. If the Agency feels that insufficient data exist to reliably quantify performance as suggested above, an alternate approach might be to use agreement with HTP data as a "vote" supporting those ESv3 nodes containing at least a few chemicals overlapping with and consistent with the HTP results. Finally, the Agency may also ask how well the ES as a model stands up to random temporary deletion or insertion of incorrect chemical-to-outcome data within its nodes. If individual decision branches or nodes are more susceptible to disruption from such 'scrambling," then these are areas of the decision tree that may warrant greater confirmatory testing to confirm actual reliability. But this 'scrambling" approach can only be applied if nodes and/or tree branches have performance scores (also referred to here as "weights") assigned as recommended above, so that the effect of the scrambling on these scores can be assessed.

Considerations for expression of the coverage in terms of percentages

The coverage of each version of the ES is given in terms of percentages. This number is useful but it may be illustrative to show the coverage graphically, *e.g.*, the distributions of chemicals, and coverage in terms of 2-D or 3-D plot of descriptors. This could have the advantages of (a) being simpler to comprehend and (b) illustrating where there may be large areas of the chemical universe that have not been covered in this approach.

Coverage is an important metric, but as currently used, it implicitly assumes uniform predictive accuracy of the ES across the covered space. It is possible that some chemical categories (*i.e.*, nodes within the ES decision tree) may have greater predictive accuracy than others, and this is not quantified or expressed; whereas, the output from the ES is an assignment of each chemical into one of only three distinct categories, *i.e.*, active, inactive, or unknown. The Panel stated that it should be possible to assign confidences to each such assignment and rank chemicals. The question is whether or not potency in the assays could be used to relatively rank substances to distinguish highest priorities for testing. This would distinguish areas of complete but less reliable coverage from those that are more reliable, as supported by retrospective or empirical evidence, such as using a commercial available tool such as Spotfire® gene expression and chemical properties (Kaushal and Naeve 2004a, b).

Level of confidence in a prediction –weight of evidence approach

The level of confidence in a prediction based on an alternative method (*e.g., in vitro* or *in silico*) is often centered on the summary experience with the method and the robustness of the data that it has provided. Increasing confidence in a prediction is related to having transparency and a mechanistically probable rationale for the prediction and increasing the weight-of-evidence supporting the prediction. There is growing agreement that the mechanistic basis for a prediction is best provided by an AOP. The weight-of-evidence analysis is based on the relevance of each key event to the final *in vivo* endpoint in question, the number of methods used to assess each key event, the relevance of each method to the particular event, and the reliability of the databases supporting each method. In the case of ER-alpha AOP, the rtER binding and liver slice GA assays are highly relevant to predicting reproductive impairment in rainbow trout (Schmeider *et al.* 2004, Schmeider *et al.* 2000). Moreover, using them in tandem increases confidence in predictions. As the ER-alpha ES TrSet is expanded, confidence in the predictions will increase.

A second aspect of confidence in a prediction is the perceived consequence of the prediction being incorrect. The net result is that there is greater confidence in accepting a positive or toxic prediction than accepting a negative or non-toxic prediction. One can take this argument one step further. For example, if both 1-butanol and 1-octanol were tested and shown to be negative in both the rtER binding and liver slice GA assays, there should be a greater confidence that 1-hexanol would also be a negative chemical. However, extending that prediction to 1,12-dihydroxydodecane, another aliphatic alcohol but one with 2 polar groups which can be configured 10.2 A apart, should be made with less confidence. The weak point in the approach is how similarity is used to develop the chemical groups. Confidence may vary with the rationale and structural boundaries of the chemical group on which the prediction is based.

Current and emerging computational approaches may be incorporated as a 'second pass" follow-up screen for chemicals that ESv2 and ESv3 assign an "unknown binding potential." These include Naïve Bayesian classifiers (Besnard *et al.* 2012, Paolini *et al.* 2006, Nidhi *et al.* 2006; Rogers *et al.* 2005), and statistical similarity methods such as the Similarity Ensemble Approach (Keiser *et al.* 2009; Lounkine *et al.* 2012) or others (Campillos *et al.* 2008, Vidal *et al.* 2011). These approaches are less susceptible to the QSAR limitations described on page 24 of Appendix B, and may defensibly be applied across the "diverse structures with widely varying physical-chemical properties" (Appendix B, p. 27, white paper) that characterize the inventories of interest. By contrast, the current ESv1-3 must rely on an "exact match" for chemicals such as those from the mixed organics (MO) and mixed phenols (MP) categories, which lack general rules (Appendix B, p. 34, white paper). An "exact match" strategy alone may not scale well because it relies on chemical-by-chemical "brute force" experimental testing. When used as an augmenting or 'second pass" analysis, these chemical informatics and systems pharmacology computational methods may increase the recovery rate of compounds that would otherwise remain UnkBP chemicals or false negatives.

Considerations for software platform chosen

The Panel noted that the software platform for the ES is developed from a proprietary system. While the platform is a robust and proven system, and one that the Panel applauds the Agency for providing to the public, the Agency should also consider other chemoinfomatics approaches (see response to charge question #2), and analytical tools (platforms) that are not proprietary and offer greater flexibility, *e.g.*, multivariate data analysis tools including hierarchical analysis tools such as the KNIME (Konstanz Information Miner) platform (KNIME.com GmbH, Zurich, Switzerland). KNIME workflows are available free-of-charge from KNIME.com. They offer greater flexibility (than the EPA's platform) and provide the possibility of developing complex structural alerts (*e.g.*, for grouping) as well as links to a wide variety of chemoinformatic tools and databases (Saubern *et al*. 2011). KNIME workflows have been used for toxicity prediction (Nelms *et al.* 2012). Another possible platform is the commercial software, Spotfire® (TIBCO Software). Spotfire® is commonly used in the pharmaceutical industry to identity chemical/biological properties and performs clustering analysis in a highly visual and accessible format. The Panel commented that without freely available versions of the ES (non-proprietary software) then distribution and transparency are more problematic.

As noted by the Panel, ligand-dependent activation of ER is highly complex and dependent on ligand structure, ER subtype and intracellular location, promoter and cell context. Different ER cell and

promoter contexts can explain the dramatic functional differences among ER agonists, partial agonists and antagonists (Gould *et al*. 1998, McDonnell *et al*. 1995, Yoon *et al*. 2000, Yoon *et al*. 2001). Transactivation of genes with an estrogen response element (ERE) motif depends not only on the core consensus ERE, but also the flanking regions, histone modifications, cofactor expression (*e.g.*, cell context dependent differences) and other variables. For example, the pS2[1] (TFF1) gene contains a non-consensus ERE (Berry *et al.* 1989), but in some cell lines activation of this gene is dependent in part on the activator protein 1 (AP1) transcription factor response element (Barkhem *et al.* 2002a). In another example, transcriptional synergism of the pS2 gene promoter between a160 coactivator and ER-alpha depends on the coactivator subtype, the type of ERE, and the promoter context (Barkem *et al.* 2002b). The recombinant reporter gene construct, ERE-luciferase (Luc), has been used as a sensitive indicator of estrogenic activity for many years (discussed in Zacharewski, 1997). McDonnell and coworkers, who were among the first to develop screening assays to distinguish between selective estrogen modulators (SERMs), used the complement C3 gene promoter complex (complement 3-luciferase construct containing a complement 3 gene promoter sequence linked to a luciferase report gene) as a detection system (McDonnell *et al.* 1995). Safe and coworkers have used this gene promoter to screen endocrine disruptors and demonstrate that they too are SERMS (Yoon *et al.* 2000, Yoon *et al.* 2001).

In addition to ligand-bound ERs binding directly to EREs in the promoters of target genes (ERE-dependent), they can also interact with other transcription factors such as Fos/Jun (responsive elements part of AP1-transcription factor complex) or specificity protein 1 (Sp1) and thereby regulate the transcription of genes whose promoters do not harbor EREs (ERE-independent) (Safe and Kim 2008). Another example of an ERE-independent response is the interaction between ER-alpha and nuclear factor-kappaB (NFκB) complex. This interaction prevents NFκB from activating interleukin-6 (IL-6) promoter and cause immunomodulation. In addition to the well-studied transcriptional effects of estrogen, there are also rapid effects that lead to the activation of kinases and phosphatases (Marino *et al*. 2006). The Panel indicated that all physiological mechanisms by which ERs regulate the expression of target genes are important; however, pathways may differ among cell types and tissues.

The Panel acknowledged that further study of the mechanisms and proteins associated with coactivation and repression of ER/transcription factor complexes is warranted to elucidate their role in mediating the effects of estrogens. The Panel suggested that more complex and "realistic" promoter–reporter constructs be considered to study the effect of different ER-promoter complexes on gene activation among chemicals that are ER agonists, partial agonists, and antagonists. Also, assays for ER-

[1] pS2, also known as Trefoil factor-1 (TFF1), belongs to the family of TFFs that possess a characteristic three-loop structure named trefoil or P domain. The human TFF1 protein is expressed predominantly in the foveolar epithelial surface cells of the gastric mucosa and demonstrates a pattern of increased expression during inflammation. TFF1 and other TFF family members are up-regulated and secreted in an autocrine and/or paracrine manner in response to inflammation and gastrointestinal damage and are also expressed in tumors.

.

beta and membrane-bound estrogen receptors should be incorporated as these may be somewhat different in their transactivation potentials compared to ER-alpha. Some assays, *e.g.*, BG1-Luc (OECD TG 457) also have some ER beta.

The systems biology network approach, used in pharmacological research, may also be an adaptable approach for more complex and integrated AOP development and elucidation in the future. In drug metabolism, for example, a network approach identifies cross-interactions (mutual "cross-talks") between signaling pathways for two or more receptors (*e.g.,* ER, aryl hydrogen receptor (AhR), pregnane X receptor (PXR)) which result in different functional molecular and physiological consequences. In the ER-alpha transactivation HTP assay, expression of AhR in the cell (or also PXR) could competitively affect transactivation potency. The use of a systems biology network approach could be used to examine the interactions of ER signaling with other signaling pathways (*e.g.*, AhR, PXR) and activation of the ER-alpha AOP (or within assay cell lines*, e.g.*, a meta-analysis of expression microarray datasets, such as that for 17β estradiol in MCF-7cells, Oschner *et al.* 2009, see also discussion in CQ 8), and other relevant AOPs.

Figure 2 illustrates the complexities of estrogen dependent signaling and crosstalk with AhR as an example. AhR ligands, such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), activate inhibitory AhR-ERα crosstalk through multiple pathways, and other interactions between the two signaling pathways have been reported (see review by Matthews and Gustaffsson, 2006*).* ERs and AhR are ligand activated transcription factors and members of the nuclear receptor and basic-helix-loop-helix Per (Period)–ARNT (aryl hydrocarbon nuclear translocator)–SIM (single minded) (bHLH-PAS) superfamilies, respectively. Treatment of MCF-7 breast cancer cells with TCDD activates ER-alpha directly if AhR is silenced (Abdelrahim *et al*. 2003). Wormke *et al.* (2003) identified inhibitory AhR-ER-alpha cross talk linked to degradation of ER-alpha in which TCDD initially induces formation of a nuclear AhR complex which coordinately recruits ER-alpha and the proteasome complex, resulting in degradation of both receptors.

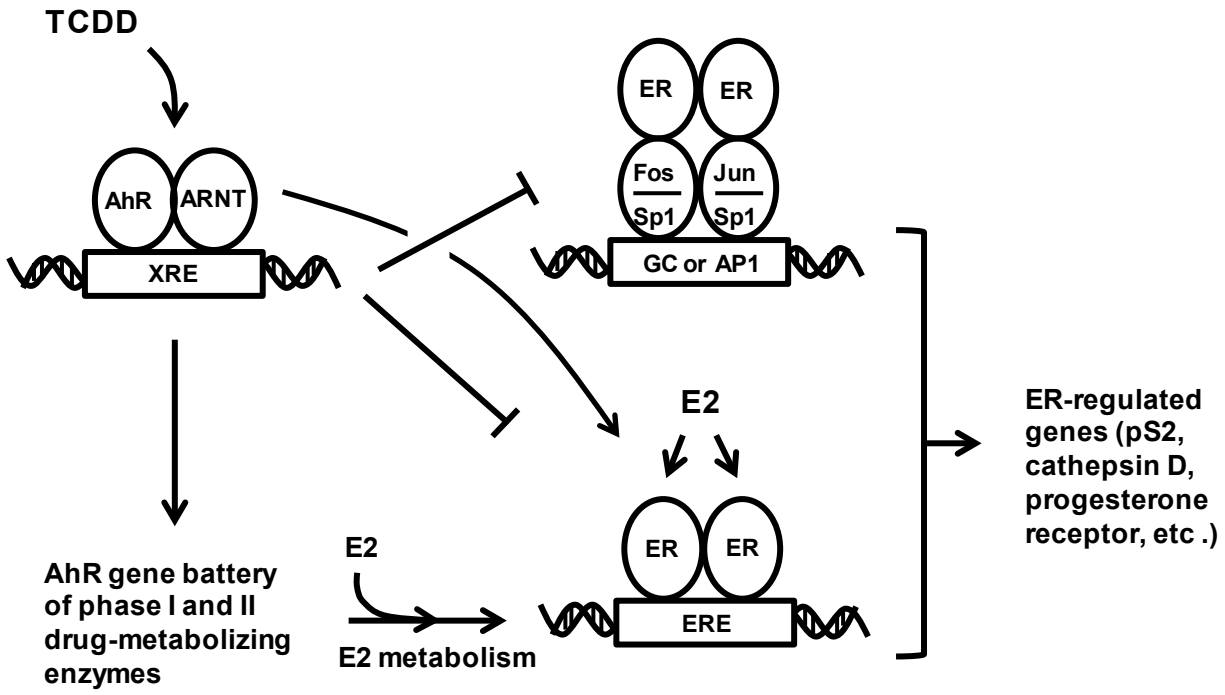

Figure 2. 2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin (TCDD) modulation of ER and AhR pathways in MCF-7 cells. TCDD is a potent AhR agonist. In this illustration, TCDD activates ER-alpha directly when AhR is silenced (Abdelrahim *et al*. 2003). The activated AhR-XRE (xenobiotic response element) ligand translocates from the cytoplasm to the nucleus where it binds to its dimerization partner ARNT. The activated AhR/ARNT heterodimer complex binds to its cognate DNA sequences (XREs), and activates the expression of AhR target genes, such as cytochrome P4501A1 (CYP1A1) and CYP1B1, and several phase II drug metabolizing enzymes. ER ligand binding with17β estradiol (E2) induces receptor homodimerization, DNA binding to estrogen response elements (ERE) in the promoter regions of target genes, recruitment of coregulators, and changes in transcription. ER also activates gene expression through interactions with other DNA bound transcription factors such as Sp1 and AP1, and membrane ER also regulates gene expression. In this illustration, activation of genes with non-ERE promoters in which ER-alpha interacts with other DNA bound transcription factors, including Fos/Jun, elements of the AP1 transcription factor complex, or specificity protein (Sp) transcription factors such as Sp1, and ARNT coactivate ER-alpha-mediated transcription. Abbreviations: AP1, activator protein 1; AhR, aryl hydrocarbon receptor; ARNT, aryl hydrocarbon receptor nuclear translocator; CYP, cytochrome; E2, 17β estradiol; ER, estrogen receptor; ERE, estrogen response element; Fos/Jun AP1, Fos/Jun dimer activator protein 1; GC, GC box element; Sp1, specificity factor 1; XRE, xenobiotic response element; TCCD, 2,3,7,8 tetrachlorodibenzo-*p*-dioxin.

3.2 Building from the ESv1 training set, please comment on the level of scientific confidence that a chemical is unlikely to initiate the ER AOP, if the in vitro *assay TrSet data shows no activity. Please comment on the extent to which the level of confidence may vary by chemical category.*

Panel Response

Preface

The Panel noted that while the ER-alpha AOP is a significant component in endocrine disruption, there are other possible MIE's that may lead to downstream estrogenic effects that were not considered by the Agency as part of their prioritization strategy. The Panel spent a considerable amount of time discussing these other MIEs, which was beyond the specific focus of this charge question. At the end of the Panel's discussion, the Agency clarified and reiterated that the initial focus of the prioritization efforts is on the ER-alpha AOP (versus other possible MIEs) and apologized for any confusion. This one specific AOP is the basis for all of the *in vitro* assays in building the ES and the HTP assays used in the prioritization scheme. The Panel's comments on the consideration of other MIEs in the overall prioritization efforts are found near the end of the Panel's response to CQ 1. To avoid confusion, the Panel recommended that the Agency state more clearly that the focus of the initial prioritization approach is on the ER-alpha AOP and that other possible MIEs and AOPs would be considered in the future as part of the prioritization approach. With this in mind, the Panel agreed that the Agency's initial focus on the ER-alpha AOP as the basis for the proposed prioritization scheme is scientifically sound, logical, and practical, but recommended that this should be clarified again, in the context of other possible MIEs that may lead to downstream estrogenic effects. Here an ER-alpha AOP roadmap including indication of potential for other AOPs to be added in future would be valuable for future Agency planning, experimental design and subsequent testing. See earlier comments in response to CQ 1.

Specific Comments

The Panel made the following comments concerning the level of confidence in the *in vitro* TrSet data as to whether or not a chemical would or would not initiate the ER-alpha AOP.

1) *Agreement with the August 2009 SAP.* The Panel agreed with the August 2009 SAP's evaluation of the ESv1 TrSet data (see SAP report, Appendix C of the white paper). As stated in the 2009 SAP report, ESv1 met the five of the OECD QSAR validation principles of a well-defined endpoint, clear mechanistic interpretation, unambiguous algorithm, goodness of fit, and domain of applicability. The ESv1 training set was reliable and robust for the applicability domain of food use inert ingredients and antimicrobial pesticides. Both *in vitro* tests, rtER binding and liver slice gene activation (GA) assays, are sensitive assays, and, when used in tandem, increase the confidence in the predictions of ER binding and activation of the ER-alpha AOP. Whilst gene activation, such as vitellogenin mRNA activation, may not be indicative of an adverse outcome, a downstream event such as vitellogenin protein induction may represent an ER-alpha AOP. Therefore, vitellogenin mRNA presence is only indicative of an AOP if the protein is confirmed, as shown in the liver slice assay.

2) *Lack of data for some chemical categories leads to uncertainty.* The level of scientific confidence in a prediction based on a level 1 or 2 test method (c.f. revised OECD ED conceptual framework, see OECD 2012a) such as the *in vitro* assays the TrSet depends upon the proven or demonstrated reliability of the method and the robustness of the data which it has provided. The Panel agreed that the level of confidence in some chemical categories could not be defined because there was a lack of *in vitro* data, *e.g.*, fragrances. See the Panel's response to CQ 3.1. As the ES TrSet is expanded, confidence will increase.

3) *Size of chemical category will affect the level of confidence.* If the categories are to be used to assign guilt by association for unknown binders, then the size of the category is going to be an important consideration of your confidence in a prediction. Small categories aren't going to have appropriate sampling for confidence calculations, and very large categories with a broad range of binding potential (binders and nonbinders) should perhaps be considered for splitting to better capture the discriminating features using informatics approaches instead of supervised approaches using 2D structural features.

The Panel recommended that the definitions of categories and rule sets need further refinement (see response to CQ 3.1). Perhaps a protocol could be developed to quantify the level of confidence (also referred to as node "performance" in CQ 3.1) per category.

4) *Choice of test chemical, relevant MIE, and assay chosen will affect the level of certainty.* The Panel commented that ER-alpha binding is dependent on the intrinsic properties of the test chemical, the relevant MIE, and the assay chosen. There are also species-specific differences in sensitivities of ER-alpha to hormones and environmental chemicals that could influence the level of certainty about species specific classification of endocrine disruptors. For example, the HTP gene transactivation assay is based on a frog consensus ERE for vitellogenin, which is rarely expressed in any other ER responsive gene (other species). The human ER-alpha is not the same as the rainbow trout ER-alpha, about 60% similarity overall. There is higher similarity in the DNA binding domain (95%) than in the ligand binding domain. There are also differences in the affinity of agonists for the ER ligand binding domains between fish and humans, with higher affinity seen for human ERs. In addition, there are chemicals that bind differently to ER-alpha and to ER-beta.

5) The Panel noted that no data were provided for antagonism, although the OECD method, BG1-Luc ER transactivation test method for identifying ER agonists and antagonists assay (TG 457, OECD 2012c) can address this. The Panel also understands that antagonism/partial antagonism HTP data, *etc*. are available from some of the Tox21 assays (not presented in the white paper), but antagonism should also be incorporated into the ES, as chemicals that are antagonists could interfere the ER-alpha AOP leading to adverse effects.

While the focus of the prioritization approach is on the ER-alpha AOP, the Panel noted that some of the assays considered are inclusive of other estrogenic endpoints, *e.g.*, liver slice *in vitro*

assay (could include membrane receptor and ER-beta), selected ToxCast and Tox21 HTP assays. Broadening of the ES to these other estrogenic endpoints should include existing HTP assays for ER-beta, receptor dimerization, co-activators, *etc*.

6) *A chemical class might not fit into the classical ER-alpha binding AOP, but may however be relevant to estrogenic AOPs.* The phthalates, for example, might be a chemical class that might not fit into the ER-alpha AOP, but fits in a different estrogenic or steroidal AOP, to which the ER MIE belongs. Many anti-androgens trigger the same AOPs as estrogens, but through another mechanism. Thus, a negative response in the TrSet data set does not necessarily predict that a chemical is not an estrogen, since it may be activating a different estrogenic AOP, for example, through ER-beta or through the membrane ER, not the same mechanism as incorporated by the ES. The ES is still highly appropriate for prioritizing for Tier 1 screening, and can potentially identify a majority of potentially higher risk chemicals that will require further testing, but with the caveat that there might be occasions in which some "positive" chemicals are not identified and become part of the group of chemicals with unknown binding potential. That is, they may be identified as chemicals with "no binding activity" based on the selected *in vitro* assays, but may be able to initiate another adverse estrogenic pathway via another MIE and adverse endocrine pathway.

7) *Category approach does not provide equal emphasis on chemicals that bind to the B site.* The category approach used identifies chemicals that have the probability to bind the A site (ER-A), but not the B site of the ER (ER-B). The Panel indicated that it is important to know the probability of binding to the ER-A site versus ER-B site (Katzenellenbogen *et al.* 2003). The Panel believes it would also be important to measure binding to the C site, not at all captured by the ES.

The Panel also made the following suggestions as to how to determine whether an assay is a good prognosticator.

1) *Sources of collated ER modulator data from the literature*. The scientific literature contains a large number of chemicals reported to modulate the ER-alpha. However, as the quality of such data can vary widely, the Panel noted that the task of manually curating ER-alpha binders from this primary literature—especially those being reported in multiple journals on a monthly basis—would be substantial. One resource that may augment or side-step this effort, however, is the freely-available ChEMBL database, downloadable at any time at https://www.ebi.ac.uk/chembl/. This database collates and annotates chemical-to-molecular-target information from many scientific journals. As such, it may serve as an initial starting place to confirm that the inventory chemicals queued for testing have not *already* been reported to bind the ER-alpha as well as means to track new testing results from other labs as these data grow over time.

2) *Use of selective estrogen receptor modulators*. If chemicals are considered as selective estrogen receptor modulators (SERMs) of the DNA binding domain for example, one suggestion would be to test this further by modifying or knocking-out the DNA binding domain and examine

consequent downstream events that would depend upon the particular chemical SERM modulation. See also CQ 3.1 and CQ 5.3 for further relevant discussion.

Charge Question 4: HTP ER Binding and ER Transactivation Data (Sections 4-6)

The Agency *is proposing to use the HTP data to expand the ER expert system (i.e., ESv3). Therefore, it is important to understand how the HTP data are generated and interpreted for the intended use. As endorsed by the 2010 SAP on Integrated Approaches to Testing and Assessment Strategy: Use of New Computational and Molecular Tools, in evaluating computational tools, the* Agency *will consider internationally accepted science principles including the OECD QSAR validation principles of "Clear, Defined Biological Endpoint, Mechanistic Interpretation, Unambiguous Algorithm, Goodness of Fit and Domain of Applicability." These principles are consistent with and complement the recommendations of the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) and considered flexible to be broadly applicable to predictive models or computational tools which may be based on HTP or other in vitro data. The principles particularly applicable to this charge question are: a well-defined endpoint and mechanistic interpretation. A "well defined endpoint" is intended to ensure clarity in the endpoint data used to build predictive models, chemical categories, or read-across. Any predictions made from the data will inherently contain all uncertainties and limitations in the data measurement and interpretation. The intent of "mechanistic interpretation" is to ensure a mechanistic association between the attributes of a chemical and its interaction with the biological system resulting in the measured endpoint, to the degree possible. The following questions relate to the transparency of the HTP techniques (e.g., how the data are processed and interpreted and the extent of reliability and consistency within an assay and among HTP assays).*

Panel Response

General comments

The Panel has been requested to provide advice to the EPA on a prioritization approach for the universe of EDSP chemicals based on potential impacts of the ER-alpha AOP. The ER-alpha AOP involves an estrogen/estrogenic compound binding to the ER where the molecular initiating event is activation of estrogen receptor alpha via ligand binding and subsequent transcriptional activation via estrogen response elements in promoter regions. While the ER-alpha AOP is a significant pathway for xenoestrogen action, the Panel recognizes that there are other pathways as well (see comments in CQ 1).

In CQ 4, the Panel was asked to address the use of HTP data to expand the ES (*i.e.*, from ESv2 to ESv3). If the intent of the ER-alpha ES is to validate chemicals that are estrogen agonists/antagonists, then the HTP assays evaluated may not capture all compounds that behave as estrogens or anti-estrogens, as the assays in the ES are specific to only one aspect of estrogen action, *i.e.*, mediated through ER-alpha and canonical ERE's in the DNA. Not included are effects through ER-beta, the membrane receptor and non-genomic estrogen action. Even for ER-alpha, not all aspects of its transactivation were tested, for example specific steps in the process that occur after binding of the ligand to the receptor were not necessarily captured by the assays. There are HTP assays available in the

Tox21 or ToxCast repertoire that capture these other steps, as was indicated in the presentations, but these assays were not used in this context, *e.g.*, receptor dimerization, receptor binding to other response elements, coactivators, *etc*. The Panel indicated that the ES could be broadened to include these end points by using additional HTP assays already available in Tox21 and ToxCast.

The ESv3, as currently defined, will focus only on ER-alpha binding to a canonical ERE and thus would have a more narrow focus of the number of estrogen ligands binned for testing. This narrower focus has the following advantages that ESs for other endocrine disruption mechanisms currently lack.

1) The ERES is based on the well-documented ER-alpha AOP, where the downstream effects at the different levels of biological organization are "hard-wired" and easily scaled. The net result of having a hard-wired AOP is that only a few key events (*e.g.*, rtER binding and liver slice gene activation (GA)) need to be assessed with large numbers of chemicals.

2) The ER-alpha MIE is a well-established MIE with two decades of studies supporting its application. In addition, GA is an established key event being examined in a variety of protocols (*e.g.*, yeast, mammalian cells) that also support its application in this context. More specifically, the liver slice assay dating from 2004 has withstood the scrutiny of time. Thus, the mechanistic interpretation and the reliability and consistency of data derived from the two *in vitro* assays are acceptable for this specific ER-alpha AOP.

4.1 Given the importance of a well-defined endpoint to determining if/how endpoint data are appropriately used, the panel is asked to comment on:

Panel Response

(a) Whether the assays are sufficiently described so that others can reconstruct the assay conditions and data analysis and to what extent additional information would be useful?

There was considerable discussion about how well the HTP assays were described. In general, the Agency described each of the assays well for how the assays worked, the detection system used for each, the endpoint tested, how the assays dealt with cytotoxicity, acceptance criteria, and how known issues affected the performance of the assays. However, several panel members criticized the Agency for not providing the original references for the assays, which were difficult to find and in one case, the reference was wrong. The primary description of the assays needs to be better referenced. For example, the primary reference for the BG1 assay was a four page abstract although there was also a link provided to a full SOP.

There was also considerable discussion about the use of proprietary assays that only one vendor/laboratory can perform and for which the performance of the assay cannot be tested by a third party. This means that the assay performance and reliability cannot be assessed independently of the vendor/laboratory. The Panel suggested that the assay performance could be tested for repeatability by scrambling samples and keeping the identity of samples blinded to the vendor, but this process would need to be made transparent to gain universal acceptance. However, this process would not test reproducibility by different vendors or different interested individuals.

The Panel was not clear on why the eight assays discussed in the white paper were selected from among the 21 assays available from the Tox21 program. The selection criteria for the eight assays should be more transparent. There was some concern expressed by the Panel that the eight assays picked were selected to test basically the same steps in the process and therefore would give the same type of data rather than selecting assays that would test different steps in the process which could then provide additional criteria for deciding that a chemical is not an estrogen. The Panel suggested that it might have been better to select assays that would not necessarily test the exact same step, to broaden the information used for grouping chemicals into whether they are/are not an estrogen mimic. For example, one could select a panel of different gene activation assays that recognize different promoter elements and this may add more information for the ERES.

(*b) Whether EPA has described sufficiently the important experimental conditions that affect the assay measurements, and whether EPA has sufficiently discussed their potential impact on assay results and interpretation of the results?*

There was considerable discussion about this point.

- Some panel members commented that more information should have been provided about the cellular characterization background for each of the transactivation assays. Interpretation of results should be made in terms of the cellular characterization background as estrogen receptors require cofactors for function and these cofactors vary by cell type. Thus information on the characterization of the cell line is needed to account for differential binding and activation, and this is noted in the performance based OECD *in vitro* test guideline TG 455 (OECD a2012b).

- The Panel believes that there was insufficient discussion about the potential metabolism of chemicals in the different cellular backgrounds and how this might compare to an *in vivo* situation. For example, the Panel asked the question, "Were all chemicals that can normally be metabolized into estrogen-active derivatives, metabolized correctly in the cells?"

- The Panel appreciated the efforts that were put into understanding the effect of protein concentration on the binding assays included in the HTP assay screens. This was very well done; however, there were no plans mentioned to improve the experimental design of the assays with respect to, for example, the maintenance of correct protein concentration.

- Assays that rely on cell cultures may be subject to bioavailability constraints that media constituents may impart. These constraints may result in false negatives since protein constituents may bind and reduce bioavailability. Alternatively, protein binding may enhance material uptake depending on the mechanism of transport from the medium into the cell. Klaassen *et al.* (1999) reported that binding of retinoids to bovine serum albumin (BSA) significantly reduced their uptake by keratinocytes. The permeability of mannitol was significantly different in Caco-2 cells cultured in fetal bovine serum versus iron-supplemented calf serum (Lentz *et al.* 2000). Note: It seems likely that some of these effects might be predicted from physical/chemical characteristics of the test substance and hence could be further scrutinized when and if that compound is tested.

(c) Whether the chemical library (e.g., purity and analysis), chemical exposure (e.g., solvents used, chemical dilutions), and plates and plate layouts provides information necessary for data interpretation?

While the Panel appreciates that it is difficult to obtain all chemicals at high enough purity, they caution against using chemicals that are not fully characterized. The purity of the test chemical and information on its purity is very important to understand whether the reported results are due to the test chemical or to trace levels of contaminants, *e.g.*, dioxins, in the case of non-dioxin-like (NDL) polychlorinated biphenyls (PCBs) (Hamers *et al*. 2011), which might give false negative or false positive results, or due to a different form of the chemical, *e.g.*, mixtures of straight chain and branched chain nonylphenol.

(d) The adequacy of test concentrations (maximum and minimum concentration tested) and cutoffs, and whether the concentration cutoffs affected some assays results more than others, especially in the context of false negatives.

The Panel noted that the test concentration range and cut offs used in some of the HTP assay datasets discussed in the white paper were lower than those used in the rtER binding and GA assays used as gold standards in detecting ER activation. Because of this, the HTP assay datasets discussed in the white paper do not provide adequate or appropriate data for performance comparison of the two technologies (HTP *in vitro* assays vs. manual *in vitro* rainbow trout studies) or for assessing whether the HTP assays have the potential to be sensitive detectors of ER activators. The use of only one test concentration at 25 µM in many of the HTP assays discussed in the white paper was considered insufficient, resulting in misclassification of low affinity ligands as ER-alpha negative. It would be better to have full dose response curves, even for estradiol as the positive control, where in some cases the estradiol positive control was performed with only two concentrations. Additionally, other HTP assays discussed in the white paper were performed to a maximum concentration of 100 µM, far lower than that examined in the rtER assays, which were optimized to detect low potency ER-alpha activators. See additional discussion in response to CQ 6.

The Panel appreciated the care that was taken for calculations and rejection of outliers. The very conservative threshold calculation for the lowest effective concentration (LEC) could lead to identifying a chemical as non-active when it is actually active at high concentration, *i.e.*, increased false negatives. This could explain some of the discrepancy observed between the assays.

4.2 With respect to data interpretation, please comment on the approach for defining an active chemical (i.e., to initiate the ER AOP) and an inactive chemical (i.e., unlikely to initiate the ER AOP) compound and whether the method of data interpretation is adequately described and if the rationale for the approach is sufficiently presented.

Panel Response

The Panel agreed that the method of data interpretation (RBA cut off value of 0.00001%) is adequately described and the rationale for the approach (experimental rtER binding and liver slice GA)

is sufficiently presented in the white paper. The ESv1, shown in Figure 3.1.2 (white paper), is an expandable rule-based decision tree that was built using data from the rtER binding and liver slice GA assays (Figure 3.1.2, white paper) and follows the OECD QSAR guidelines. The scientific rationale for each of the decision rules in ESv1 was reviewed and accepted by the 2009 SAP (see 2009 SAP report in Appendix C). Using ESv1, a chemical is classified as "active" if it has a RBA, as compared with the endogenous ER-alpha ligand (17-β estradiol), of greater than 0.00001%, or "inactive" if it has an RBA <0.00001%. Subsequent to the 2009 SAP review, EPA expanded and modified ESv1 to create ESv2 (Figure 3.2a.1, white paper) using the same approach. This approach is consistent with current thinking of other organizations, *e.g.*, OECD, and has withstood the test of time, (albeit short). Moreover, as demonstrated in screen-shots of Figures 3.1.3A and 3.1.3B of the white paper, the ES is amenable to automation. Additionally, the supporting information provided by EPA in the white paper (lines 643 to 862) is sufficient to support the expansion of the ESv2 decision tree to ESv3.

Please include in your comments:

(a) The adequacy of the approach to generate the data plots; how and what parameters are calculated, etc.

The Panel unanimously agreed that the approach was clear and concise.

(b) How the background data were used to establish the control level.

The Panel indicated that the background information permits the unambiguous establishment of control criteria. However, DMSO can be weakly estrogenic (*e.g.,* Mortensen and Arukwe 2006, Rasmussen and Nielsen 2002, Timm 2013, Immonen *et al*., 2009) and should be called a "vehicle control" rather than a "negative control." The Panel recommended that the mean value for the dimethyl sulfoxide (DMSO) vehicle control (VC) should be calculated, if not already done, and then the mean value of the DMSO VC from each well value should be subtracted to normalize the data as noted in the OECD validated performance-based ER-alpha stably transfected transactivation *in vitro* test guidelines (TG 455) (OECD 2012b). Indeed all vehicles should be normalized.

(c) Whether the process for identifying assay interference is adequately described

The Panel agreed the process for identifying assay and data interference, which is based on the ER-alpha AOP, is adequately described. However, this process does not describe fully how to deal with the assay interferences. The current plan appears to exclude data that comes from the interference, but does not offer any method to fix it. There is also no explanation of how assay interference could affect false negatives.

(d) Adequacy of data normalization, outlier identification, curve fitting, background subtractions, and all other data processing and calculation techniques, and appropriateness of the statistical analyses.

The Panel stated that the data analyses, *etc*. meet the currently accepted standards, especially as stated by OECD. However there was no mention of when the chemical testing is conducted in a blind manner, *i.e.*, without prior knowledge of the chemical identity.

4.3 Would you recommend other considerations or approaches to analyzing the HTP data?

Panel Response

While data analysis will always be subject to other techniques, the approaches used are transparent, mechanistically probable, allow for a clear statement of the weight-of-evidence and allow for hypothesis-based, targeted testing. However, some of the Panel noted that there were additional software packages that could help in data analysis, *e.g.*, Spotfire® (TIBCO Software) and hierarchical cluster analysis software (see **Table 2**), to help aggregate the results by chemical class. There are also agglomerative techniques that can pull in multiple types of analyses, including weighting schemes.

Each assay was well-controlled and validated. However, it was difficult to understand how the data were integrated into a decision point; The Panel suggested that the addition of a validation set would increase the confidence that the analysis was done correctly. The Panel also suggested that standard protocols for data analysis be developed when the actual data are released to the public for better transparency of the analytical process.

Question 5. Performance Evaluation of the HTP ER Binding and Transcriptional Activation Assays Against a Set of Reference Chemicals (Section 6)

An important aspect of the Agency's *Proposed Chemical Prioritization Approach for the EDSP is to minimize the occurrence of false negatives and to have confidence that all potentially active chemicals (i.e., those that can trigger the MIE) are identified. Therefore, the performance of each of the computational toxicology components of the prioritization process is being evaluated. The performance of each HTP assay has been evaluated using a set of well-documented reference chemicals that represent a diversity of chemical classes and range of potencies. Additional performance metrics have also been evaluated that take into account a number of parameters that are unique to HTP technology.*

5.1 Please comment on the selection of reference chemicals and whether they are sufficient to assess the performance of each HTP assay for ER agonists. How well do the reference chemicals represent the range of potencies needed to establish the reliability and relevance of these assays for use in EDSP chemical prioritization for the inert ingredients, fragrances, and SDWA chemicals?

Panel Response

The Panel commented on the white paper discussion of HTP assay performance assessment against the selected reference chemicals and classification performance of the assays to predict activity of the AOP in the documents and presentations, and agreed that clarification was needed.

The reference chemicals chosen represent a broad range of ER activity, classified from inactive to weakly active to strongly active. Reference chemicals for test development are as derived for the OECD performance based test guideline for TG 457 (OECD 2012c), and are based on validated data generated in the development of the test guideline. The Panel appreciates that this OECD reference chemical list was derived to have at least 25% negatives, and a range of weak, to stronger potencies, with structural

diversity, for the ER agonist stably transfected transactivation assay (STTA) assay (TG 455, OECD 2012b) and the ER agonist/antagonist BG-1 Luc assay (TG 457, OECD 2012c). Less clear is how well the reference chemicals represent the range of potencies, particularly negatives/inactives, that are needed to establish the reliability and relevance of the HTP assays under discussion, and ES models based on the results from these assays for use in EDSP chemical prioritization for the inert ingredients, fragrances, and SDWA chemicals. The reference chemical list used was not developed for such HTP assays, although the data can be used as a basis to take forward a reference chemical list more specifically designed for the HTP assays. The Panel felt, therefore, that the reference chemical list should be extensively expanded beyond that used for OECD TG 457 (OECD 2012c). This should be conducted by utilizing relevant data from the literature, and in particular, with respect to more inactive chemicals (to give an increase to a least 50% of the reference chemicals). It is also important to conduct a literature search on chemicals and chemical families to identify lead compounds, so that important classes of xenoestrogenic compounds are not missed, *e.g.*, PAH chemicals (see Abdelrahim *et al.* 2006, Lui *et al.* 2006, Kummer *et al.* 2008, Swedenborg *et al.* 2007).

The Panel also noted that the HTP assay performance was evaluated only on a binary scale of "active" or "inactive", which is more relevant for statistical analysis than for describing potency. The Panel believes it would be useful to evaluate assay performance and reproducibility against the "gold" standard chemical and positive control, 17β-estradiol, because this chemical is used to normalize all of the raw data in each plate. Assays that do not reproducibly produce high quality data for the gold standard should be removed rather than simply normalized and compared to other assays. Assay sensitivity should then be evaluated separately for each assay using weakly active to moderately active standards to establish the dynamic range of the assay. In addition, the Panel recommended that there should be an adequately balanced statistical model to calculate sensitivity and specificity.

Statistical evaluation of the performance of the assays requires a different set of "reference" chemicals because the model is evaluating "active" versus "inactive" classes. In this case, size and composition of the "truth set" is insufficient because only eight "inactive" chemicals are included. Additionally, many tested chemicals have substantial literature to support classification as "active" or "inactive", which could be used to increase the size of the truth set for the statistical evaluation of the assays. An equal balance of classes is necessary for accurate estimates of sensitivity and specificity, so that false negatives can be minimized and weakly active chemicals captured.

With respect to the HTP data, insufficient information was provided to explain the basis for the reference chemicals list or how it could be expanded given the total number of "known" activity chemicals in the ToxCast/Tox21 database. The Panel recommended that the reference list should be expanded to over 100 chemicals, including more diverse chemical classes that are known to be ER-alpha active in the literature and chemicals identified on the basis of the Agency's own data produced as a result of the ToxCast/Tox21 research programs.

In summary, the Panel felt there should be a lot of information coming from the HTP assays on inactive chemicals, (and also could be collated from the literature) and this is specifically important for the model building aspects that the HTP data can be utilized for. It would be relevant to conduct a

comparison of the inactive compounds between the ESv1 predictions, and the "inactives" data results generated via HTP screening.

The Panel assumed that the OECD validated ER-alpha antagonist assay available for BG-1 Luc (TG 457) would be used, but this was not specifically denoted in the white paper.

5.2 Please comment on whether the Agency's evaluation of the performance of the HTP ER binding and activation assays has considered and accurately assessed all relevant aspects of the assays. For example, have signal-to-noise ratios, background subtraction and interferences been adequately characterized when describing the strengths and limitations of the assay?

Panel Response

The Panel indicated that while careful analysis had been conducted, the results and explanations for poor data were not always adequate. No documentation was provided regarding the algorithm for dealing with flares. Standard protocol for monitoring assay validation at the contractor sites would be useful and it was unclear if this was going on in practice. The Panel was initially unclear whether and which HTP assays were being run in-house or by contractors. During the meeting, the EPA clarified that contractors were running the HTP assays.

5.3 Based on analyses in sections 5 and 6, please comment on the comparative performance of the 8 HTP assays for detecting ER reference agonists (e.g., do some assays perform differently, and if so to what extent?) Was sufficiently detailed information provided explaining the likely reasons for assay differences, when they are observed (e.g., cutoffs, assay interferences, background differences)?

Panel Response

The Panel stated that the EPA had provided extremely useful information on the comparative performance of the eight HTP assays for detecting ER-alpha reference agonists, but that it was preliminary and far from complete, as recognized and stated in the white paper. The Agency provided detailed information on potential assay differences with respect to cutoffs and testing only a single concentration (25uM) as reported for the ToxCast data. As noted in the Panel's response to charge CQ 4, a full dose response curve for all HTP assays is needed. In some HTP assays, however, a complete dose response curve is not possible, *e.g.,* some transfection assays can be conducted in a full dose response mode, but can provide a partial dose response in some test chemical cases, but they are still useful for screening purposes.

The Agency's evaluation appears to be mostly a statistical approach which did not sufficiently describe the effects of thresholds, limits of detection or sensitivity of each HTP assay to detect the range of estrogen receptor activity for the referenced chemicals.

Alternative options for identification or assessment of assay batteries to detect endocrine disruptors and prioritize testing

Although beyond the scope of CQ 5.3, the Panel had a number of comments and suggestions with respect to the comparative performance of the eight HTP assays for detecting ER reference agonists and improving the selection, development and interpretation of assays.

The Agency's evaluation of differences between the eight HTP assays was based on the assumption that they provided redundant measures of the same biological activity. The Panel indicated that these assays represent complementary, but distinct components of the MIE for the ER-alpha AOP. Some of the remaining 13 assays may represent additional distinct, but related MIE's. When combined, the 21 HTP assays, with multiple modes of action, would improve predictions of "active" and "inactive" compounds. Overall, the Panel indicated that the eight HTP assays chosen represent a distinct, but limited survey of the potential MIEs for ER adverse outcome pathways.

Performance differences are to be expected when looking at different assays. Concordance analyses are also relevant, and are often conducted when comparing *in vitro* ER-alpha mechanistic data with *in vivo* screening data obtained from the uterotrophic assay, (*e.g.,* Wang *et al.* 2012). Therefore, integration of assay performance should be considered for future predictive models.

The Panel noted that observing complete concordance across cell lines, where every compound exhibited comparable induction in two different cell lines, would indicate redundancy. In practice, one would expect differences when testing selective estrogen receptor modulators (SERMs) in different cellular contexts. Therefore, the fact that the HTP assays were not completely concordant across contexts suggests that different assays or combinations of assays might emerge, in the future, as being prognostic. Each cell line used in the assay should be characterized to account for differential binding and activation due to the presence of other relevant receptors, cofactors, and promoters. These other factors might affect the interaction of the chemical with the target MIE and induction of the AOP, *e.g.*, see discussion in the performance based OECD *in vitro* test guideline TG 457 (OECD 2012c). Additional discussion on the importance of cellular context to activity is found in response to CQ 4.

The Panel suggested that systematic integrative aggregate assays should be conducted to assess the overall utility of the assays, *e.g.*, withhold an assay, see how data set performs and then save resources and stop working with poor performers. This sort of approach is used extensively in the pharmaceutical industry and there are some case studies that might be useful models. The Panel also noted that different cellular context can result in performance differences (see earlier discussion in response to CQ 4).

The EU 5th framework program, *Dysregulation of endogenous steroid metabolism potentially alters neuronal and reproductive system development: effects of environmental plasticisers (ENDOMET),* includes a case study for use of the "screening battery of tests" approach. The selected chemicals represented three limited, but different classes, *i.e.*, phthalates, alkyl phenols and adipate classes. Their effects were measured using five different tests selected from 23 *in vitro* tests and analyzed by cluster analysis to identify potential "bottleneck" AOPs for steroid metabolism and function. The analyses

identified four different endocrine disruptor pathways for which the *in vitro* test battery of the five assays could potentially be used as indicators. In this example, critical pathways were shown to include estrogen synthesis and sulfonation, synthesis of "active" sulphate/PAPS (3'-phosphoadenosine-5'-phosphosulphate), and thyroid hormone regulation. The five assays: sulphite oxidase (SOX) activity;

sulfotransferase (SULT)1A1 activity; PAPS synthetase (PAPSS) activity; aromatase; and the thyroid hormone screening assay (T-screen assay)[2] (see discussion of the assays in Waring *et al.* 2012; Sung *et al.* 2012).These five assays utilize human or rat cell lines relevant to assessing endocrine disruption in humans. The initial 23 tests included the thyroid as a target, neuronal cell lines, also tests for steroid synthesis and metabolism as well as the classical nuclear receptors (NRs). While this group of tests still need further optimization and more substance testing (*e.g.*, so far they have only been tested for a range of phthalates, alkyl phenols, adipate and miscellaneous endocrine disruptors (Waring *et al.* 2012; Sung *et al.* 2012), this project demonstrates how cluster and multivariate analyses can be used to interpret the results of test batteries and how it could be used for designing specific test batteries for different chemicals (Jacobs *et al.* 2013).

A systems biology predictive modeling approach may be used to predict target-to-biological effects. As noted above, many AOPs are integrated into complex networks instead of single, isolated linear pathways and these networks may be highly complex. Computational approaches that are now available can predict possible data-driven linkages between molecular and gene expression data with, in this case, observed phenotypes (*e.g.*, Dudley *et al* 2011). Although this particular example is not specific to endocrine pathways or endocrine AOPs, it uses a general computational approach that may be adaptable to endocrine related chemical risk assessment.

Question 6. Analysis of ER Expert Training Set and HTP in vitro Assay Data (Section 7)

To better understand assay responses to a chemical and whether it has the potential to activate the ER-mediated pathway, the Agency compared a set of chemicals evaluated with the ER HTP in vitro assays with data on the same set of chemicals included as part of the in vitro training set data used to build the ER ESv1/2 (ES TrSet).

6.1 Please comment on the Agency's assessment of how HTP in vitro assays performed with respect to expectations of performance for chemicals within chemical categories when compared with the in vitro training set data used to build the ESv1/2.

Panel Response

The Agency's assessment of HTP assay performance was clear and able to provide explanations for most of the instances where assays performed differently from ES TrSet. In general, the largest cause of

[2] The T-Screen is an *in vitro* bioassay used to study interference of compounds with thyroid hormone at the cellular level; thus, bridging the gap between limitations of assays using either isolated molecules (enzymes, transport proteins) or complex *in vivo* experiments with all the complex feedback mechanisms present (Gutleb *et al.* 2005).

discordance between the two data sets was due to the high potency bias of the HTP assay data sets vs. the rainbow trout assay datasets that were optimized for low potency chemicals. The inability of the HTP assay studies discussed in the white paper to detect weak ER ligands/activators is a result of the HTP data having been generated by ToxCast, which focused on efficient identification of high potency chemicals using HTP assays with a lower maximum concentration than used in the TrSet (noted in CQ 4.1.d). As a result, the rainbow trout liver assays are able to detect weak ER ligands/activators (as well as strong ER ligands/activators). As could be expected, a high degree of false negatives were observed in the HTP assay studies, in comparison to the rainbow trout assays. The impact of different concentration ranges tested in HTP vs. TrSet on classification of a chemical as active or not could be minimized, as noted both in the EPA White paper and in public comments by Dr. Geter, Bayer CropScience, by shifting the concentration maximum of the rainbow trout liver assays to that tested in the HTP assays. As expected, this leads to the HTP assays being in greater agreement with the TrSet results. However, this would also lead to a low detection of weak ER ligands/activators in both data sets, and would be contrary to the objectives of good quality data generation.

The Panel also indicated that that expectations that chemicals within categories should all behave similarly is unreasonable as chemicals in categories were compared using the mismatched concentration-range between the rainbow trout *in vitro* assays and the HTP assays. This was highlighted in the white paper (pp. 98-101) in the 2-hydroxybenzophenone chemical class where two of the chemicals were not ER active and two others were active by HTP assays. However, all four chemicals were active in the trout assays done at higher concentrations. This example reinforces previous observations of the Panel and others that it is critical to compare assays across an equivalent concentration range and more effort in the future should be placed in testing chemicals in HTP assays so that the data are compatible with the ES TrSet.

Some panel members noted that the Agency could incorporate existing HTP data, once successfully validated into the current scheme for identification of highly potent ER-alpha activators. Shifting the concentration maximum of the rainbow trout liver assays to that tested in the HTP assays and comparing the results could be used to assess concordance with the HTP data. The ES (all versions) could then be used to prioritize remaining chemicals for further testing. An effective HTP approach would employ both biochemical (*e.g.*, ER-alpha binding and gene activation assays) and a limited number of cell-based assays that would allow identification of active compounds with a low probability of false negatives.

Given the capabilities of HTP assays in general, and access to fully validated HTP assays in the future, a portion of the Panel suggested that the Agency focus on screening as many compounds in the EDSP "universe" that are amenable to HTP testing. These panel members stated that having actual ER binding data (HTP data) rather than just predicted binding data (ES predictions) empirical information should be pursued. The Panel recommended that more effort should be placed in constructing the HTP data set so that it is compatible with the ES TrSet prior to comparison by testing chemicals in HTP assays at higher concentration ranges. However, care must be taken in application of mixed data, since the Agency is looking at the ES as a QSAR and mixing test endpoints (rainbow trout ER binding vs. a variety of ER activation and estrogenic MIE HTP assays) is a violation of QSAR principles as established by the OECD (OECD 2004, 2007). However, data can be mixed if one considers the ES to be a series of structural alerts based on the relative activity for both ER-binding and gene expression. In

order to apply this approach, one should compare results from a series of compounds tested in parallel using the same concentration ranges for both the rainbow trout assays and HTP assays.

6.2 Please comment on EPA's approach to comparing the ES TrSet data with the HTP data and the interpretations regarding discordant results between the two assay approaches. Please recommend other considerations or approaches to conducting this comparative analysis?

Panel Response

The Panel indicated that the Agency's approach is reasonable given the inherent mismatch of the HTP assays vs. ES TrSet assay sensitivities due to differences in the experimental design (*i.e.*, use of different concentration ranges). Identification of protein interference through adsorption of chemicals was a good explanation of some discordant results. Once the concentration range differences are accounted for between the two methods, there appears to be relatively good concordance for the two methods to predict inactive chemicals. However, more comparison between ES TrSet results and HTP results would be useful, since the ESv1 is trained based upon these data. The Panel believes it would be useful to identify whether those chemicals that were "inactive" in the ES TrSet data and "active" in the HTP assay data, were predominantly the result of assay interference (*e.g.*, fluorescence, cytotoxicity or flash sites).

As noted previously in the response to CQ 6.1, ultimately datasets are only comparable over the same concentration range. When compared directly, the concordance of positive results is extremely poor and appears to be the result of two inconsistencies between the methods: (1) use of the single concentration 25µM for the screening of the HTP assays clearly misses the concentrations necessary to measure response of the weakly active chemicals, as seen in the ES TrSet results and (2) the RBA <0.00001% threshold considered as positive by the ESv1 needs to be confirmed as measurable by both Tier1 assays and properly designed (dose range, optimization to detect weakly active chemicals) HTP assays to expect that they will be concordant. If this activity level is not measurable, then the Agency should consider adjusting the threshold accordingly.

Other considerations or approaches to conducting a comparative analysis of the ES TrSet data with the HTP data

The Panel believes that the HTP and rainbow trout assays must be put into context (*e.g.*, isolated receptor vs. transfection expression assay vs. native receptor and native receptor vs. mRNA responses). Given that the assays have different cellular contexts, the Agency should not expect complete agreement within receptor assays, transcriptional activation assays or between either, even if similar concentration ranges are examined.

The comparative analysis may be more efficiently performed using the larger HTP assay data set (*e.g.*, all 21 ER-alpha relevant HTP assays) to compare to TrSet to enable identification of assays/or combinations of assays that capture TrSet results. The use of alternative assays may help reduce uncertainty. This would enable ranking of assays using a prognostic approach where assays are identified that are predictive, but not necessarily directly linked. In this approach the Agency should use

all 21 HTP assays related to the ER-alpha activation AOP. A panel of HTP assays could then be identified that are prognostic of ES TrSet results.

See also response to CQ 5.3.

6.3 Given that there may be varying occurrences of false negatives and false positives based on different sources of training set data, please include in your comments options for addressing these uncertainties for different categories in the context of building ESv3? Note: Agency wanted organization of response into 2 responses. What can Agency do now? What can they do in future?

Panel Response

What can the Agency do now?

The Panel suggests that the Agency examine effects of overfitting/overtraining on data, weighting of data depending on source, reliability of data, number of assays, and different chemical categorization strategies. The inclusion of standard assay methods or guidance to assure that similar concentration ranges are used in different data sets, standard control assays and standard chemicals across all data sets would be useful in integrating future data sets from different sources.

Another approach to be sure that different data sources are compatible or useable would be to make sure data are in open data format so that different platforms can access them (Marshall *et al.* 2012, Sansone *et al.* 2012). The data can be placed in a universally accessible context where the format is usable across time and computing platform, amenable to meta-analysis, incorporates keyword triplets, and permits use of semantic searching to agglomerate different data types (*e.g.,* Hastings *et al.* 2012, Courtot *et al.* 2011, Williams *et al.* 2006, Madin *et al.* 2008, Marshall *et al.* 2012, Sansone *et al.* 2012). This would have the added advantage of providing transparency of the raw and analyzed data sets so that interested stakeholders from industry, academia, government, and public interest groups would be able to access the EPA data and see how it was used in the regulatory decision-making process.

What can the Agency do in the future?

A number of approaches were identified by the Panel that would improve comparisons and prioritization efforts.

HTP assays could effectively be used to categorize chemicals in a tiered manner or through the use of a decision tree. For example, a decision tree could be based on single assays as decision points. Since the filtering process performed on HTP data may eliminate valuable data, the Panel recommended that the Agency consider analyzing results of combined or separate assays with and without filters to minimize false negatives.

Alternatively the Agency could use one assay to decide “yes” (ER-alpha active) or “no” (not active). If not active, the chemical could continue to further tests to refine uncertainty that it is not an ER-alpha active chemical. Since the current assay format identifies high potency ER-alpha activators, further

testing across a larger concentration range could focus on those negative for the assays conducted at low concentration ranges.

Further improvement of comparisons may be achieved by utilizing exposure potential to prioritize chemicals for further testing by a higher concentration range. For example, existing analysis of physico-chemical properties or environmental exposure factors as production volume could be used to identify chemicals that are more likely to come in contact with humans and the environment and these would be higher priority chemicals for testing using more sensitive assays or a higher dose range.

In the future, the Panel suggested that Agency efforts should be focused on using additional HTP assays targeting different effects related to ER-alpha AOP key events to add more context and connectivity to the AOP. Assuming that the artifactual sensitivity differences between TrSet and the HTP data sets are addressed, closer concordance between the TrSet and the HTP data sets may be achieved if the Agency enhanced the ES using HTP data, *e.g.*, categorize compounds based on both HTP activity AND *in vitro* rainbow trout assay data, in addition to physico-chemico properties to form categories of chemicals. Incorporation of HTP data should result in an approach that more truly addresses structure activity relationships through incorporation of more biological activities. Approximately 3,000 chemicals with biological activities in ToxCast are available to incorporate and expand the ES. The Panel noted that care should be taken with assays performed using low concentration ranges. For example, a new category rule, based on HTP data, could be added as a node in the decision tree. This would permit expansion of the existing effects based rule set in ESv2.

If select ER-alpha AOP specific HTP assays are optimized for weakly active chemicals, conducted over a broad range of concentrations, and validated on the small scale, then these could provide critical effects data necessary for prioritization in conjunction with development of ESv3. Of course, this would be dependent upon time, funding, and availability of chemicals. The Panel repeated their concern with the Agency's difficulties in obtaining chemicals (see discussion in CQ 1 and CQ 7.2.3). The Panel believes that approximately 70% of the chemicals in the EDSP universe could be obtained with more focused efforts. ESv3 will be designed to detect ER-alpha activation via binding and transcriptional activation. HTP screening of about 5,000 chemicals is relatively inexpensive and rapid. If the chemicals can be obtained (or synthesized), the Agency should consider testing all chemicals in the space (*e.g.*, the 4,957 inerts and SDWA chemicals) via HTP assays and compare the predictive ability of the HTP assays versus the predictions of ESv2 or v3, as well as examining different combinations of the two. Use of simple, robust HTP assays could potentially be as predictive as ESv3.

Finally, interpretations regarding discordant results highlighted differences in the approaches but did not address a path forward or strategy to fix the problem or validate the HTP assay performance reproducibility and reliability, and, or comparatively, against the results of EDSP Tier 1 assays. The Panel recommended that the Agency establish criteria for how accurately the *in vitro* assay results (ES TrSet and HTP) predict *in vivo* activity that could lead to an adverse outcome.

Question 7. Category-Based Approach and Strategic Testing to Expand the ES applicability domain with additional in vitro data (Section 8)

Section 8 describes a category based strategy (as used to build the original ESv1) to expand coverage of the training set data for the entire EDSP inventory (i.e., the OECD principle that addresses the "domain of application"). The approach described in Section 8 is considered relevant to any in vitro assay data that is applicable for measuring a well defined end point within an AOP under study (in this case, an ER-mediated pathway).

The initial EDSP inventory was based on the 324 antimicrobial active ingredients and the 1536 food use inert ingredients. This ESv1 domain was expanded to an ESv2 domain by strategic testing and assessment of relevant data. The task then became one of designing an approach that would embrace the chemical space of other inventories, including the fragrance use inert ingredients and the Safe Drinking Water Act substances, which expanded the chemicals under consideration to over 10,000, a nearly 5-fold increase.

7.1 Please comment on the adequacy and efficiency of the category based approach to select chemicals for testing to expand the training sets in terms of covering the ~1700 chemicals not covered in the domain of ESv2.

Panel Response

In Section 8 of the white paper, the EPA describes the category-based approach and strategic testing to cover the remaining unknowns after application of ESv2. The strategy for expanding the training set was as follows:

1) consider what chemicals in the new inventories reside outside the tested chemical space as defined by both structural fragments and log Kow ranges,

2) group by extended structural fragments of untested chemicals into common chemical categories; and,

3) then consider relevant experimental data to assist in assessing the new chemical categories (again taking into consideration log Kow ranges).

The intent is to gather data on a few members of each category and log Kow range to extend coverage of the untested space with a minimal amount of chemical testing (*e.g.*, 100 chemicals).

Table 8.1.1 (p. 123) of the white paper summarizes the structural distribution of the 1,704 chemicals of the 5, 975 that were outside of the ESv2 expanded domain and subsequently assigned UnkBP (final decision node VII outcome is "Unknown Binding Potential") using the subpocket theory of ER-alpha binding. This means that it was not known whether or not these particular chemicals could bind in the ER alpha sub-pocket, or not. While in some cases there are specific subgroups in the ESv2 decision tree (Figure 3.2a.1, p. 23, white paper), such as mixed phenols, potential A-site binders, and mixed organics, potential B-site binders, the vast majority of the 1,704 chemicals with unknown binding potential are not

covered in the domain of ESv2. This UnkBP group consists of 82% of the mixed phenols (133 of 161) and 98% of the mixed organics (1368 of 1400) of the mixed organics, which had no available structural information.

Briefly, EPA used structural fragments common to multiple chemicals in the "Mixed Organics" or "Mixed Phenol" chemical groups with unknown binding potential to propose 16 "new chemical groups" for ESv3 expansion (Figure 8.1.1, p.128, white paper). Test information available from the ToxCast or Tox21 ER-alpha binding related HTP assays were linked to each new group of sub-categories defined by structural complexity and physico-chemical property cutoffs and assessed for adequacy in covering the group. If the coverage was deemed inadequate, new tests were proposed. The poorly covered sub-categories would be subjected to targeted-testing in an iterative format, and subsequently, the test results would be used to further refine category definitions. In theory, examining new chemical categories is an adequate and efficient approach to expand the training sets (with caveats as noted below) particularly where it is possible to clearly define new chemical classes, as with those proposed for ESv3 (*e.g.*, Figure 8.1.1 , p. 128, white paper).

The Panel considered it important to avoid very conservative definitions of categories (*e.g.*, including only one or two chemicals) that do not allow for reduced testing. Similarly, very broadly-based groups (*i.e.*, those defined by simple structural alerts such as "cOH" for phenols) will be of limited value due to the likelihood of a large number of false positives. It will be important then, to assess the tested or screened domain in relation to the possible structural space of the category.

For example, within the fragrance use inert ingredients, the new chemical category pyridines and quinolines (see Figure 8.1.1, p. 128, white paper) would include pyridine [CAS 110-86-1], quinoline [CAS 91-22-5], isoquinoline [CAS 119-65-3], and 4-phenylpyridine [CAS 939-23-1], all of which have been tested and shown to be non-active in the yeast estrogen assay. Since the fragrance-related pyridines and quinolines are mainly alkyl-substituted, the above noted data may be sufficient to cover fragrance related N-heterocyclic compounds like 2-pentylpyridine [CAS 2294-76-0], 2,6-dimethylpyridine [CAS 108-48-5] and 6-isopropyl quinoline [CAS 135-79-5]. The test data may even be sufficient to cover other alkyl-containing N-heterocycles like pyrazines {N1CCNCC1} or pyrroles {C1=CC=CN1}. However, there will be less confidence that other fragrance-related pyridines and quinolines (*e.g.*, 3-acetylpyridine [CAS 350-03-8]) are within the new category. More interestingly, testing with the yeast assay has shown 6-hydroxyquinoline [CAS 580-16-5], 7-hydroxyisoquinoline [CAS 7651-83-4], and 8-hydroxyquinoline [CAS 148-24-3] to be active, but 2-hydroxyquinoline [CAS 59-31-4] and 3-hydroxyisoquinoline [CAS 7651-81-2] to be inactive.

Uncertainty associated with classifying the universe of chemicals into smaller groups as a basis to select "representative" chemicals for a training set should be explicitly recognized. Characterization based on structure and log Kow is necessarily subjective, identifying groupings that may be unrelated to ER-alpha effects (*e.g.*, low affinity binders). Similarly, selection of representative chemicals from within the upper and lower bounds of a class, *e.g.*, characterized on the basis of log Kow, may result in mixed responses or missing the effects of intermediate chemicals.

In addition, it was unclear to what extent the predicted binders were validated so confidence can be derived as more than just "active" or "inactive" using measurable thresholds. Specifically the *in vitro* assays should be verified to have sufficient sensitivity to measure the RBA < 0.00001% of 17β-estradiol activity threshold.

Also, for compounds remaining outside of clearly defined groups or HTP data, *e.g.*, the 576 ungrouped UnkBP MO & MP compounds in Table 8.1.5 (white paper, p. 132), category-based strategies may be inadequate. The Panel listed several examples from Appendix B of the white paper where EPA remarked on this deficiency.

1) p. 20. EPA notes that "inventories of concern contain diverse structures with widely varying physical-chemical properties," that may confound straightforward categorization.
2) p. 29. For ESv1, EPA observed that, "there are not sufficient data at this time to further categorize [MO & MP compounds] into subgroups which bind ER and those which do not, or to hypothesize structural parameters associated with the activity of those compounds."
3) p. 37, line 743. Within such categories, EPA remarked that ESv2 saw "only modest increases," such as a 3% increase for fragrances.
4) p. 125, line 1961. ESv3 may be expected to increase the applicability domain coverage for many UnkBP chemicals (of which 88% are MO & MPs), but 576 ungrouped UnkBP MO and MP chemicals will still remain (34%).

The Panel suggested that EPA consider a broader range of options for selecting and prioritizing derivatives for assessing new categories, especially as molecular complexity increases.

Second pass computational approaches. One option may be the incorporation of supporting or "second pass" chemical informatics and systems pharmacology computational approaches, particularly for recalcitrant and diverse UnkBP compounds. This would be consistent with the 2009 SAP recommendations to consider SMILES, SMARTS, or structural-similarity approaches (Appendix C, white paper). Relevant tools include Naïve Bayesian classifiers (Besnard *et al.* 2012, Paolini *et al.* 2006, Nidhi *et al.* 2006; Rogers *et al.* 2005), statistical similarity methods such as the Similarity Ensemble Approach (Keiser *et al.* 2009; Lounkine *et al.* 2012) or others (Campillos *et al.* 2008, Vidal *et al.* 2011, see also response to CQ 3) and stacked neural network classifiers (Hinton and Salakhutdinov 2006, Hinton *et al.* 2006; see also their recent use in winning the November 2012 Merck competition for drug-activity prediction: http://blog.kaggle.com/2012/11/01/deep-learning-how-i-did-it-merck-1st-place-interview). The second two approaches, in particular, are less susceptible to the QSAR limitations (described on p. 24, Appendix B). It is recommended then to apply unsupervised grouping (especially on the remaining 576 UnkBP), followed by other predictive methods on the groups. This is an ideal test case for these new predictive methods.

Opportunistic HTP prescreening. Consideration of immediate expansion of HTP screening to address the entire chemical space prior to ES analysis was also suggested as a basis to complement the categories training data set approach, unless precluded by constraints such as time and resources. In this scenario, only chemicals that cannot be obtained would need expert analysis and hence would not be available for a training set.

The Panel recommended that the Agency better articulate the strategy and rationale in defining the categories and selecting derivatives to assess them *a priori* before any additional development of categories and derivatives. These include criteria for both definition (extending beyond structural to a range of relevant physico-chemical properties) and appropriate bounding based on incorporation of a broader range of computational tools and HTP testing assays, for which *in vitro* assays could be verified to have sufficient sensitivity to measure the RBA <0.00001% of 17β-estradiol activity threshold.

7.2 Based on the HTP assay performance (analysis of reference chemicals and comparative analysis with the ES TrSet), please comment on the adequacy of the HTP data for advancing the ER expert system's rules to cover the additional groups of chemicals in ESv3.

Panel Response

One of the strengths of integrated testing and assessment strategies is the breadth of information that can be incorporated often from *in vitro* and HTP methods (See also CQ 3.2). These assays are normally designed to provide data about a single biological property or event, which can be interpreted in the context of in vivo toxicity through consideration of the adverse outcome pathway.

Expected greater variability in results from *in vitro* methods, as compared to those *in vivo*, is a function generally of results of the former being associated largely with a single factor or limited set of factors. As a result, statistical-based acceptance of an *in vitro* method (*e.g.*, r^2= 0.88) is less important than the mechanistic relevance of each *in vitro* method to a given key event and the relevance of the key event to the final adverse effect in question in the consideration of AOPs. All of the HTP methods proposed for use by the Agency are standardized well-established assays that typically quantify ER-alpha binding and/or gene activation with a reporter system (see white paper, Appendix F). Therefore, all are relevant to ER-alpha binding and gene activation which were used in the assessment of ESv1 and ESv2.

Variability in results from different *in vitro* methods is largely a function of aspects of protocol including controllable abiotic factors such as solubility and volatility, and biological aspects such as cytotoxicity, cellular context and characterization (See also CQ 4.1) and metabolic capability, which are more difficult, but possible to standardize.

The weight-of-evidence for considering results from any mechanistic, transparent, integrated approach, is typically assessed by examining several factors, including the:

- the key events that comprise the scheme,
- Relevance of each key event to the final adverse effect in question,
- Number of methods used to assess each key event,
- Relevance of each method to a given key event, and
- Reliability of the data measured or otherwise determined for each method.

The Agency in their presentations noted the OECD QSAR principles (*e.g.*, see white paper, Section 3, p. 15) were applied to the development of the ER-alpha binding ES (see Figure 3.1.2, p. 16, white paper). Yet in Section 7 (white paper), the Agency proposed to expand their ES by using information

from HTP screening assays. On the surface this appears to be a violation of the OECD principle of having "a well-defined endpoint". While it is certainly true that the binding assay utilizing rainbow trout liver cytosol ER-alpha (*i.e,*. rtER binding assay) is not the same as any of the three HTP Novascreen Biosciences Corporation (NVS) ER-alpha binding assays discussed in section 7 (white paper), in the context of the AOP used to provide the mechanistic understanding for the ES (see Figure 3.1.1, p. 16, white paper), the four ER-alpha binding assays provide similar data (RBA measured by displacement of [^{3}H]-E2 from the ER-alpha receptor for the same key event).

The Panel noted that in the context of an ES where "relative" measurements are used, mixing data from more than one assay/protocol is acceptable. This type of argument can be used to rationalize the mixing of results from different gene expression assays (*e.g.,* rtVtg GA, BG1-Luc and Tox21 BLA). Again relative, rather than absolute expression will need to be examined.

The HTP assay data may be useful, then, to cover additional groups of chemicals in ESv3, when used in full dose-response format. Single-concentration data appear to be inadequate, as in the NVS HTP assays. The white paper states, "the number of prescreen negatives that would have been binders if run in full dose-response mode would argue that the prescreening procedure be reassessed for improvements or that running in a full dose-response mode would be more appropriate for a prioritization approach" (white paper, p. 121).

The HTP assay sensitivity for detecting weakly active reference chemicals also remains to be fully evaluated and as a result does not advance the ER-alpha ES predictions for weakly active chemicals. However, missed "active" chemicals (false negatives from ESv2) should be used to evaluate limitations of the current ES. Are there common points of departure in the rule-based ES that do not apply for these missed chemicals? What are the commonalities and differences between these points of 'breakdown', and how do they compare between the missed chemicals and their classes?

The HTP data should also be weighted, taking into account that where HTP data are included with results from the TrSet assays, statistical power grows but each observation becomes less certain. As a result, quantification of the retrospective predictive sensitivity and specificity of a particular HTP assay should be explicitly considered, by applying a corresponding scalar weighting or likelihood score to any ES category or decision rule that was derived or partially derived from HTP data. It was also noted that broader integration of data from better targeted, designed for purpose HTP testing would likely significantly enhance the ERES.

The Panel also suggested the following exploratory techniques to integrate ESv3 with HTP data. The presented HTP data (where sufficiently available) might be integrated with the proposed ESv3 in two main ways: Either as a *parallel input* alongside the ES, where both contribute to a single overarching prioritization score (section 7.2.1), or as an *overlay* to the ES itself (section 7.2.2). The approaches are not necessarily mutually exclusive.

7.2.1. Consensus method: Using ES and HTP as parallel inputs. As expressed by EPA during a request for clarification, the ES and HTP methods may both be thought of as models: the first as a "virtual" (or computational) model, and the second as a "physical" model. Both, as yet, are only first approximations of the full ER-alpha AOP, and indeed only approximate the ER binding MIE that the

two TrSet assays are meant to address. Thus they can be used to prioritize TrSet-level testing, as TrSet testing is used to prioritize Tier-1 EDSP testing. Currently only the ES, the "virtual" model, is used to prioritize TrSet testing.

Since each model relies on different sets of assumptions, their errors may be less correlated than would be expected for two models of the same type. To illustrate, examples of "physical" assumptions leading to error might be the choice of particular binding assay parameters and test system design for the "physical" HTP model, whereas "virtual" assumptions such as the particular expert's organization of chemicals into categories based on structure or possibly imperfect calculation of physical properties may introduce error into the "virtual" ES model.

Thus, ES prediction and HTP screening could be run in parallel, and the results of each combined into a single higher-confidence decision metric. Each model, be it a "physical" single HTP assay, or a "virtual" model such as the ESv3 or another computational method, should then be assigned a weighting or confidence score. HTP assays with high retrospective performance would have a higher weighting, and thus contribute more, to the single final prioritization score, than assays with lower retrospective performance. One accepted metric for retrospective performance might be a receiver operating characteristic (ROC) area-under-the-curve (AUC); another might simply instead be sensitivity, if this is the primary programmatic concern. In this way, ESv3 and HTP may be quantifiably combined in parallel into a single higher-confidence prioritization score that leverages both approaches.

7.2.2. Overlay method. ESv3 relies on the molecular structures and known or calculated physical properties (*e.g.,* log Kow) of inventory or new chemicals to group them into categories. Representative chemicals within each category have been, or will be, subjected to ES TrSet testing to assign each category a positive or negative prioritization value overall. Thus the ES's rule sets (*i.e.*, "diamonds" in ES schematic diagram) operate primarily over physical properties or structural patterns, to channel a chemical into a single chemical category. This is a strength of the ES rule sets because the system can assign a new chemical to an existing category (or categorize it instead as UnkBP) solely on the basis of its structure.

In response to CQ 7.1, the Panel suggested immediately expanding at least some subset of HTP testing to the full chemical inventory. If feasible, this would generate a new property that could also be exploited to build new ES rules, namely, the chemical's observed "HTP-fingerprint". There are multiple ways to express such a fingerprint; one example is as an "affinity fingerprint" (Bender *et al.* 2006) that would here derive from the observed HTP affinities instead of the Bayesian models used in the cited reference. The motivation is that no one HTP assay is infallible, especially for detection of weak binders, but when multiple HTP assays are considered together, a weak binder may be identified, especially when this information is incorporated in the ES decision logic.

Much as a compound's topological structure may be expressed as a "fingerprint" or "molecular descriptor", and such fingerprints are commonly used to compare molecules, an HTP affinity "fingerprint" could be used in the classification rule sets of the ES's decision tree—if such HTP data are available. Using the HTP affinity "fingerprint" concept fully, one may even calculate Tanimoto coefficients and use these within the ES decision rules. This would enable chemical groupings (categories) and decision rules within the ES that derive from and leverage the "effect space" properties

of a compound. In this usage, "effect space" is used to mean the "HTP affinity profile" of a compound. Thus, rather than the proposed HTP-ESv3 rule sets being comprised solely of "contains a particular chemical group", the rule sets in HTP-ESv3 might be extended to "contains a particular chemical group AND is active in particular HTP assays". As a result, the Panel indicated that a HTP-ESv3 would extend chemical activity prediction beyond a traditional QSAR. Since HTP data may be available for only a subset of chemicals, HTP-fingerprints could be applied solely in a separate ES *overlay* or as a parallel "HTP-ESv3". This would preserve the capability of ESv3 to make ER-binding predictions on chemicals for which no HTP data are yet present.

7.2.3. Problems with obtaining compounds (HTP assays). The Panel was concerned that ~5,000 compounds in the library were not available for testing. The Panel recommended that the Agency put significant priority on obtaining samples of these compounds since they are essential for this initial prioritization effort as well as extension of this strategy to MIEs in other endocrine disrupting AOPs. Difficulty in accessing certain compounds may be related to the fact that the chemical is not in current use. This could easily be addressed by having aspects of potential for exposure, *i.e.* information on production and use, integrated into the *in silico* screening protocol prior to the ES. As mentioned in CQ 1, the Agency should have a "roadmap" on how various components of exposure and hazard will be integrated to prioritize chemicals for screening and assessment. A significant opportunity to meaningfully reduce the numbers of chemicals that need to be tested may be missed because simple elements of exposure are not being addressed up front in the prioritization process. The Agency should clearly state the rationale for acquiring chemicals and why compounds potentially not produced or used should be screened in the EDSP. The Agency could also consider contract synthesis as an approach to obtain chemicals difficult to acquire, or not currently produced, but anticipated to be produced in the future. The ES could be useful in prioritization of compounds for acquisition (perhaps by contract synthesis). ES analysis could then be more valuable in prioritizing those chemicals that should be synthesized for HTP analysis.

7.3 To the extent there are differences between the ES TrSet and HTP data in detecting the ability of chemicals to initiate the ER AOP, especially in the context of minimizing "false negatives" for low potency compounds, please comment on the strengths and limitations of combining data from the different assays to generate training set data for building the structure based rules within the ESv3.What are the strengths and limitations of using training set data from assays that have the same or similar degree of sensitivity?

Panel Response

When using alternative methods to assess chemicals for toxicity, the issue of false negatives is always of concern because of the consequences of being wrong. An examination of the data in Table 6.2.1 (white paper) reveals false negatives are more likely to be an issue for chemicals that are "weak" or extremely weak" binders. Using the RBA of 0.00001% as the cutoff value appears to minimize the false negatives to an acceptable level.

The Panel cautioned against simply combining the HTP receptor binding and transcription activation data with the ES TrSet data. While each set of assays may be equally sensitive to a ER-alpha activating

chemical, the current data can only be combined if the different concentration ranges tested by HTP assays (25 μM for initial screening and 100 μM for transcriptional activation assays) is taken into account. The HTP data appear to have a high false positive rate but also a very low false negative rate when interference thresholds are ignored. In order to capture the chemicals with low potency, the filtering thresholds should be re-evaluated and possibly eliminated if minimizing false negatives is the goal. A combination of assays should be explored as opposed to individual assay performance. In addition, strategies to augment the HTP ER-alpha binding data with at least a single higher concentration for all negative chemicals should be explored in order to synchronize with the higher concentrations used for the ES TrSet (see also CQ 6). Alternatively, current HTP data could be used to identify potent ER-alpha active chemicals. Chemicals that are negative at the low concentration range could be tested in a second round of HTP assays at higher concentration range, similar to that found in the trout liver assays. This would eliminate the problematic identification of weak ER-alpha activators in the ES TrSet, but not in the HTP assays.

Data from different assays with the same or similar degrees of sensitivity can be combined to generate the ESv3 training set data. However these need to assess chemicals within a comparable concentration range. One approach is to prioritize a compound for testing via the full ES Tr Set rainbow trout ER-alpha binding and Vtg gene expression liver slice assays (as "gold standards"), if that compound was first found to be positive in any two of the HTP assays to which it was subjected. Essentially "reliable" HTP hits could be used, where available, to allocate chemicals to the top of a "to test with the TrSet assays" list. A weakness is that this may increase the false positive rate for the initial prioritization step, but given the key programmatic goal of minimizing "false negatives" (white paper, p. 120, line 1847), it seems acceptable. Depending on testing resources and encountered false positive rates, the weakness might also be mitigated by adopting instead a "consensus-based" method, wherein a weighted average across all assay positives contributes to a single aggregate score by which further testing is prioritized. Considering the careful work that has been done to build the ESv1 and ESv2, it would seem sensible that ESv3 TrSet compounds be subject to confirmatory rtER and GA liver slice assay for all positives or borderline cases.

More assays covering other key events in the ER-alpha AOP would likely increase confidence that low affinity binders were not excluded. The inclusion of other assays or increased dose range of existing assays would help address low affinity binding chemicals. Additionally, assay specific weighting of data could also be used to inform prioritization and address potency issues. This is also discussed earlier in CQs 3 and 5.

Question 8. In Vitro Testing and Computer Based Simulations: Addressing Active Metabolites

(Note: This question is looking toward the future and there is no proposal presented in the white paper. Discussion of the issue of active metabolites in examples are presented in Section 7.4.3)

In the context of building a category rule-based approach, it is important to understand if the biological activity is attributed to the parent compound or a metabolite. It is also important to understand if a metabolite of an "inactive" parent compound may be able to trigger the ER MIE. The Agency *would like to take advantage of the growing knowledge of assays systems as well as existing*

metabolic profiles, computation techniques for predicting metabolites and understanding of ER-binding structure activity relationships to address this area of uncertainty.

8.1 Please provide any initial thoughts on how varying empirical and/or computational techniques could be employed to account for metabolites that may trigger the ER MIE.

Panel Response

The Panel acknowledged that the toxicity of chemicals which directly elicit a specific MIE (*e.g.*, ER-alpha binding) is more easily explained than those that are mediated by metabolism. Thus, if the Agency wants to increase the confidence that it has identified the complete universe of chemicals (including daughter products) requiring prioritization for further testing and evaluation, it will need to improve the capabilities of *in silico* metabolic simulations in addition to using HTP assays incorporating metabolism. The Panel discussed computational techniques that might be useful in predicting metabolites.

The Panel noted that new capabilities in the next generation of metabolic simulators, such as the ability to identify the most likely/probable metabolite(s) based on the structure of the parent compound, could assist the EDSP, especially with respect to the ER-alpha binding ES. This would be an improvement over the list of possible metabolites generated by current simulators (*e.g.*, the OECD Toolbox). Further improvements in metabolic simulators might include designing simulators that predict the percent, *e.g.*, 90%, 75% 50% 25% 10%) of parent compound that is transformed to the most likely metabolite(s) (*e.g.,* top three metabolites), and designing simulators that predict the rate of metabolism, *e.g.,* extremely rapid (sec to min), rapid (less that 1 hr), slow (hours) and extremely slow (days) from structure would be tool that may assist the Agency, especially in relationship to the exposure filter (see Figure 2.2.1, white paper). The prediction of metabolites has been demonstrated to be better if several methods are combined. For instance a recent study (Piechota *et al.*, 2013) has demonstrated that combining predictions from a rules-based system (*e.g.,* Meteor) with systems (*e.g.,* MetaPrint2D-React, SMARTCyp) that predict sites of action provide a better coverage of predicted metabolites.

The Panel suggests that a strategy for using predicted stable metabolites be investigated. There are at least two possible strategies. The first is to include the predicted metabolites within the categories. The second is a cyclical process where metabolites would be predicted and then run through the ES. Each strategy has its advantages; however, the second strategy may be more practical based on current ES. Additionally, the Panel indicated that predictions of reactivity (*e.g.*, half life in water) in the exposure filter may also serve as early indication that metabolism is likely to be of concern. Other rule-based or computational chemistry approaches may be useful in identifying reactivity and metabolic potential. These characteristics could be used to identify chemical candidates with which to empirically examine metabolism effects on potency.

Although not specific to endocrine disruption, the Panel pointed to an EFSA opinion (EFSA 2012) on the prediction of pesticide metabolite toxicity that may be relevant to the Agency's applicability of QSAR tools, combining grouping, and read-across approaches in the evaluation of developmental (and neurotoxic) effects of pesticide metabolites. Briefly, QSAR tools alone appeared insufficiently reliable to predict developmental effects, due to their low sensitivity and specificity, but a stepwise approach

involving QSAR analysis and read-across, resulted in an improvement in the identification of potential developmental toxicants (Cronin and Worth 2008, EFSA 2012). The opinion concludes that predictivity for neurotoxicity of the QSAR models tested alone or in combination is currently inadequate to be applied for pesticide metabolites. The results of the QSAR projects allowed the EFSA Panel on Plant Protection Products and their Residues (PPR) to propose the application of computational methods, involving the separate or sequential use of QSAR and read-across in the prediction of developmental toxicity, (and genotoxicity) to complement the Threshold of Toxicological Concern (TTC) approach in the assessment scheme for pesticide metabolite exposure.

More specific to the ER-alpha ES, the Panel noted the following, while ER-alpha binding must take into consideration the "dynamic and plastic character of the ER" as described in the sub-pocket theory, the energetic and steric constraints determined by the ER-alpha defines the domain of chemical structures that can bind to the receptor. Based on empirical data for chemical interactions with the ER-alpha, the structures of chemicals that can bind to the ER-alpha appear, with noted exceptions, *e.g.*, fluoro-substituted compounds, to be limited to cyclic compounds. Moreover, only two sub-pockets, the A- and B-site, are polar in character, and only a limited number of hydrogen-bonding groups (*e.g.*, OH and NH2 moieties) interact at these sub-pockets.

The Panel noted this information may be useful for selecting empirical metabolic pathway data (*e.g.*, oxidation of cyclic compounds) for consideration to identify pro-ER-alpha ligands. The limitations noted in the previous paragraph can also be used to design targeted *in silico* metabolic simulators. For example, by first restricting simulation to cyclic compounds with an unencumbered hydrogen-substituted carbon atom in a 5- or 6-membered ring system and then restricting the reactions to those which only lead to selected polar moieties, it is likely that simulation predictions can be refined to lead to identification of metabolites of high concern for the ER-alpha binding ES as defined within the context of the AOP. The Panel acknowledges that the approach outlined above may be less applicable for the 'special case" bindings where the mechanistic rationale for ER-alpha binding is less apparent.

The Panel further noted as a bench-level *in vitro* assay, the rainbow trout tissue slice GA assay was employed within the AOP to confirm ER-alpha-mediated gene activation of potential ER-alpha ligands and established a biologically and toxicologically relevant endpoint at the tissue/organ level. Since rainbow trout liver slices are considered a metabolically-competent system and the liver slice assay has been shown to respond positively to appropriate parent chemicals or their metabolites, examination of chemicals which are negative in the rtER binding assay and positive in the liver slice gene expression assay provides a means of identifying chemical structures where metabolism is likely to be of concern. The negative/positive response sequence provides a hypothesis-based testing profile that can be used to validate metabolic simulator-based predictions.

Other HTP cell-based systems incorporating metabolism (*e.g.*, ATG assays) could be paired with similar endpoint assays (*e.g.*, transcriptional activation of ER-alpha targeted genes to empirically examine metabolic activation. Potential areas include co-cultivation of metabolically active cells with reporter assay cells, performing assays in a cellular context with a characterized metabolic potential, or initial treatment of chemicals with S9 fractions. The S9 fraction is a supernatant fraction obtained from liver homogenate that contains cytochrome P450 enzymes (phase I metabolism) and other enzymes such

as transferases (phase II metabolism). If achievable in a HTP format, a few different metabolically active HTP assays could provide a range of potential metabolic conditions that would provide increased confidence that potentially ER-alpha active metabolites are or are not formed by the chemicals in question. Such metabolically activated HTP results could be paired with analytic confirmation to confirm loss of parent compound and formation of daughter chemicals and with computational analysis of reactivity/ metabolism of the parent compound to further reduce uncertainty as to effects caused by the parent vs. daughter chemicals.

Some Panel members raised concern that the rainbow trout ER-alpha AOP-based prediction with metabolic simulation may not be directly applicable to mammals where metabolite stability and bioavailability may be more important. Briefly, the different forms of cytochromes P450 will vary among species so metabolism in a system representing one species may not be applicable to other species. Additionally different species may have different capacities to inactivate an active metabolite, such as by differences in Phase 2 reaction capabilities. The stability and the bioavailability of an active metabolite will also be a factor to consider among species. For example, an active metabolite formed in the most likely bioactivation in vertebrates, the liver, will be formed in relatively close proximity to the site of action in a fish where the liver is the site of vitellogenin synthesis, whereas in a mammal the metabolite would need to exit the liver, be transported in the blood where it might bind to plasma proteins and then enter a target tissue, so the effective concentration in the target organ is likely to be less than the concentration produced in the liver. In addition some of the xenobiotic metabolizing enzymes are relatively inefficient, so the kinetics of bioactivation to the active metabolite and subsequent detoxification of the active metabolite will need to be considered in understanding the significance of metabolism and in predicting biologically effective concentrations of an active metabolite

Criteria for interpreting results from combined computational and empirical (*in vitro)* testing are needed. Such criteria are discussed in a forthcoming publication from members of the OECD Validation Management Group Non-Animal (VMG-NA). (Jacobs *et al.* 2013, in press). Use of principal component, multivariate and cluster analyses is useful for study designs that can optimize test batteries addressing metabolism. For example, meta-analyses of gene expression of micro array data sets, such as that used for 17β-estradiol (E2) in MCF-7 cells (Ochsner *et al*., 2009), combined with a characterisation of the metabolic capacity of MCF-7 cells, and the integration of such genome experiments (Lachmann *et al.* 2010) for other cell lines and reference chemicals, could aid the iterative improvement of transcription factor regulation data and metabolism inferences for the *in vitro* tests.

A freely available reference collection of gene-expression profiles from cultured human MCF-7 cells, the "Connectivity Map" (http://www.broadinstitute.org/cmap/), contains data for over 1500 substances, some of which may be metabolites, and includes pattern-matching software for data mining. It was demonstrated that the map can identify substances sharing a mechanism of action in MCF-7 cells (Lamb *et al.* 2006). Profiling new compounds and subsequent pattern-matching can assist in the prioritization of chemicals for further hazard testing.

Question 9. Use of the AOP and Category-Based Testing Strategies for Other Endocrine Pathways (Chapter 9)

The Agency plans to implement an EDSP chemical prioritization scheme that will use both exposure and effect based metrics. The current paper focused on development of an effect-based approach using the ER pathway and relevant in vitro assays. This prioritization methodology draws on several fundamental concepts generally applicable to the development of computational methods. These concepts include understanding inherent chemical properties as well as chemical structure and bioactivity relationships, use of the AOP concept and category-based read across methods to guide the development and application of in vitro testing using lower or higher throughput methods and computer-based simulations.

Panel Response

General comments

The Panel supported the use of the ER-alpha-mediated fish reproductive impairment AOP as described and discussed during this SAP to pre-screen inventories and prioritize chemicals for further testing can be used as a roadmap for other MIEs related to endocrine disruption.

The Panel noted that when the Agency is developing additional Expert Systems to model a specific MIE/AOP, there Agency needs to be a transparent scientific rationale for why the particular MIE/AOP was selected as the basis for the pre-screening and prioritization scheme. This explanation will need to include details on why selected intermediate events are key events and what assays/endpoints can be used in category formation. Especially essential to building the ES will be providing enough detail in the AOP on species of concern (*e.g.*, human, rat, mouse, fish, frog) and assay models (*e.g.*, *in chemico, in vitro, ex vivo*) such that it is transparent what the relationship of each assay/endpoint is to the AOP and why it is scientifically feasible and defensible to use that assay or battery of assays to generate data used in the ES.

Some panel members believe this could also be a useful tool in highlighting approaches for weighting assay and data. For example, assays more closely related to and having a clearer connection to an apical effect would have more significance than very early measures (*e.g.*, receptor binding *in vitro* < (less significant than) ER specific transcriptional activation in a cell <transcriptional activation in a tissue <acute test *in vivo*).

The Panel emphasized that one challenge for moving forward to other MIE/AOP based ESs is the choice of the MIE(s). As further noted by the Panel, a second challenge is identification of true-positive chemicals for system development and validation. One Panel member noted that a quick survey of the current ChEMBL-14 database (https://www.ebi.ac.uk/chembldb), a database derived from the scientific literature, demonstrates more information is available for ER-alpha-mediated effects than for any other receptor-mediated effects (**Table 3**) including those mediated by the androgen receptor (AR) and

thyroid-hormone receptor (THRa and b) and even more so for the xenobiotic and steroid nuclear receptors, *e.g.*, pregnane X receptor (PXR), and the constitutive androstane receptor (CAR):

Table 3. Comparison of receptor-mediated bioactivities (data ChEMBL-14 database)

Receptor	Number of Bioactivities[1]	Number of Compounds
ER	12,000	5,857
AR:	5,700	2,527
THRa	641	411
THRb:	7,511	6,032
PXR:	876	710
CAR:	109	42

[1]The Panel notes that the bioactivities from this database have not all been curated by experts in the field, and in some cases may be discarded during such an expert review process. The counts shown above, therefore, are simply to illustrate the overall trend. In moving forward, the Panel encourages the Agency to use publically available data to support their choice of MIE/AOP for other endocrine disruptors.

9.1 Please comment on whether the principles and concepts used to develop the ER expert system are generally applicable for any category-based prioritization system for other molecular initiating events (e.g., other AOPs for perturbing estrogen, androgen receptor, and thyroid hormone systems)?

Panel Response

The development process of the ER-alpha prioritization model has highlighted the need for iterative steps of model-building, experimental testing, and model extension/refinement (*e.g.*, see white paper, p. 43, line 808; p. 125, line 1964; p. 134, line 2095). The growth of the ES from ESv1 to ESv2 is an example of this process, as is the proposal for ESv3. Going forward, model building for the androgen pathway should anticipate the need to build iterative datasets and models, no matter how well the initial models and datasets are constructed, and incorporate a rapid-turnaround version of this process into its design, such that each step of the iteration need not necessarily wait on explicit review by the Panel once a satisfactory protocol for model development and refinement has been put in place.

As noted earlier, the Panel agreed that the principles and concepts used to develop the ER-alpha AOP ES are generally applicable to formulating category-based prioritization schemes for other MIEs leading to endocrine disruption-related final endpoints. However, a number of issues relating to uncertainty must still be overcome. A key limitation of the system is its dependence on chemical "exemplars" for the AOP, especially the MIE.

The Panel believes there is growing agreement on what issues reduce uncertainty in any ES used to pre-screen inventories. These issues include:

- Transparency, mechanistic plausibility/probability,
- Weight-of-evidence, and,
- Ability to do hypothesis-based testing, especially with rapid and inexpensive methods.

The mechanistic-basis for justifying testing and assessment schemes is best provided by well-documented AOPs. An AOP is a framework comprised of the events at the different levels of biological organization and other key dimensions (*e.g.*, gender, life stage, *etc*.) and their causal relationship to the final adverse effect under consideration.

Critical issues in developing useful AOPs include:

- Separating events that control toxic potency from events that report symptoms,
- Attaining acceptance of the key events included in the AOP,
- Attaining agreement on the method(s) used to assess these key events, and,
- Attaining agreement on the *in vivo* data used as the standard for verifying methods/data for key events.

The importance of weight-of-evidence is directed toward accepting the results based on a mechanistically-transparent AOP. The weight-of-evidence is typically assessed by examining several factors, including the:

- Key events that comprise the AOP,
- Relevance of each key event to the final adverse effect in question,
- Number of methods used to assess each key event,
- Relevance of each method to a given key event, and,
- Reliability of the data measured or otherwise determined for each method.

The Panel believes that attention to these issues during the development and justification of any new ES will aid the Agency in their efforts to garner support for new system. Again it is worth noting that a significant number of Panel members supported the option of first just screening with HTP for activity, especially with an assay(s) closely linked to the MIE of interest. They noted that as long as there is a low false negative rate, the inexpensive nature and rapidity of HTP offset any high false positive rate.

9.2 What lessons have been learned from the development of the ER focused prioritization model that will facilitate more efficient and effective development of an effect-based prioritization model for the androgen pathway.

Panel Response

The primary AR pathway leading to endocrine disruption will be similar to the ER-alpha pathway, in that the seminal event will be receptor binding, though be it antagonist versus agonist in activity, and the downstream events, especially at higher levels of biological organization will be largely a cascade of events that do little to mediate the final outcome. If this assumption is true, the scheme developed for the ER-alpha pathway is directly applicable to the AR-pathway. For example, what has been learned from work on the ER-alpha pathway and the associated chemical groups defined by the knowledge-based ES can be applied to the AR-mediated fish growth impairment AOP developed by Volz *et al*. (2011) and shown in **Figure 3**. Specifically, this includes:

1) The importance of having a well-developed AOP.

2) The importance of being able to demonstrate the typical responses observed at various levels of biological organization (*e.g.*, AR-binding, alterations in growth of most susceptible cell types, malformations/dysfunction in the most likely organs/organ systems *etc*.) in following the AOP.

3) The importance of having developed and standardized assays (*in vitro* and/or *in vivo*) from which data can be gathered in a systematic manner.

4) The importance of being able to apply mechanistic information and data to define chemical categories that can be developed into a structure-based ES which, in turn, be can be validated with hypothesis-based tested.

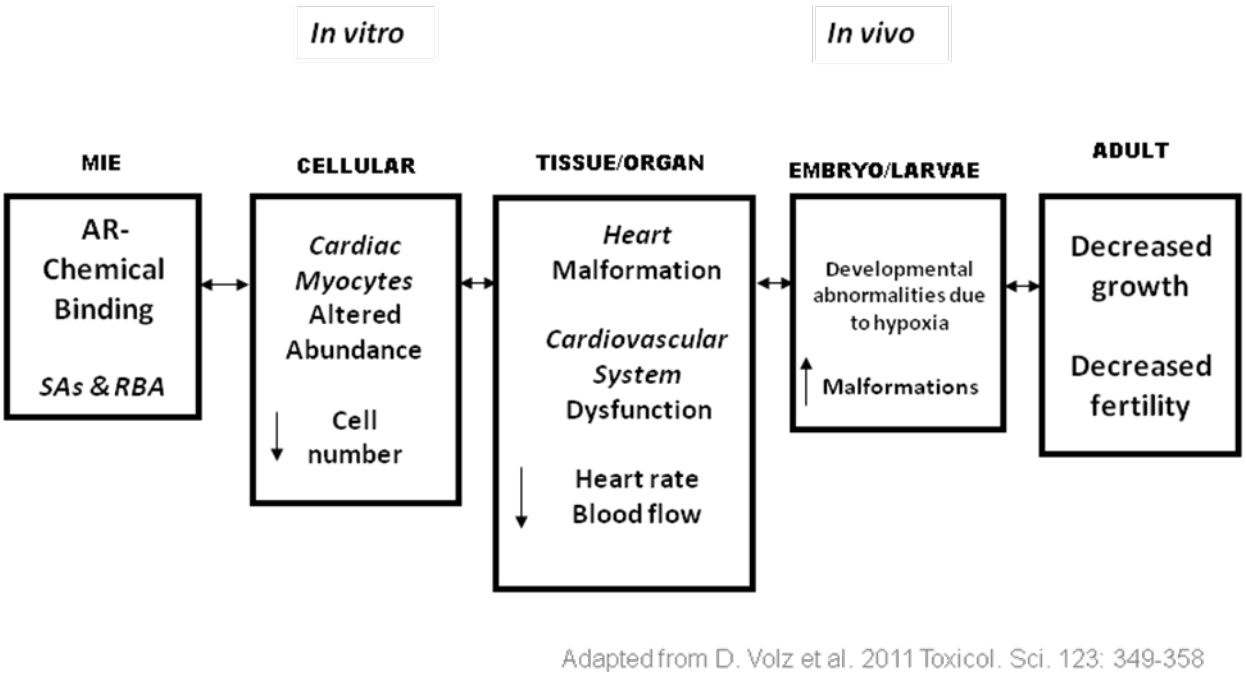

Figure 3. An AOP for AR-mediated Fish Growth Impairment. This figure illustrates the typical responses observed at various levels of biological organization, *i.e.,* AR-binding, alterations in growth of most susceptible cell types, malformations/dysfunction in the most likely organs/organ systems, developmental abnormalities in the embryo/larvae, and decreased growth/fertility in adults.

The Panel noted that there were several ongoing international activities related to the androgenic and thyroid-dependent pathways that would be of use to the Agency. With respect to validation activities, the Japanese androgen receptor stably transfected *in vitro* assay is currently undergoing additional validation work following the recommendations from the OECD peer review report (OECD 2011). The Danish National Food Institute has published environmental chemical relevant models on ER-alpha and AR (Jensen *et al.* 2008, Vinggaard *et al.* 2008), as well as PXR (Dybdahl *et al.* 2012). Combinations of 3D modeling tools, such as 3D QSAR models, are also available for steroid hormone receptors such as AR, glucocorticoid receptor (GR), PXR, peroxisome proliferator-activated receptor-gamma (PPAR-gamma), progesterone receptor (PR), thyroid hormone receptors (TRs), and the cytosolic aryl hydrocarbon receptor (AhR) (*e.g.*, Vedani *et al*. 2012). The Open Virtual Toxlab also provides a free platform for public research (Vedani *et al.* 2012). There are models published for other endocrine

disruption modalities, such as the retinoic X receptor (RXR) and estradiol sulfotransferase, but are not sufficient to be of general use in chemical risk assessment.

The Panel also indicated that the Androgenized Female Stickleback Screen (AFSS) could be quite useful for measuring androgenic activity. Briefly this is a 21-day *in vivo* assay for identifying endocrine active chemicals with (anti)androgenic activity in fish using female sticklebacks (*Gasterosteus aculeatus*) (OECD 2011a). The presence of the specific biomarker for androgens, "spiggin", is present only in this species.

Additionally, the current OECD TG 229 and to some extent TG 230 can detect androgen antagonism, in addition to other endocrine disrupting actions; however, the activity detected is not always clearly specific to androgen antagonism (OECD 2011b). The Hershberger assay (OECD 2009) detects androgenic activity in castrated rats.

9.3 What will be the challenges in developing a prioritization using in vitro methods and computer based simulations for the thyroid hormone system?

Panel Response

A thyroid disruptor may stimulate or inhibit the thyroid through multiple mechanisms and pathways. The Panel noted that the development of any ES to prioritize testing for alterations of the thyroid hormone system may need to consider a variety of potential MIEs (as illustrated in **Figure 4)**, which generally have less data to use in building a model. The Panel recognized that because of the increased complexity and data limitations of the thyroid system, that this might lead, at least initially, to a weaker ability to categorize TDs.

While mediating receptor binding is generally agreed to be the primary initiating event for ER-alpha and AR-related endocrine disruption, thyroid receptor (TR)-binding may not be the primary means of a chemical altering the thyroid hormone system. Thus, the scheme developed for the ER-alpha pathway, discussed during this SAP meeting, may not be directly applicable to the thyroid disruptors (TDs). Since thyroid hormone homeostasis in mammalian systems is important in the development of the central nervous system, general growth of the individual and plays a role in metabolism, the final endpoint for TDs may vary.

Given the choice of multiple MIEs and AOPs, the Panel indicated that it would be difficult to scientifically justify the choice of one specific thyroid-related AOP from another as the "most important" AOP (*e.g.*, the AOP that will have the greatest chemical space for active compounds). The challenges include finding or developing experimental evidence from *in vivo* studies that define the AOP and the mechanistic understanding of how one or a few model chemicals perturb biological pathways associated with the final outcome.

During the FIFRA SAP Consultation of May 24-26, 2011 entitled "Integrated Approaches to Testing and Assessment Strategy: Use of New Computational and Molecular Tools" (see SAP report, EPA-HQ-OPP-2011-0284-0006.pdf), the Panel reviewed a proposed AOP for the molecular initiating event that leads to activation of the pregnane-X-receptor (PXR) and/or the constitutive androstane receptor (CAR)

in the liver, which then leads to the final adverse outcomes of impaired neurodevelopment in children (shown in **Figure 4)**. Briefly, the activation of PXR and CAR leads to up-regulation of hepatic Phase I and Phase II enzymes and hepatic transporters, which results in increased catabolism of thyroxine (T4) and a concomitant decrease in circulating T4, tissue T4, and tissue triiodothyronine (T3).

As shown in Figure 4, the activation of PXR and CAR and indirect stimulation of enzymes in the liver is only one pathway. Thyroid disruptors exert their action directly on the thyroid by stimulating or inhibiting thyroid receptors or by stimulating or inhibiting thyroid enzymes (*e.g.,* thyroid peroxidase) that inactivate the thyroid hormone or displace it from binding proteins in the blood.

Thyroid receptor alpha-1 (TRα1), TR alpha-2 (TRα2) and TR beta-1(TRβ1) activation are all possible MIEs. The ligand binding domains of these receptors or the impact on the recruitment/release of co-activator/co-repressor factors of the thyroid receptor are potential key events. Tissue triiodothyronine T3 or T4 binding or competitive inhibition between chemicals and the serum protein carriers, thyroxine binding globulin or transthyretin, are other possible MIEs. Similarly, thyroid peroxidase inhibition, suppression of iodination or iodide uptake, or the inhibition of the sodium/iodide symporter (NIS) also may be possible MIEs. The Panel suggests the Agency explore some of these other MIEs, also reviewed in a recent OECD detailed review paper on new endpoints (OECD 2012d), and establish a priority list of pathways.

The Panel noted that the methodologies used to develop a thyroid disruptor ES are likely to be different in design than those for the ER-alpha or AR binding systems. For example, human neural progenitor cell systems exposed to potential TDs and the *in vitro* measurements of proliferation, migration and differentiation are possible cellular key events. *Ex vivo* assays with Zebrafish larvae (or other fish species) or with *Xenopus laevis* embryos, which study events in very early stages of development, might also be relevant in screening potential thyroid-disrupting chemicals for higher level effects.

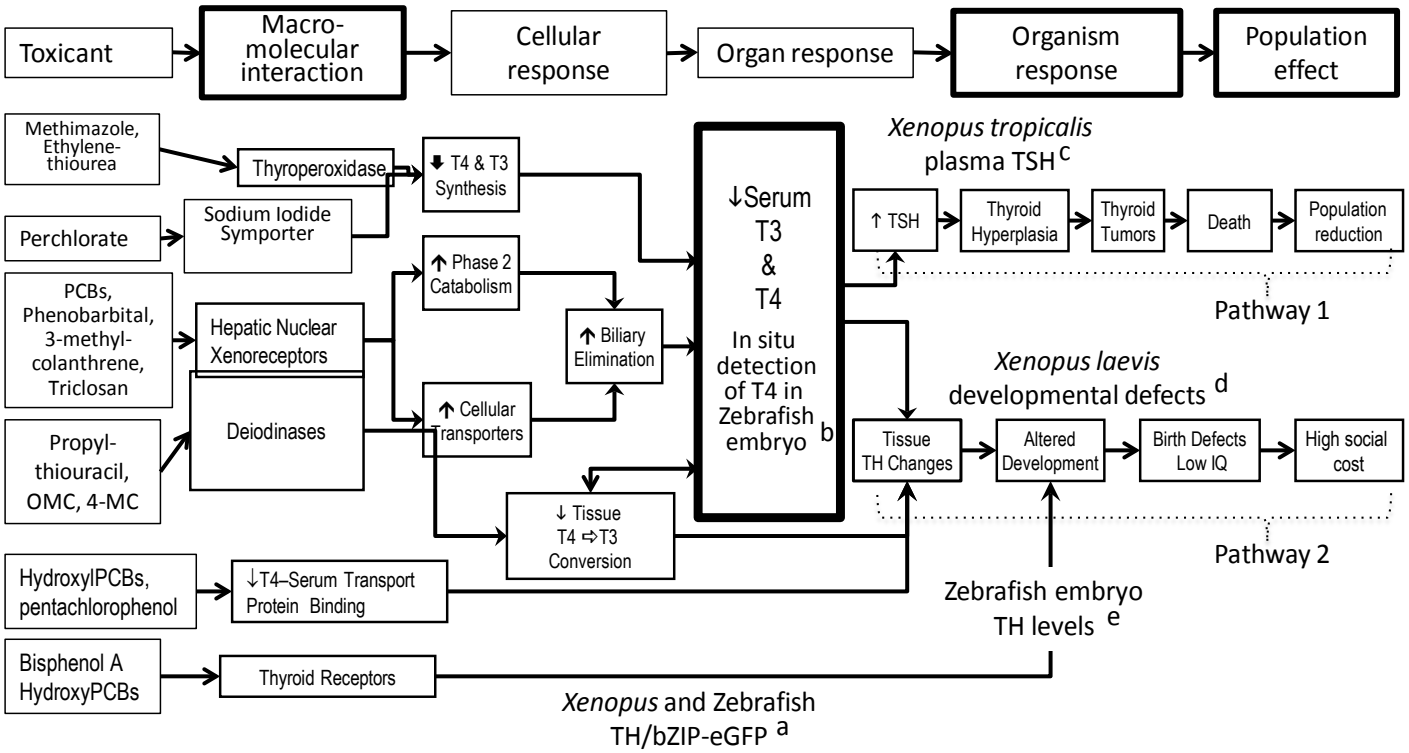

Figure 4. Major Adverse outcome pathways for thyroid disruption with example toxicants, alternative species and non-animal models applicable to human and ecological risk assessment. The black box indicates the critical event of serum level concentrations of thyroid hormones. 1. Rat pathway leading to tumors via thyroid hyperplasia. 2. Principle pathway of concern affecting humans. Figure modified from Miller et al. 2009.

a. **Reporter gene (eGFP) detection of Thyroid Receptor (TR) activity (Fini *et al.* 2007)**
b. **Detection of developmental defects with zebrafish embryos (Nesan and Vijayan 2012)**
c. **Quantification of plasma thyroid Stimulating Hormone levels (Korte *et al.* 2011)**
d. **Detection of developmental defects with *Xenopus laevis* metamorphosis assay (Degitz *et al.* 2005; OECD 2004)**
e. **Direct quantification of thyroid hormone concentrations in zebrafish embryos (Thienpont *et al.* 2011)**

REFERENCES

Abdelrahim M, Ariazi E, Kim K, Khan S, Barhoumi R, Burghardt R, Liu S, Hill D, Finnell R, Wlodarczyk B, Jordan VC, Safe S. (2006) 3-Methylcholanthrene and other aryl hydrocarbon receptor agonists directly activate estrogen receptor alpha. Cancer Res. 2006 Feb 15;66(4):2459-67.

Abdelrahim M, Smith R 3rd, Safe S. (2003) Aryl hydrocarbon receptor gene silencing with small inhibitory RNA differentially modulates Ah-responsiveness in MCF-7 and HepG2 cancer cells. Mol Pharmacol. 2003 Jun;63(6):1373-81.

Barkhem T, Haldosén LA, Gustafsson JA, Nilsson S. (2002a) Transcriptional synergism on the pS2 gene promoter between a p160 coactivator and estrogen receptor-alpha depends on the coactivator subtype, the type of estrogen response element, and the promoter context. Mol Endocrinol. 2002 Nov;16(11):2571-81.

Barkhem T, Haldosén LA, Gustafsson JA, Nilsson S. (2002b) pS2 Gene expression in HepG2 cells: complex regulation through crosstalk between the estrogen receptor alpha, an estrogen-responsive element, and the activator protein 1 response element. Mol Pharmacol. 2002 Jun;61(6):1273-83.

Bender A, Jenkins JL, Glick M, Deng Z, Nettles JH, Davies JW. (2006) “Bayes affinity fingerprints" improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? J Chem Inf Model. 2006 Nov-Dec;46(6):2445-56.

Berry M, Nunez AM, Chambon P. (1989) Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc Natl Acad Sci U S A. 1989 Feb;86(4):1218-22.

Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, *et al.* (2012) Automated design of ligands to polypharmacological profiles. Nature. 2012 Dec 13;492(7428):215-20. doi: 10.1038/nature11691.

Blaauboer BJ. (2010) Biokinetic modeling and *in vitro*–in vivo extrapolations. Journal of Toxicology and Environmental Health, Part B 13:242-252.

Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. (2008) Drug target identification using side-effect similarity. Science. Jul 11;321(5886):263-6.

Burns KA, Korach KS. Estrogen receptors and human disease: an update. ArchToxicol. 2012 Oct;86(10):1491-504. doi: 10.1007/s00204-012-0868-5. Epub 2012 May 31

Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, *et al*. (2011) Controlled vocabularies and semantics in systems biology. Mol Syst Biol. 7:543.

Cronin MTD, Worth AP. (2008) (Q)SARs for predicting effects relating to reproductive toxicity. QSAR and Combinatorial Science 27: 91-100.

Dearden JC, Rotureau P, Fayet G. (2013) QSPR prediction of physico-chemical properties for REACH. SAR and QSAR in Environmental Research, in press.

Degitz SJ, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, Tietge JE. (2005) Progress towards development of an amphibian-based thyroid screening assay using *Xenopus laevis*. Organismal and thyroidal responses to the model compounds 6-propylthiouracil, methimazole, and thyroxine. Toxicol Sci. 87:353–364.

Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3(96): 96ra76.

Dybdahl M, Nikolov NG, Wedebye EB, Jonsdottir SO, Niemela JR (2012) QSAR model for human pregnane X receptor (PXR) binding: screening of environmental chemicals and correlations with genotoxicity, endocrine disruption and teratogenicity. Toxicology and Applied Pharmacology 262:301-309.

Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC). (1998) EDSTAC Final report. August 1998. U.S. Environmental Protection Agency. Available online: http://www.epa.gov/endo/pubs/edspoverview/finalrpt.htm.

EFSA (European Food Safety Authority) (2012) Scientific Opinion on Evaluation of the Toxicological Relevance of Pesticide Metabolites for Dietary Risk Assessment EFSA Journal 2012;10(7):2799 [187 pp.].

Ellison CM, Madden JC, Judson P, Cronin MTD. (2010) Using *in silico* tools in a weight of evidence approach to aid toxicological assessment. Molecular Informatics 29: 97–110.

Enoch SJ, Ellison CM, Schultz TW, Cronin MTD. (2011) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. Critical Reviews in Toxicology 41:783-802.

Fini J-B, Le Mével S, Turque N, Palmier K, Zalko D, Cravedi P, Demeneix BA. (2007)An *in vivo* multiwell-based fluorescent screen for monitoring vertebrate thyroid hormone disruption. Environ. Sci. Technol., 2007 Jul 20; 41 (16), pp. 5908–5914. DOI: 10.1021/es0704129

Gould JC, Leonard LS, Maness SC, Wagner BL, Conner K, Zacharewski T, *et al.* (1998) Bisphenol A interacts with the estrogen receptor alpha in a distinct manner from estradiol. Mol Cell Endocrinol. 1998 Jul 25;142(1-2):203-14.

Gutleb AC, Meerts IATM, Bergsma JH, Schriks M, Murk AJ.(2005) T-Screen as a tool to identify thyroid hormone receptor active compounds. Environ. Toxicol. Pharmacol., 19 (2005), pp. 231–238.

Hamers T, Kamstra JH, Cenijn PH, Pencikova K, Palkova L, Simeckova P, *et al.* (2011) In vitro toxicity profiling of ultrapure non-dioxin-like polychlorinated biphenyl congeners and their relative toxic contribution to PCB mixtures in humans. Toxicol Sci. 2011 May;121(1):88-100.

Hastings J, Magka D, Batchelor C, Duan L, Stevens R, Ennis M, Steinbeck C. (2012) Structure-based classification and ontology in chemistry. J Cheminform. 4:8.

Hewitt SC, Korach KS. (2003) Oestrogen receptor knockout mice: roles for oestrogen receptors alpha and beta in reproductive tissues. Reproduction. 2003 Feb;125(2):143-9. Review. PubMed PMID: 12578528.

Immonen E, Serpi R, Vähäkangas K, Myllynen P. (2009) Responses of PhIP (2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine) in MCF-7 cells are culture condition dependent. Chem Biol Interact. 2009 Nov 10;182(1):73-83. doi: 10.1016/j.cbi.2009.07.020. Epub 2009 Aug 6.

Jacobs MN. (2005) Nuclear receptors and dietary ligands. In J Zemplini, K Dakshinamurti, (eds.) Nutrients and Cell Signaling. Oxidative Stress and Disease series. CRC Taylor and Francis Inc, New York. Spring 2005. pp 35-90

Jacobs MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R., *et al.* (2008) The use of metabolising systems for *in vitro* testing of endocrine disruptors. Curr. Drug Metab. 8:796-826.

Jacobs MN, Laws S, Willett C, Schmieder P, Odum J, Bovee T. (2013) *In vitro* metabolism and bioavailability tests for endocrine active substances: What is needed next for regulatory purposes. ALTEX, in press.

Jeng YJ, Kochukov M, Watson CS. (2010) Combinations of physiologic estrogens with xenoestrogens alter calcium and kinase responses, prolactin release, and membrane estrogen receptor trafficking in rat pituitary cells. Environ Health. 2010 Oct 15;9:61. doi: 10.1186/1476-069X-9-61.

Jensen GE, Niemela JR, Wedebye EB, Nikolov NG. (2008) QSAR models for reproductive toxicity and endocrine disruption in regulatory use-a preliminary investigation. SAR and QSAR in Environmental Research 19: 631-641.

Katzenellenbogen J, Muthyala R, Katzenellenbogen BS. (2003) Nature of the ligand-binding pocket of estrogen receptor α and ß: The search for subtype-selective ligands and implications for the prediction of estrogenic activity. Pure and Applied Chemistry. 75:2397-2403.

Kim J, Petz LN, Ziegler YS, Wood JR, Potthoff SJ, Nardulli AM. (2000) Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells.J Steroid Biochem Mol Biol. 2000 Nov 15;74(4):157-68.

Klaassen I, Brakenhoff RH, Smeets SJ, Snow GB, Braakhuis BJM. 1999. Considerations for *in vitro* experiments: importance of protein interaction. Bichemica et Biophysica Acta 1427: 265-275.

Kaushal D, Naeve CW. (2004a) Analyzing and visualizing expression data with Spotfire. Curr Protoc Bioinformatics. 2004 Oct;Chapter 7:Unit 7.9

Kaushal D, Naeve CW. (2004b) Loading and preparing data for analysis in Spotfire. Curr Protoc Bioinformatics. 2004 Sep;Chapter 7:Unit 7.8.

Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, *et al*. (2009) Predicting new molecular targets for known drugs. Nature. 2009 Nov 12;462(7270):175-81. doi: 10.1038/nature08506. Epub 2009 Nov 1.

Korte JJ, Sternberg RM, Serrano JA, Thoemke KR, Moen SM, Lillegard KE, *et al.* (2011) Thyroid-stimulating hormone (TSH): Measurement of intracellular, secreted, and circulating hormone in

Xenopus laevis and *Xenopus tropicalis.* General and Comparative Endocrinology. 171(3):319-325. http://dx.doi.org/10.1016/j.ygcen.2011.02.017.

Kummer V, Masková J, Zralý Z, Neca J, Simecková P, Vondrácek J, Machala M. (2008) Estrogenic activity of environmental polycyclic aromatic hydrocarbons in uterus of immature Wistar rats. Toxicol Lett. 2008 Aug 28;180(3):212-21.

Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom, AR, Ma'ayan A. (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 26(19):2438-2444.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006 Sep 29;313(5795):1929-35.

Leitman DC, Paruthiyil S, Yuan C, Herber CB, Olshansky M, Tagliaferri M, *et al*. (2012) Tissue-specific regulation of genes by estrogen receptors. Semin Reprod Med. 2012 Jan;30(1):14-22. doi: 10.1055/s-0031-1299593. Epub 2012 Jan 23. Review.

Leitman DC, Paruthiyil S, Vivar OI, Saunier EF, Herber CB, Cohen I, *et al.* (2010) Regulation of specific target genes and biological responses by estrogen receptor subtype agonists. Curr Opin Pharmacol. 2010 Dec;10(6):629-36. doi: 10.1016/j.coph.2010.09.009. Epub 2010 Oct 14. Review.

Lentz KA, Hayashi J, Lucisano LJ, Polli JE. (2000) Development of a more rapid, reduced serum culture system for Caco-2 monolayers and application to the biopharmaceutics classification system. International Journal of Pharmaceutics 200:41-51.

Li Y, Sun L, Zhang Y, Wang D, Wang F, Liang J, Gui B, Shang Y. (2011) The histone modifications governing TFF1 transcription mediated by estrogen receptor. J Biol Chem. 2011 Apr 22;286(16):13925-36.

Lipinski CA, Lombardo, F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3-25

Liu S, Abdelrahim M, Khan S, Ariazi E, Jordan VC, Safe S. (2006) Aryl hydrocarbon receptor agonists directly activate estrogen receptor alpha in MCF-7 breast cancer cells. Biol Chem. 2006 Sep;387(9):1209-13.

Lo R, Matthews J. (2010) A new class of estrogen receptor beta-selective activators. Mol Interv. 2010 Jun;10(3):133-6. doi: 10.1124/mi.10.3.3.

Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, *et al*. (2012) Large-scale prediction and testing of drug activity on side-effect targets. Nature. 2012 Jun 10;486(7403):361-7. doi: 10.1038/nature11159.

Madden JC. (2010a)Toxicokinetic Considerations in Predicting Toxicity. In MTD Cronin, JC Madden (eds) *In Silico* Toxicology. RSC Publishing. pp. 531-557.

Madden JC (2010b) *In silico* approaches for predicting ADME properties. In T Puzyn, J Leszczynski, MTD Cronin (eds) Recent Advances in QSAR Studies - Methods and Applications. Springer, Dordrecht, pp 283-304.

Madin J, Bowers S, Schildhauer M, Jones M. (2008) Advancing ecological research with ontologies. Trends in Ecology and Evolution Vol.23: 159-168.

Marino M, Galluzzo P, Ascenzi P. (2006) Estrogen signaling multiple pathways to impact gene transcription. Curr Genomics. 7(8):497-508.

Marino M, Pellegrini M, La Rosa P, Acconcia F. (2012) Susceptibility of estrogen receptor rapid responses to xenoestrogens: Physiological outcomes. Steroids. 2012 Aug;77(10):910-7. doi: 10.1016/j.steroids.2012.02.019. Epub 2012 Mar 5.

Marshall MS, Boyce R, Deus HF, Zhao J, Willighagen EL, Samwald M, Pichler E, Hajagos J, Prud"hommeaux E, Stephens S. 2012. Emerging practices for mapping and linking life sciences data using RDF – a case series. J Web Semantics. Vol 14.

McDonnell DP, Clemm DL, Hermann T, Goldman ME, Pike JW. (1995) Analysis of estrogen receptor function *in vitro* reveals three distinct classes of antiestrogens. Mol Endocrinol. 1995 Jun;9(6):659-69.

Miller MD, Crofton KM, Rice DC, Zoeller RT. (2009) Thyroid-disrupting chemicals: interpreting upstream biomarkers of adverse outcomes. Environ Health Perspect. 2009 July; 117(7):1033–1041. Published online 2009 February 12. doi: 10.1289/ehp.0800247

Mortensen AS, Arukwe A. Dimethyl sulfoxide is a potent modulator of estrogen receptor isoforms and xenoestrogen biomarker responses in primary culture of salmon hepatocytes. Aquat Toxicol. 2006 Aug 12;79(1):99-103. Epub 2006 Jun 3.

Mueller SO. Xenoestrogens: mechanisms of action and detection methods. Anal Bioanal Chem. 2004 Feb;378(3):582-7. Epub 2003 Oct 16.

Nelms MD, Enoch SJ, Fioravanzo E, Madden JC, Meinl T, Richarz A-N, *et al.* (2012) Strategies to Form Chemical Categories from Adverse Outcome Pathways. In: Gocht T, Schwarz M (eds) Seurat-1 Annual Report. Volume 2. pp. 263-267. Available from: http://www.seurat-1.eu/pages/library/seurat-1-annual-report.php.

Nesan D, Vijayan DM. (2012) Embryo exposure to elevated cortisol level leads to cardiac performance dysfunction in zebrafish, Mol Cell Endocrinol, 363: 85-91.

Nidhi, Glick M, Davies JW, Jenkins JL. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model. 2006 May-Jun;46(3):1124-33.

Ochsner, SA, Steffen DL, Hilsenbeck SG, Chen ES, Watkins C, McKenna, NJ. (2009) GEMS (Gene Expression MetaSignatures), a Web resource for querying meta-analysis of expression microarray datasets: 17beta-estradiol in MCF-7 cells. Cancer Res. 69(1): 23-26.

OECD (Organization for Economic Co-operation and Development) (2009). OECD Guideline for Testing of Chemicals: Test No 441: Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti)Androgenic Properties.

OECD (Organization for Economic Co-operation and Development) (2011a) OECD Guidance Document on the androgenised female stickleback screen. Series on Testing and Assessment: No 148: ENV/JM/MONO(2011) 29. http://www.oecd.org/env/ehs/testing/48532191.pdf

OECD (Organization for Economic Co-operation and Development) (2011b) Peer Review Report for the Validation of the Stably Transfected Transcriptional Activation Assay for the Detection Androgenic and Anti-Androgenic Activity of Chemicals. Series on Testing and Assessment 161: ENV/JM/MONO(2011)46.

OECD (Organisation for Economic Co-operation and Development) (2012a).Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment, No. 150, OECD, Paris. http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2012)22&doclanguage=en.

OECD (Organisation for Economic Co-operation and Development) (2012b) OECD Guideline for Testing of Chemicals Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists. OECD, Paris.

OECD (Organisation for Economic Co-operation and Development (2012c) OECD Guideline for Testing of Chemicals: Test No 457: BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists.

OECD (Organisation for Economic Co-operation and Development) (2012d) Series on Testing and Assessment: No 178: Detailed Review Paper on the State of the Science on Novel *In vitro* and *In vivo* Screening and Testing Methods and Endpoints for Evaluating Endocrine Disruptors. ENV/JM/MONO(2012)23, 213 pp.

Piechota P, Cronin MTD, Hewitt M, Madden JC. (2013) Pragmatic approaches to predicting xenobiotic metabolism. Journal of Chemical Informatics and Modeling. (Submitted)

Rasmussen TH, Nielsen JB. (2002) Critical parameters in the MCF-7 cell proliferation bioassay (E-Screen). Biomarkers. 2002 Jul-Aug;7(4):322-36.

Rogers D, Brown RD, Hahn M. (2005) Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. J Biomol Screen. 2005 Oct;10(7):682-6. Epub 2005 Sep 16.

Safe S, Kim K. (2008) Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. J Mol Endocrinol. 2008 Nov;41(5):263-75. doi: 10.1677/JME-08-0103. Epub 2008 Sep 4.

Safe SH, Pallaroni L, Yoon K, Gaido K, Ross S, Saville B, McDonnell CD. (2001)Toxicology of environmental estrogens. Reprod Fertil Dev. 2001;13(4):307-15.

Sansone S-A, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, *et al*. (2012) Toward interoperable bioscience data. Nature Publishing Group 44:121–126; doi:10.1038/ng.1054.

Saubern S, Guha R, Baell JB. (2011) KNIME workflow to assess PAINS filters in SMARTS format. Comparison of RDKit and Indigo Cheminformatics Libraries. Molecular Informatics 30:847- 850.

Schmieder PK, Tapper MA, Denny JS, Kolanczyk RC, Sheedy BR, Henry TR, Veith GD. (2004) Use of trout liver slices to enhance mechanistic interpretation of estrogen receptor binding for cost-effective prioritization of chemicals within large inventories. Environ Sci Technol. 2004 Dec 1;38(23):6333-42.

Schmieder P, Tapper M, Linnum A, Denny J, Kolanczyk R, Johnson R. (2000) Optimization of a precision-cut trout liver tissue slice assay as a screen for vitellogenin induction: comparison of slice incubation techniques. Aquat Toxicol. 2000 Jul 1;49(4):251-268.

Schwöbel JAH, Koleva YK, Enoch SJ, Bajot F, Hewitt M, Madden JC, *et al.* (2011) Measurement and estimation of electrophilic reactivity for predictive toxicology. Chemical Reviews, 111:2562–2596.

Swedenborg E, Rüegg J, Hillenweck A, Rehnmark S, Faulds MH, Zalko D, Pongratz I, Pettersson K. (2008) 3-Methylcholanthrene displays dual effects on estrogen receptor (ER) alpha and ER beta signaling in a cell-type specific fashion. Mol Pharmacol. 2008 Feb;73(2):575-86.

Sung E, Turan N, Ho PW, *et al*. (2012). Detection of endocrine disruptors - from simple assays to whole genome scanning. Int. J. Androl. 35(3):407-14.

Thienpont B, Tingaud-Sequeira A, Prats E, Barata C, Babin PJ, Raldúa D. (2011) Zebrafish eleutheroembryos provide a suitable vertebrate model for screening chemicals that impair thyroid hormone synthesis. Environ. Sci. Technol. 45(7):7525-7532.

Timm M, Saaby L, Moesby L, Hansen EW. (2013) Considerations regarding use of solvents in *in vitro* cell based assays. Cytotechnology. 2013 Jan 18. [Epub ahead of print].

Vedani A, Dobler M, Smiesko M. (2012) VirtualToxLab - a platform for estimating the toxic potential of drugs, chemicals and natural products. Toxicology and Applied Pharmacology 261:142-153.

Vidal D, Garcia-Serna R, Mestres J. Ligand-based approaches to in silico pharmacology. Methods Mol Biol. 2011;672:489-502. doi: 10.1007/978-1-60761-839-3_19.

Vinggaard AM, Niemela J, Wedebye EB, Jensen GE. (2008) Screening of 397 chemicals and development of a quantitative structure-activity relationship model for androgen receptor antagonism. Chemical Research in Toxicology 21:813-823.

Volz DC, Belanger S, Embry M, Padilla S, Sanderson H, Schirmer K, Scholz S, Villeneuve D. (2011) Adverse outcome pathways during early fish development: A conceptual framework for identification of chemical screening and prioritization strategies. Toxicological Sciences 123: 349-358.

Wang S, Aarts JMMJG, Evers NM *et al*. (2012) Proliferation assays for estrogenicity testing with high predictive value for the *in vivo* uterotrophic effect. J. Steroid Biochem. Mol. Biol. 128:98-106.

Waring RH, Ramsden DB, Jarratt PB, Harris RM. (2012). Biomarkers of endocrine disruption: cluster analysis of effects of plasticisers on Phase 1 and Phase 2 metabolism of steroids. Int. J. of Andrology 35:415-423.

Williams RJ, Martinez N, Golbeck J. Ontologies for ecoinformatics. Web Semantics: Science, Services and Agents on the World Wide Web. Volume 4, Issue 4, December 2006, Pages 237–242.

Won Jeong K, Chodankar R, Purcell DJ, Bittencourt D, Stallcup MR. (2012) Gene-specific patterns of coregulator requirements by estrogen receptor-α in breast cancer cells. Mol Endocrinol. 2012 Jun;26(6):955-66.

Wormke M, Stoner M, Saville B, Walker K, Abdelrahim M, Burghardt R, Safe S. (2003) The aryl hydrocarbon receptor mediates degradation of estrogen receptor alpha through activation of proteasomes. Mol Cell Biol. 2003 Mar; 23(6):1843-55.

Yoon K, Pellaroni L, Ramamoorthy K, Gaido K, Safe S. (2000) Ligand structure-dependent differences in activation of estrogen receptor alpha in human HepG2 liver and U2 osteogenic cancer cell lines. Mol Cell Endocrinol. 2000 Apr 25;162(1-2):211-20.

Yoon K, Pallaroni L, Stoner M, Gaido K, Safe S. (2001) Differential activation of wild-type and variant forms of estrogen receptor alpha by synthetic and natural estrogenic compounds using a promoter containing three estrogen-responsive elements. J Steroid Biochem Mol Biol. 2001 Jul;78(1):25-32.

Zacharewski T. (1997) *In vitro* bioassays for assessing estrogenic substances. Environ. Sci. Technol. 31:613-623.