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

OFFICE OF
PREVENTION, PESTICIDES AND
TOXIC SUBSTANCES

MEMORANDUM

DATE: November 19, 2009

SUBJECT: Transmittal of the Meeting Minutes of the FIFRA SAP Meeting Held August 25-27, 2009 on the Scientific Issues Associated with "The Use of Structure Activity Relationships of Estrogen binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing"

TO: Debbie Edwards, Ph.D.
Director
Office of Pesticide Program

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

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

Frank Sanders
Director
Office of Science Coordination and Policy

Please find attached to this memorandum the meeting minutes of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) open meeting held in Arlington, Virginia on August 25-26, 2009. This report addresses a set of scientific issues associated with "The Use of Structure Activity Relationships of Estrogen binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing."

Attachment

cc:
Stephen Owens
James Jones
Betsy Shaw
Steven Bradbury, Ph.D.
Vicki Dellarco, Ph.D.
William Jordan
Margie Fehrenbach
Donald Brady, Ph.D.
Keith Matthews
William Diamond
Joan Harrigan-Farrelly
Tina Levine, Ph.D.
Lois Rossi
Richard Keigwin
Jack Housenger
Vicki Dellarco, Ph.D.
Patricia Schmeider, Ph.D.
Enesta Jones
Doug Parsons
Vanessa Vu, Ph.D. (SAB)
OPP Docket

FIFRA SAP Members
Janice Chambers, Ph.D., DABT, ATS
(FIFRA SAP Session Chair)
Steven Heeringa, Ph.D. (FIFRA SAP Chair)
John Bucher, Ph.D., DABT
Carey Pope, Ph.D.
Kenneth Portier, Ph.D.
Daniel Schlenk, Ph.D.

FQPA Science Review Board Members

Mark Cronin, Ph.D.
Nancy Denslow, Ph.D.
Miriam Jacobs, Ph.D.
Bette Meek, M.Sc.
Reynaldo Patiño, Ph.D.
Terry Schultz, Ph.D.
Leming Shi, Ph.D.

SAP Minutes No. 2009-08

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

The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing

August 25-26, 2009
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 or disseminated 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. 2009-08

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

The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing

August 25-26, 2009
FIFRA Scientific Advisory Panel Meeting
held at
One Potomac Yard
Arlington, Virginia

Janice E. Chambers, Ph.D., DABT, ATS
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel

Janice E. Chambers

Date: 11/19/09

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

Date: 11/19/09.

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
August 25-26, 2009

Scientific Issues Associated with The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing

PARTICIPANTS

FIFRA SAP Session Chair

Janice E. Chambers, Ph.D., DABT, ATS
William L. Giles Distinguished Professor
Director, Center for Environmental Health Sciences
College of Veterinary Medicine
Mississippi State University
Mississippi State, Mississippi

Designated Federal Official

Sharlene Matten, Ph.D.
US 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, Fax: 202-564-8382, matten.sharlene@epa.gov

FIFRA Scientific Advisory Panel Members

John R. Bucher, Ph.D., DABT
Associate Director
National Toxicology Program
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina

Carey N. Pope, Ph.D.
Professor, Head & Sitlington Chair of Toxicology
Department of Physiological Sciences
Oklahoma State University College of Veterinary Medicine
Stillwater, Oklahoma

Kenneth M. Portier, Ph.D.
Program Director, Statistics
American Cancer Society
National Home Office
Atlanta, Georgia

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

FQPA Science Review Board Members

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

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

Miriam Jacobs, Ph.D.
Scientific Officer
European Food Safety Authority - EFSA
Assessment Methodology Unit
Parma, ITALY

Bette Meek, M.Sc.
Associate Director
Chemical Risk Assessment
McLaughlin Centre for Population Health Risk Assessment
University of Ottawa
Ottawa, Ontario, CANADA

Reynaldo Patiño, Ph.D.
Unit Leader, U.S. Geological Service, Cooperative Fish & Wildlife Research Unit
& Professor, Texas Tech University
Lubbock, Texas

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

Leming Shi, Ph.D.
Computational Chemist – Principal Investigator
National Center for Toxicological Research
U.S. Food and Drug Administration
Jefferson, Arkansas

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed its review of the Agency's analysis of **Scientific Issues Associated with The Use of Structure Activity Relationships of Estrogen Binding Affinity to Support Prioritization of Pesticide Inert Ingredients and Antimicrobial Pesticides for Screening and Testing**. Advance notice of the SAP meeting was published in the *Federal Register* on **June 3, 2009**. The review was conducted in an open Panel meeting August 25-26, 2009 held at One Potomac Yard, Arlington, Virginia. Dr. Janice Chambers chaired the meeting. Dr. Sharlene Matten served as the Designated Federal Official. Mr. Stephen Owens, Assistant Administrator, Office of Pesticides, Prevention, and Toxic Substances; Ms. Laura Bailey, Executive Secretary, FIFRA SAP, Office of Science and Coordination Policy (OSCP); and Dr. Stephen Bradbury, Deputy Office Director for Programs, Office of Pesticide Programs (OPP) provided opening remarks at the meeting. Presentations of technical background materials were provided by Dr. Stephen Bradbury, Deputy Office Director for Programs, OPP; Dr. Patricia Schmieder and Mr. Richard Kolanczyk, Mid-Continent Ecology Division, National Health and Environmental Effects Research Laboratory (NHEERL)-Duluth, Office of Research and Development; and Mr. Gary Timm, OSCP.

The Office of Pesticide Programs (OPP) has a number of ongoing activities to broaden the suite of computer-based methods to better predict potential hazards and improve the effectiveness of risk assessment and risk management by focusing the generation of new data on likely risks of concern. The use of quantitative structure-activity relationships ((Q)SARs) in early hazard identification is one such cost effective prioritization tool that can guide the systematic collection of key test data. The U.S. EPA is required by the FQPA of 1996 to screen all pesticide chemicals, which includes active ingredients and other ("inert") ingredients for potential endocrine disruption (ED). The Agency must prioritize hundreds of chemicals for ED screening and testing in biologically complex and resource intensive assays. This FIFRA SAP review focused on a (Q)SAR approach that was developed for pesticide food use inert ingredients and antimicrobial pesticides to help prioritize candidate chemicals for the Tier 1 Endocrine Disruptor Screening Program (EDSP) specific to an estrogen-mediated toxicity pathway. A particularly challenging issue is the development of prioritization techniques for chemicals, such as inert ingredients and some antimicrobial active ingredients that have minimal existing toxicological information. This SAR/Rule-Based Expert System (expert system) for predicting estrogen receptor (ER) binding affinity is based on data derived from two *in vitro* assays: one optimized to measure the potential of chemicals to bind rainbow trout estrogen receptors (rtER), and a second that measures gene activation through production of vitellogenin.

The expert system was guided by the Organization for Economic Cooperation and Development (OECD) Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models (OECD, 2004), found at http://www.oecd.org/dataoecd/33/37/37849783.pdf). The five OECD validation principles are: a well defined endpoint; a mechanistic interpretation of the model; a defined domain of the model applicability; an unambiguous algorithm and appropriate measures of goodness-of-fit, robustness, and predictivity. These principles are intended to guide the (Q)SAR validation process undertaken within OECD member countries and to facilitate their regulatory

acceptance by defining the types of information that regulators would find useful when considering the acceptability of individual (Q)SARs. These principles are also consistent with, and complement, recommendations of the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) (USEPA, 1998). Two important characteristics of (Q)SAR validations are addressed by OECD to enhance regulatory acceptance for using estimated values for priority-setting. The first characteristic is transparency of the (Q)SAR estimate, not so much in terms of the methods used, but rather in terms of how the estimate can be explained mechanistically and how the estimate is reasonable based on data for comparable chemicals. The second major characteristic for (Q)SAR acceptance is usefulness of a particular (Q)SAR model for estimating endpoints of regulatory relevance for all compounds within specified chemical inventories. A prototype of the expert system was the subject of an OECD-convened peer consultation in February 2009, at which time the system was evaluated using the (Q)SAR validation principles (OECD, 2009). Based on input from this peer consultation, the Agency further refined the expert system, particularly as it related to the OECD validation principles.

The SAP was asked to address specific issues concerning the (Q)SAR validation principles and the subject expert system in the context of its use to determine the order in which chemicals, i.e., food use inert ingredients and antimicrobial pesticides, will be subject to Tier 1 screening specific to an estrogen-mediated toxicity pathway under EPA's Endocrine Disrupter Screening Program (EDSP), *i.e.*, for prioritization. The SAP also discussed the applicability of ER binding affinity findings for acyclic compounds in the food use inerts and antimicrobial pesticide inventories to other acyclic compounds in other chemical inventories. Finally, the SAP commented on the cross species applicability of an Expert System based primarily on trout estrogen binding affinity data for predicting relative binding affinity to human estrogen receptors to assist in prioritizing food use inert ingredients and antimicrobial pesticides for EDSP Tier 1 screening.

PUBLIC COMMENTERS

Oral statements were presented by:

1) **On behalf of CropLife America:** Erik Janus, Director; Human Health Policy

2) **On behalf of the American Chemistry Council, Biocides Panel:** Elizabeth Brown, Ph.D., Senior Technical & Regulatory Analyst, Steptoe & Johnson, LLP.

Written statements were provided by:

On behalf of the American Chemistry Council, Biocides Panel and Regulatory and Technical Affairs Department: Elizabeth Brown, Ph.D., Senior Technical & Regulatory Analyst, Steptoe & Johnson, LLP.

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

1. Evaluation of the Expert System in the Context of the Organization for Economic Cooperation and Development (OECD) Validation Principles

The components to Charge Question 1 address specific issues concerning the (Q)SAR validation principles and the subject expert system in the context of its use to determine the order in which chemicals (i.e., food use inert ingredients and antimicrobial pesticides) will be subject to Tier 1 screening under EPA's Endocrine Disrupter Screening Program (EDSP) (*i.e.*, for prioritization).

a. Biological Endpoint

Question a.1. *Please comment on the methods employed and their adequacy for detecting and measuring ER binding affinity for the compounds in the two chemical inventories of immediate interest.*

The overall conclusion of the Panel was that the two methods for detecting and measuring ER binding affinity of the compounds in the food use inerts (FI) and antimicrobial (AM) pesticide inventories were appropriate for helping EPA set testing priorities for these chemicals. The two methods provide a logical approach to evaluate compounds with a specific mechanism of action (estrogen receptor (ER)-binding) and then linking this to a transcriptional activation outcome, vitellogenin (Vtg) production. The Panel stated that it was imperative to note that ER-binding or Vtg induction should not be equated with higher order reproductive effects. They recommended that EPA consider adding more discussion in the White Paper to better define the limitations of the assays used to generate binding and biological response data, *i.e.*, what can and cannot be concluded from these data.

Question a.2. *In the context of chemical prioritization for Tier 1 screening, please comment on the decision to measure binding affinity up to the maximum concentration based on the solution properties of the chemical, rather than using ligand concentration 'cut-off' values of -4 Log Molar to -3 Log Molar, which have typically been used to conclude a compound does not bind to the ER.*

There was general agreement among the panel members that the chemicals of concern (FI and AM pesticides) would be of low binding affinity. The Panel indicated that the cut-off values of -4 log Molar to -3 log Molar were established in the pharmaceutical industry as a screen for strong estrogens and are not applicable for the purposes of screening and prioritization. They recommended that the FI and AM pesticides should be tested using concentrations up to their solubility in the media. In addition, the concentrations tested should be in concordance with concentrations that are biologically relevant and cut-off values should be set more broadly to be as inclusive of chemicals that might be able to

bind to estrogen receptors. These chemicals would then be tested further using more in-depth assays.

Question a.3. *Please also comment on how any in vivo studies that are available for compounds with low receptor binding affinity could be used to provide a relative binding affinity 'cut-off' value either alone or in combination with cut-off values based on the maximum solubility of a ligand in the buffer solution.*

Panel members interpreted this charge question in different ways. As a result, the Panel provided several *in vivo* and *in vitro* methods in response to this question.

1) **Distribution of compounds within tissues using dosimetry.** Fathead minnow adults (male or female) and embryo tests could be used to determine whether or not compounds with low receptor binding affinity are bioaccumulated in tissues of endocrine importance, for example, liver, brain, pituitary, gonad, and kidney.
2) **Vitellogenin induction *in vivo* experiments.** Vitellogenin induction, *in vivo*, has been generally accepted as a good biomarker for compounds that bind to and activate ERs in fish.
3) **MCF-7 cell proliferation assays.** These assays use an ERα human breast cancer cell line and monitor cell proliferation following exposure to both estrogen and estrogen mimics that bind through estrogen receptors to induce activity.
4) **48-hour transcriptomics experiments.** Microarrays are now available for a large number of fish species that are routinely tested for toxicity experiments. Genes that are important for reproduction, growth and susceptibility to disease are now known and this assay could be used to determine whether any of these are affected by low concentrations of test chemicals mixed in the food at levels that would be normally experienced by humans.
5) ***In vivo* test for ova-testis.** This assay would look for downstream effects of exposure to estrogenic compounds. The end point would be histological determination of ova-testis.
6) **Fish recrudescence assay.** Cost effective *in vivo* studies would include fish recrudescence assays (Tier 1, *i.e.*, 21-day fish reproductive assay) that partner well with the *in vitro* approach using the rainbow trout estrogen receptor.
7) ***In vitro* assays with other endpoints.** *In vitro* assays could be conducted with regard to other biological endpoints, *e.g.*, estrogen sulfotransferase, aromatase, and other receptors.
8) **Other *in vivo* test guidelines.** Other *in vivo* test guidelines such as the "enhanced" OECD conventional 28-day repeat dose toxicity test (OECD TG 407) and the updated draft physiologically-based pharmacokinetic (PBPK) test guidelines when available would also be useful.
9) **PBPK modeling.** Dosimetry PBPK studies would allow estimates of hepatic concentrations that could be partnered with concentrations

identified from *in vitro* responses (ER-binding/Vtg mRNA induction and Vtg production).

b. Mechanistic Interpretation

Question b.1. *Please comment on the adequacy of this training set for supporting the expert system's rule that acyclic compounds do not bind to the ER.*

The Panel determined that the current acyclic compound training set is adequate for the structures within the two inventories, especially for the intended purpose of screening, *i.e.*, prioritization. The Panel, however, considered the current training set to be inadequate to cover all chemicals. To illustrate this point, the Panel noted that the acyclic compounds, perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA), were shown to bind ER and induce Vtg in at least two fish species (Ishibashi *et al.*, 2008; Liu *et al.*, 2007). The Panel thought that for the acyclic rule to gain more universal acceptance, it would have to undergo further verification.

Question b.2. *Please comment on strengths and limitations of this mechanistic interpretation for selecting chemicals in the training sets and for interpreting the observed binding data.*

The Panel concurred with the mechanistic interpretation of ER-binding used by the Agency as consistent with the structural characteristics of the ER-binding domain discussed in the literature (*e.g.*, Katzenellenbogen *et al.*, 2003 and references therein) and informs testable hypotheses of how chemicals bind to the ER.

The Panel provided several strengths of the mechanistic interpretation of the ER-binding.

1) There is abundant knowledge to support the chemical basis for defining the chemical structure space(s).
2) There is a good understanding of the energetic and steric characteristics of the ERα binding domain.
3) The mechanistic interpretation focuses on chemistry with specific biological relevance.

The Panel noted several limitations of the mechanistic interpretation of the ER-binding.

1) Current knowledge largely addresses ER-binding at the A and B sites and not other sites. The expert system may miss some compounds that bind to the ER via other binding sites not included in the model.
2) "Mixed phenols" rules need further development and will involve further testing. The same criticism was stated for the "mixed organics" rules.

Question b.3. *Please comment on the clarity of the White Paper in describing the differences in (Q)SAR development when the goal is to predict in vitro ER receptor binding from chemical structure vs. when the goal is to predict in vivo reproductive/developmental responses from chemical structure. Please indicate if additional discussion in the White Paper is needed to establish the relevance of ER binding affinity (either measured or predicted) to interpret the potential for in vivo outcomes.*

The Panel agreed that the ER-binding adverse outcome pathway illustrated in Figure 1 of the White Paper is transparent and mechanistically sound. The Panel stressed that the White Paper makes it very clear that the goal of the expert system is to provide for the use of "the molecular initiating event as a basis for prioritizing chemicals for further screening with EDSP Tier I assays, which incorporate endpoints at higher levels of biological organization" (see page 12 of the White Paper). While the prediction of *in vivo* reproductive/developmental responses from *in vitro* responses is very tantalizing, the Panel concurred that ER-binding will not necessarily translate into predicting these effects *in vivo*. The Panel emphasized that ER-binding is only one of many key events that might lead to reproductive impairment. DNA activation events that occur downstream from ligand binding, agonist/antagonists or partial agonist/antagonists are also critical to the manifestation of reproductive effects.

Panel members offered some suggestions to improve the clarity of the White Paper with regard to the relevance of ER-binding and *in vivo* reproductive/developmental effects.

1) Repeatedly stress that the expert system currently only allows prioritization of FI and AM pesticides for EDSP Tier 1 testing.
2) Add a discussion of how the mechanistic/mode of action approach can be used in mixture assessments, and how this information can be partnered with ongoing research in EPA's Office of Research and Development (ORD), *i.e.*, genomics/dosimetry.
3) Add some examples of non-receptor mediated estrogenic activity to the White Paper as ER-binding activation may not pick up all estrogenic compounds, *e.g.*, nonylphenols which have several non-ER mechanisms of estrogenic activity.

c. Model Domain

Question. *Please comment on the adequacy of the approach that was used to select chemicals for the training sets in terms of these two inventories.*

The Panel agreed that the iterative approach to select chemicals for the trainings sets of these two inventories was adequate and congratulated EPA for conducting this important work. This approach involved strategic selection of chemicals within chemical groups, assessment of the results in the context of the ER-binding

hypotheses, and testing additional chemicals until sufficient information is collected to make a prediction for all the chemicals in the inventory. The Panel suggested that the large body of literature data on ER-binding could be used in a meta-data analysis format either for developing or validating the decision rules implemented in the ER-binding expert system.

d. Algorithm

Question d.1. *Please make suggestions for improvements in presenting the expert rules and their underlying rationale, especially with regard to groups with multiple functional groups.*

The Panel made several suggestions for improvements in presenting the expert rules and their underlying rationale, especially with regard to groups with multiple functional groups.

1) **Mixed phenols.** The Panel recommended developing a systematic testing program based on, for example, Simplified Molecular Input Line Entry Specification (SMILES) or Smiles Aritary Target Specification (SMARTS) strings. A SMILES or SMARTS string could be used as an input into something like a classification tree methodology which could then be used to determine if certain "structures" could be used to predict activity level.
2) **Mixed organics.** The Panel thought that a small number of mixed organic compounds could act as "skeleton keys" to indicate binding to the ER.
3) **Substructure coding.** The Panel indicated that EPA could pick out substructure characteristics that would be important by just looking at the structure, *e.g.,* bisphenol A-like, dichlorodiphenyltrichloroethane (DDT)-like, chemicals with a long fatty acid tail, and phthalate-like chemicals.
4) **Nature of hydrogen (H) bonds.** The Panel suggested that EPA more closely examine chlorinated compounds/pesticides and the nature of H bonds.
5) **Decision rules.** Three suggestions were made by the Panel to clarify the presentation of the rules.
    a. The first suggestion was to use an automatic approach to derive the decision rules.
    b. The second suggestion was to construct the expert system decision tree as chemical classes using log Kow modifier rules.
    c. The third suggestion would be to have decision rules that center on chemical properties. Properties of the sub-classes beyond log Kow would determine relative binding affinity (RBA) activation class.

Question d.2. *Please also comment on the ability of the expert rules to identify chemicals outside the model domain.*

In general, the Panel agreed that the expert system is expected to perform reasonably well for certain chemicals outside the model domain. Some panel members presented examples of a limited number of chemicals that appear to be exceptions to some of the rules, especially the acyclic one. The Panel commented that the model domain may mean different things for different people and that better definitions are needed in the White Paper. EPA has defined the model domain by chemical classes and log P limits on the class. As a consequence of using the class-wise approach, the model domain is very broad. The Panel suggested one way to increase the expert system's utility for untested chemicals in other chemical inventories would be to convert the seven specific decision or domains in the expert system to seven chemical property rules, which would better describe the "model domain" as limits on chemical properties. The Panel remarked that the "acyclic rule" is well-described in the Agency's White Paper for the two regulatory inventories of interest; however, it may be difficult to extend this rule to other chemical inventories because of the likelihood of different structural activity relationships.

The Panel encouraged EPA to further evaluate the expert system's predictivity using large literature data sets of ER binding and/or the generation of new relevant data to include a wide variety of chemicals representing all the chemical classes or subgroups within the inventory in question, and at the same time, covering the relevant physico-chemical property ranges of each class or subgroup. The Panel indicated that the EPA had implemented these strategies for the two specific inventories in question, *i.e.*, FI and AM pesticides. Chemical selection might be based on the presence of organic functional groups in particular structural environments. This can be accomplished by developing a systematic testing program based on, for example, use of SMILES or SMARTS strings (see above).

e. Goodness-of-fit, Robustness, and Predictivity

Question e.1. *Please comment on the adequacy of information presented in the White Paper to evaluate the scientific rationale of how a chemical is processed through the decision logic; i.e., how a chemical is assigned to a subgroup with an associated binding affinity value; the mechanistic rationale for estimates of binding affinity data, including data for related chemicals; and how it is determined that a chemical is outside of the domain of the expert system.*

As a general principle, all panel members agreed that the information in the White Paper was adequate to evaluate the scientific rationale of how a chemical is processed through the decision logic. That is, the Agency addressed, in detail, all of the descriptive elements needed to provide a clear and transparent description of the expert system, *i.e.*, how a chemical is assigned to a subgroup with an associated binding affinity value; the mechanistic rationale for estimates of binding affinity data, including data for related chemicals; and how it is determined that a chemical is outside the domain of the expert system. The

Agency's description of the expert system conforms to the OECD Principles of (Q)SAR validation for goodness-of-fit, robustness, and predictivity. The Panel noted, however, that the Agency's expert system is not a (Q)SAR model by the traditional understanding of regression-based or discriminant analysis model (Q)SARs. Statistical methods such as those used to assess QSARs based on regression models are not applicable to the Agency's expert system. Therefore, the description of the expert system in the White Paper reflects this difference and focuses on the transparency and validity of the decision logic.

Question e.2. *While to date the Agency is not aware of statistical approaches that would provide the means to assess goodness-of-fit or predictivity of expert systems such as the one described here, is the SAP aware of any statistical approaches or data presentations that could be amenable for such evaluations?*

The Panel agreed that the rule-based expert system for predicting the activity of chemicals is not a classical regression-based or discriminant analysis regression or discriminant analysis (Q)SAR model. Traditional statistical methods used to measure goodness-of-fit for these types of models are not applicable in the case of a rule-based expert system. The Panel agreed that other types of statistical evaluation of the expert system would provide additional confidence to the user. They stated that one method of statistical evaluation could be an assessment of predictivity for compounds not in the training set. The validation data for compounds not in the training set, but within the applicability domain of the model, could be obtained from the literature and can be used to assess the predictivity of the expert system. The Panel thought that using misclassification rates, as well as sensitivity and specificity of the validation data, would provide further confidence of goodness-of-fit in the predictivity of the expert system.

f. Transparency and Clarity of the Expert System

Question f.1. *Please provide any additional comment on how well the White Paper's summary of the expert system conforms to the OECD validation principles and provide suggestions, as appropriate, to enhance the clarity or transparency of the expert system's development and intended use with regard to the validation principles.*

The Panel concurred that the Agency's expert system conforms to the five OECD Principles of (Q)SAR validation which include demonstration of:

1) a well defined endpoint,
2) mechanistic interpretation of the model,
3) defined domain of model applicability,
4) an unambiguous algorithm, and
5) appropriate measures of goodness-of-fit, robustness, and predictivity.

Overall, the Panel commended EPA on their delineation of the basis for the expert system, especially in a context suitable for regulatory risk assessment. The Panel recommended that the expert system be modified as new data become available.

The Panel thought the clarity of the White Paper could be improved by adding a glossary to define terms such as "model domain." The Panel also suggested that each of the seven "models" which make up the expert system should be discussed separately in the context of their applicability domain, training set, etc. See also suggestions provided by the Panel in response to question 1) f.2.

The Panel thought the transparency of the expert system could be improved by linking the training sets, especially their molecular structures, to each of the seven model domains. This would be particularly helpful if done in an electronic format linked with automation of the expert system. Transparency would be further enhanced with some form of validation or verification exercise (see discussion in section 1) e. They recommended that this exercise be performed for each of the seven models in the expert system as some are better documented than others. One panel member recommended that a standard reporting format, such as (Q)SAR Modeling Report Format (QMRF), would be considered as a transparent method to record information regarding the rule-based expert system, including results of any validation. The QMRF information is structured according to the OECD (Q)SAR validation principles.

Question f.2. *Please provide any suggestions for preparing the system documentation that will enhance clarity and understanding for users.*

The Panel complimented the Agency on the level of documentation of the ER binding expert system. Panel members recommended some areas for improvement, including:

1) reorganization of the White Paper. This specifically related to the information starting with page 30, "A Rule-Based Exert System to Predict ER Binding Affinity." They agreed there are essentially seven rules (*i.e.*, I-VII in Figure 9) and there should be a separate discussion for each of the seven models in context of their applicability, domain, training set etc.
2) that an Executive Summary to the White Paper would be very useful to the reader.
3) that a glossary be added to the White Paper to clarify terms such as "model domain."
4) that the two-sided arrows in Figure 1 of the White Paper be changed to one-way arrows to reflect the current understanding of the flow of responses in the Agency's expert system.
5) that EPA amplify the specific criteria by which the number of compounds tested in each of the subgroups were selected as these were only briefly described in the White Paper as being "a wide variety" selected "across the respective log Kow ranges" of each subgroup.
6) the development of illustrative training aids would provide greater clarity and understanding of the expert system, *e.g.*, animation of Figure 9 in the White Paper.

7) that automation of the expert system, including more detailed description of the individual parts, would be critical to increase availability and acceptance of the system.

2. Acyclic Compounds

Question a. *Please comment on the extent to which the finding with acyclic compounds in the FI and AM inventories may be broadly applicable to other acyclic compounds. Suggestions on an approach to empirically and efficiently assess a hypothesis that acyclic compounds will not bind to the ER in other chemical inventories would be welcomed.*

The Panel concluded that the findings for acyclic compounds in the FI and AM inventories could be extended to most acyclic compounds in other inventories, but not to all. Acyclic compounds may bind to the ER by different mechanisms. These compounds may bind to receptors at binding sites other than A and B. In a broad sense, the Panel explained, these acyclic compounds may bind directly to the ER at sites outside of the main ligand binding domains, affect binding of 17β-estradiol (E2) at the receptor site, affect structural changes related to receptor activation or co-activator protein interactions, and/or affect other downstream actions related to transcriptional activation.

They indicated that there are several databases that might prove to be useful to inform the expert system, *e.g.,* Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) database of both positive and negative compounds tested *in vitro* for endocrine receptor (ER) and androgen receptor (AR) binding and studies conducted under the Tox21 federal partnership. Tox21 studies will ultimately test approximately 10,000 chemicals for transcriptional activation and antagonism of ERα and ARs), and a dozen other nuclear receptor using high throughput assays (Collins *et al.,* 2008). The Panel thought that there may be an automated way to test for fragments that might be able to bind to the ER receptor, *e.g.,* Analog Identification Method (AIM). Some panel members suggested one way to quickly address the chemical structural requirements for binding to these receptors would be to use combinatorial chemistry, as done in the pharmaceutical industry. This is done by formulating similar chemical structures with different substructures to explore the binding space both singly and as known mixtures. The Panel cautioned that special attention should be given to halogenated acyclic compounds.

Question b. *Please comment on the extent to which the finding with acyclic compounds in the FI and AM inventories can be applied to other nuclear steroid receptors in general. Suggestions on an approach to empirically and efficiently assess a hypothesis that acyclic compounds will not bind to the androgen receptor would be welcomed.*

The Panel concluded that it would be difficult to extrapolate the finding for acyclic compounds in the FI and AM inventories to other nuclear steroid receptors because of different chemical structural-binding interactions expected for other receptors, *e.g.*, other nuclear receptors, such as thyroid hormone receptors, ARs and membrane receptors. The Panel commented that thinking about other nuclear receptors is a valid extension of the work completed by EPA on the ER. The Panel encouraged EPA to study the structure activity relationships for these other receptors with the same logic and thoroughness that was used for soluble ERs, *i.e.*, defining the optimum structure for receptor binding and activation. For the greatest efficiency, acyclic compounds could be tested using these assays as high throughput screens for both ER and AR binding. Some panel members suggested that test compounds could be synthesized by combinatorial methods to aid study of structure binding relationships of other steroid receptors as noted above and that one possible structural motif of interest would be halogenated long chain molecules.

3. Prioritization for EDSP Tier 1 Screening

Question. *Based on the characteristics of the (Q)SAR-based expert system presented in the White Paper, please comment on the Agency's view that the expert system could be employed to support "sorting and prioritizing" food use inert ingredients and antimicrobial pesticides for EDSP Tier 1 screening.*

The Panel was in strong agreement that the expert system is likely to be an important tool to support "sorting and prioritizing" FI and AM pesticides for EDSP Tier 1 screening by the ER α binding mechanism. Considerable and careful effort to cover the structural domains of these two regulatory inventories was invested in development of the expert system.

As designed, the expert system separates chemicals into two bins defined by the limit of detection of the ER binding assay, *i.e.*, 0.0001%. Chemicals with a RBA>0.00001%, *i.e.*, "the positives", would be assigned a higher priority for EDSP Tier 1 testing and chemicals whose RBA< 0.00001%, *i.e.*, "the negatives", would be assigned a lower priority for testing. The Panel indicated that the "negatives" should be characterized as chemicals in which prediction of ERα binding is not possible. The Panel suggested that sorting and prioritization aspects of the expert system might be enhanced if each of the seven model rules were weighted separately. Different weights would be given to positive events resulting from each rule.

The Panel stressed that the ER binding mechanism is only part of the estrogenic mode of action, and the estrogenic mode of action is only one component of the endocrine system. Other aspects of steroid metabolism and action pathways could also be important for prioritization for endocrine disruptor screening (see Figure 1). Therefore, the expert system should be seen in the context of the overall EDSP, i.e., use a weight of evidence approach for integrating information coming

from many possible sources including the results of various endocrine mechanism and mode of action tests, modeling such as the rule-based expert system for ER-binding, as well as other information related to the reactivity and potency of the chemical.

4. Cross Species Applicability

Question. *Given what is reported in the literature and similarities between human and rainbow trout ER binding affinity observed thus far in the research described in the White Paper, please comment on the extent to which use of an expert system based primarily on trout ER binding affinity data is a reasonable effects component for prioritizing food use inert ingredients and antimicrobial pesticides for EDSP Tier I screening.*

The Panel responded to this question in two different ways.

1) Some panel members commented in a broad sense about the extrapolation of trout cytosolic ER binding data "across species" and were not confined to comparisons between the rainbow trout and mammalian ERs. These comments were meant to consider the ecological implications of screening tests in addition to the human health implications. Overall, the Panel concluded that currently available information suggests that the trout cytosolic ER assay can be used to develop reasonable predictions of the RBA of xenobiotics for ERα across species. However, what is unclear at the present time is whether the trout cytosolic ER assay can be used for estimating RBAs of xenobiotics across all ER subtypes and all species.

2) Other panel members discussed how the trout cytosolic ER assay may fit into prioritization schemes for endocrine disruption screening tests primarily from the viewpoint of human health. In this latter context, the Panel thought that trout cytosolic ER and human ER assays would likely yield comparable results. However, in extrapolating a negative result from the trout cytosolic ER assay to human systems, there would be some level of uncertainty because some differences in performance have been reported between cytosolic and recombinant ER assays (EPA's White Paper, page 38; and associated presentations).

DETAILED RESPONSES TO CHARGE QUESTIONS

1. Evaluation of the Expert System in the Context of the Organization for Economic Cooperation and Development (OECD) Validation Principles

As discussed in the Preface and Introduction of the White Paper, EPA's development of the Quantitative Structure Activity Relationship (QSAR)-based expert system was guided by the OECD principles for (Q)SAR validation. The five principles include demonstration of:

- a well defined endpoint,
- mechanistic interpretation of the model,
- defined domain of model applicability,
- an unambiguous algorithm, and
- appropriate measures of goodness-of-fit, robustness, and predictivity.

A prototype of the expert system was the subject of an OECD-convened peer consultation in February 2009, at which time the system was evaluated using the (Q)SAR validation principles. Based on input from this peer consultation, the Agency further refined the expert system, particularly as it related to the OECD validation principles. The components to Charge Question 1 address specific issues concerning the (Q)SAR validation principles and the subject expert system in the context of its use to determine the order in which chemicals, i.e., food use inert ingredients and antimicrobial pesticides, will be subject to Tier 1 screening under EPA's EDSP, *i.e.*, for prioritization.

a. Biological Endpoint

The biological endpoint that is predicted by the expert system is relative binding affinity to the cytosolic rainbow trout estrogen receptor. Based on preliminary studies it was anticipated that food use inert ingredients and antimicrobial pesticides would have low relative binding affinities. In addition, an evaluation of the two respective inventories indicated a wide range of structures and associated physical-chemical properties (*e.g.*, solubility, Kow, etc.). Consequently, assay methods used to measure binding affinity to establish the training set were designed to detect low levels of binding affinity (*e.g.*, testing to solubility in binding assays and cytotoxicity or solubility, as appropriate, in slice assays). Confirmatory binding studies and transactivation assays were employed to systematically verify that apparent low affinity binding levels represented competitive displacement and translated to ER-mediated transactivation.

Question a.1. Please comment on the methods employed and their adequacy for detecting and measuring ER binding affinity for the compounds in the two chemical inventories of immediate interest.

Panel Response

The Panel was in general agreement that the two assays employed for detecting and measuring ER binding affinity for the compounds in the two chemical inventories were satisfactory. The methods provide a logical approach to evaluating compounds with a specific mechanism of action (ER-binding) and then linking this to Vtg transcriptional activation and Vtg production. The Panel stated that it was imperative to note that binding or Vtg induction should not be equated with higher order effects.

The rainbow trout ER binding assay is based on cytosol preparations from livers of individual immature rainbow trout, *Oncorhynchus mykiss*. In brief, tissues are homogenized in a Tris buffer, pH 7.6 (Denny *et al.*, 2005) equal to 20 times the weight of the liver tissue, and centrifuged to collect cytosolic fractions. Competition assays are based on maximum binding capacity for 5 nM for tritiated Tritiated 17β estradiol [[^{3}H]-E2] and are used to determine relative binding affinity constants for test chemicals. Free [^{3}H]-E2 is separated from bound [^{3}H]-E2 to calculate how much is displaced by the addition of test chemicals. In addition, non-specific binding is subtracted from the total bound to calculate the amount of specifically bound [^{3}H]-E2. The Panel agreed that the performance of the assay is standard and well documented in the scientific literature including researchers working for the Agency (Denny *et al.*, 2005). The Panel concurred that the Agency used sufficient replication, appropriate positive and negative controls, and the assay results were highly reproducible.

The basis of the trout liver slice assay is to use liver slices prepared from immature male rainbow trout. About 100 slices can be obtained from one liver. The slices are incubated in multi-well plates in incubation medium and the idea is look for the ability of E2, or test chemicals, to bind and activate ERs resulting in induction of Vtg mRNA and the expression of Vtg. The Panel thought that the use of this assay, in conjunction with the ER-binding assay, was a very good approach for detecting and measuring ER-binding and transactivation capacity for the compounds in the two inventories of interest. The Panel, therefore, deemed this combination of assays was highly satisfactory to validate the (Q)SAR expert system with the goal of prioritizing chemicals in the two chemical inventories, *i.e.*, FI and AM pesticides, for Tier 1 testing under the EDSP.

The main advantage of the liver slice assay is that it is used to look for activity of the estrogen receptors. The tissue slices are composed of intact tissues and are therefore closer to the function of liver *in vivo*. In addition, the Panel indicated that it was possible for compounds to be metabolized in these assays, which may be important if the active chemical is actually a metabolite of the parent compound tested. The Panel stated that the Vtg gene and resulting translated protein are accepted biomarkers of estrogen activity through soluble ERs. Both Vtg mRNA and its translated Vtg protein are well studied. It is known that the

promoter for Vtg mRNA contains estrogen response elements that would respond to active ERs, *i.e.*, ERs that have bound with E2 or an E2-mimic. The assay has a wide dynamic range with Vtg mRNA expression levels sometimes exceeding a thousand-fold that of the control depending on the concentration of estrogen or estrogen mimics.

The Panel enumerated several positive aspects of both assays.

1) The molecular event, which is based upon physicochemical data for receptor binding, is linked to subsequent confirmation of this response through transcriptional ER activation.
2) The cytosolic method simultaneously evaluates the binding of all soluble subtypes. The Panel thought it might be necessary to differentiate the binding affinities for each soluble receptor subtype particularly for interspecies calibration (see Question 4).
3) The assays are relatively easy to perform in a high throughput manner. Many cytosol fractions can be obtained from a few fish for the binding assay and about 100 slices can be obtained from one liver for the transactivation assay.
4) Tissue integrity is maintained; thus, the transactivation results should mimic *in vivo* changes more directly.

Nevertheless, the Panel pointed out several limitations of both assays.

1) The binding assay does not distinguish agonists from antagonists, as both would bind to the ERs. However, this limitation is removed by coupling the binding assay with the tissue slice assay, which will distinguish these activities.
2) The binding assay may underestimate the true binding affinity of the chemicals for ERs because the actual concentration of ER is not known in the cytosolic fractions and because other proteins in the fraction may interfere with the assay in non-specific ways. Other assays have been developed to measure binding affinities that are more closely indicative of "true" binding affinities, but these rely on test systems with recombinant ERs.
3) Neither of the assays distinguishes binding or activity of specific ER isotypes. Rainbow trout liver contains both ERα and ERβ isotypes and there are two forms of each of these isotypes for a total of four ERs (Nagler *et al.*, 2007). The Panel stated that it was not clear how many and which of the two isoforms of each nuclear ER subtype in rainbow trout is present in the cytosolic preparation, and therefore the contribution of each ER to the measured binding is unknown.

   The Panel indicated that distinguishing among binding activity of specific ER isotopes is not a trivial point. Significant differences have been reported for some species in the binding affinity of ERα and ERβ to

xenobiotics and even to natural ligands such as E2. In channel catfish, for example, ERβ binds E2 with 20-fold greater affinity than ERα (Xia *et al*, 2000; Gale *et al.*, 2004). Also, the RBA's of xenobiotics, such as 4-nonylphenol, can differ by nearly two orders of magnitude between the two catfish receptor subtypes (Gale *et al.*, 2004). Thus, selection of chemicals based on cut-off RBA values for further study could be potentially influenced one way or the other depending on which receptor subtype is used for the binding assay.

The Panel identified additional reasons why a clarification of the contribution of different ERs to the ligand binding activity of cytosol extracts would be desirable. Information provided by the White Paper suggested that the binding activity of trout cytosol behaves similarly to the rainbow trout ERα. In a study using selective agonists and antagonists of ERα and ERβ, Leaños-Castañeda and Van Der Kraak concluded that despite the fact that most ligand binding activity present in rainbow trout liver cytosol may be due to ERα under certain conditions, ERβ was responsible for mediating the vitellogenic response of the trout liver to E2 (Leaños-Castañeda and Van Der Kraak, 2007). The Panel concluded that the trout cytosol assay that was used to develop the training set of binding data and the trout tissue slice assay that was used to confirm the biological relevance of the binding data may be measuring different estrogen receptors (ERα and ERβ, respectively). In general, while both subtypes, in balance, contribute to overall estrogenic activity, ERα activation is more implicated with hazard while ERβ is implicated in protection against hyperproliferation and carcinogenesis (reviewed in, e.g., Jacobs, 2005; Macpherson *et al.*, 2006; Carruba 2007; Ellem and Risbridger, 2009; Wilson and Westberry, 2009; Bian *et al.*, 2001). Most of the Panel thought that it might be important to know their relative ratios. However, two panel members did not think that this limitation was a problem for using the assays as a screening tool.

4) ER subtype profiles may change throughout the life history and gender of the animal. The Panel suggested that identification of the relative contributions of each subtype at various life stages would reduce uncertainty.
5) The liver based assays may not be representative of action of ERs in other endocrine tissues, which may have a different composition of ER subtypes (e.g., brain and gonad, where the ratios of the two isoforms differ).
6) The two assays completely neglect membrane receptor binding (*e.g.*, G-protein-coupled receptor 30 [GPR30]).
7) Slice-to-slice variability in the liver slice assay may be high depending on the exact location in the liver core for each of the slices, *e.g.*, they may vary in hepatocyte content. However, the Panel indicated that this limitation can be overcome by performing a higher number of replicates.
8) Livers of individual fish may vary in their ER concentrations especially since E2 (and other estrogenic compounds) appears to be able to up-

regulate ERα This last point, noted by the Panel, would also be a source of variance in juvenile fish according to gender.

9) Neither of the assays measures effects of estrogen-like compounds on membrane receptors, which feed into important signaling pathways that may be linked to estrogenic effects.

In summary, the overall conclusion of the Panel was that the two assays were appropriate for helping to set testing priorities of the FI and AM chemical inventories. Having both assays together showed a logical progression in defining the toxicity pathway, *i.e.*, combining binding of a chemical to the ER, activation of the receptor in a sensitive tissue that can also metabolize chemicals. In terms of the current OECD conceptual framework for endocrine disruptor testing, the two assays represent both level 1 and level 2 testing. However, given the limitations of the assays enumerated above, a negative result of a chemical in the selected assays should not be construed as meaning that the chemical was not an endocrine disruptor or endocrine active substance. The Panel recommended that EPA consider adding more discussion in the White Paper to better define the limitations of the assays used to generate binding and biological response data (*i.e.*, what can and cannot be concluded from these data).

Question a.2. In the context chemical prioritization for Tier 1 screening, please comment on the decision to measure binding affinity up to the maximum concentration based on the solution properties of the chemical, rather than using ligand concentration 'cut-off' values of -4 Log Molar to -3 Log Molar, which have typically been used to conclude a compound does not bind to the ER.

Panel Response

There was general agreement among the panel members that the chemicals of concern (FI and AM pesticides) would be of low binding affinity. The Panel indicated that these chemicals should be tested using concentrations up to their solubility in the media. Thus, the maximum concentration tested per chemical would vary depending on their solubility and this solubility would vary with the log Kow of the test chemicals. This flexibility in testing means that false negatives would be reduced.

The Panel stated that the cut-off values of -4 log Molar to -3 log Molar were established in the pharmaceutical industry as a screen for strong estrogens and may not be applicable for screening and prioritization of FI and AM pesticides. They commented that attention should be given to possible overt toxicity at high concentrations. The Panel recommended that concentrations tested should be in line with concentrations that are biologically relevant. As a screening and prioritization mechanism, "cutoff" values should be set more broadly to be as inclusive of chemicals that might be able to bind to ER's so that they can be further tested using more in depth assays.

Question a.3. Please also comment on how any in vivo studies that are available for compounds with low receptor binding affinity could be used to provide a relative binding affinity 'cut-off' value either alone or in combination with cut-off values based on the maximum solubility of a ligand in the buffer solution.

Panel Response

Panel members interpreted this charge question in different ways. As a result, the Panel provided several *in vivo* and *in vitro* methods in response to this question.

1) **Distribution of compounds within tissues -- dosimetry**
   Fathead minnow adults (male or female) and embryo tests could be used to determine whether or not compounds with low receptor binding affinity are bioaccumulated in tissues of endocrine importance, for example, liver, brain, pituitary, gonad, and kidney. Such *in vivo* assays can be performed by either adding the test chemicals to the water or mixed in the food. This information on tissue distribution could be useful to determine whether assaying high concentrations of the test chemicals is essential.

2) **Vitellogenin induction experiments**
   Vitellogenin induction, *in vivo*, has been generally accepted as a good biomarker for compounds that bind to and activate ERs in fish. The assay is very sensitive and is performed by exposing fish to estrogen or estrogen mimics in the water for 7 days (Brian et *al.*, 2005), or by mixing the compounds with the food. This assay can be used to show general concentration cut-offs for a single compound. For mixtures of compounds that act through the same mechanism of action, it is possible to show additivity of effects even when each individual compound is below its No Observed Effective Concentration (NOEC). This assay can also be used to measure anti-estrogenic compounds if used in female fish. Thresholds of toxicological concern have been discussed and developed for estrogenic agonists in aquatic environments (Crane *et al.*, 2009). These thresholds could be used to determine cut-off values for test chemicals in aquatic environments using regulatory priority lists and existing databases.

3) **MCF-7 cell proliferation assays**
   These assays depend on using a MCF-7, a human breast adenocarcinoma cell line, to monitor cell proliferation following exposure to estrogen or to estrogen-like substances. The assay is very sensitive to estrogen and estrogen-like susbstances, but does not work solely through the ERα as other estrogenic pathways are involved.

4) **48-hour transcriptomics experiments**
   Microarrays are now available for a large number of fish species that are routinely tested for toxicity experiments. For example, there are arrays now for rainbow trout, zebrafish, medaka, three-spined stickleback and

fathead minnow, largemouth bass among others. Recent research has shown that a 48 hour exposure of fish to chemicals is sufficient to induce gene transcription in various endocrine tissues including brain, liver, and gonad. This assay would not just measure changes based on estrogen receptor binding, but would evaluate changes in toxicity pathways of concern for all endocrine pathways including other sex hormone receptors and pathways that alter the concentration of endogenous hormones (either synthesis or metabolism). There are definite thresholds for gene transcription. Some genes are more sensitive than others. More genes are changed with higher concentrations. Test compounds can be administered in the water or mixed in the fish food. Genes that are important for reproduction, growth and susceptibility to disease are now known and this assay could be used to determine whether any of these are affected by low concentrations of test chemicals mixed in the food at levels that would be normally experienced by humans.

Transcriptomics (and proteomics) could also be accomplished on the fish liver slices to look for estrogen fingerprints and to determine what other pathways are being affected by the exposure. If funds are not a problem, the microarray studies can be coupled to fish recrudescence assays (21-day fish reproductive assays) (see #6 below).

5) ***In vivo* test for ova-testis**
This assay would look for downstream effects of exposure to estrogenic compounds. The end point would be histological determination of ova-testis.

6) **Fish recrudescence assay**
Cost effective *in vivo* studies would include fish recrudescence assays (Tier 1, *i.e.*, 21-day fish reproductive assay) that partner well with the *in vitro* approach using the rainbow trout estrogen receptor. These are already approved for use.

7) ***In vitro* assays with other endpoints**
In vitro assays could be conducted with regard to other biological endpoints, *e.g.*, estrogen sulfotransferase, aromatase, and other receptors.

8) **Other *in vivo* test guidelines.** Other *in vivo* test guidelines such as the "enhanced" OECD conventional 28-day repeat dose toxicity test (OECD TG 407) and the updated draft of physiologically-based pharmacokinetic (PBPK) test guidelines (when available) would also be useful.

9) **Physiologically-based pharmacokinetic (PBPK) modeling**
Dosimetry PBPK studies would allow estimates of hepatic concentrations that could be partnered with concentrations identified from *in vitro* responses (ER-binding/Vtg mRNA induction).

Given the occurrence of natural ligand-receptor interactions and in vivo metabolism, the prediction of *in vivo* responses from *in vitro* data currently available may be quite difficult or not currently possible. Compounds with low receptor binding affinity may cause *in vivo* endocrine responses by non-receptor mechanisms that involve modulation of the natural ligand for the receptor. Examples would include interactions with gonadotropin release, hormone biosynthesis, or impairment of hormone clearance. Extrapolation from *in vitro* data would need to take this into account (reviewed in Jacobs *et al.*, 2008).

b. Mechanistic Interpretation

Question b.1. Numerous studies have established the alignment of estrogen and other high affinity ligands within the ER binding domain and indicate that a distance of 10.2 to 11A between the two H-bonding sites and stable (non-flexible) ring structure is optimal for binding. These and other studies lead to the assumption that acyclic compounds would not bind to the ER, although a systematic analysis across a diversity of structures was not available in the literature. In the current study 25 acyclic compounds across 10 classes present in the inventories were evaluated in the training set and none were found to bind at a Relative Binding Affinity (RBA) detection limit of 0.00001%.

Please comment on the adequacy of this training set for supporting the expert system's rule that acyclic compounds do not bind to the ER.

Panel Response

General comments:

The intent of OECD (Q)SAR Validation Principle 2 (*i.e.*, "mechanistic interpretation of the model") is to make sure that consideration is given to the possibility of a mechanistic association between both the descriptors used in a model and the endpoint being predicted, and to make certain that this association is documented in the expert system.

To meet the mechanistic principle three questions need to be addressed:

1) In the case of a SAR, is there a description of the molecular events that underlie the reactivity of the molecule? In this case is there a description of how substructural features could form part or all of a receptor binding region?
2) In the case of a (Q)SAR, do the descriptors have a physicochemical interpretation that is consistent with a known mechanism (of biological action)?

3) Are any literature references cited in support of the purported mechanistic basis of the (Q)SAR?

For the ER-binding expert system, the Panel answered all three of these questions in the affirmative.

Specific comments:

The ER-binding training set for acyclic compounds includes many of the major chemical classes (*e.g.*, alcohols, amines, aldehydes, carboxylic esters, and ketones) and covers the structural domain of the two inventories (*i.e.*, FI and AM pesticides) evaluated in this exercise. The Panel determined that the current acyclic compound training set is adequate for the structures within the two inventories, especially for the intended purpose of screening (*i.e.*, prioritization). The Panel, however, considered the current training set to be inadequate to cover all chemicals. To illustrate this point, the Panel noted that the acyclic compounds, perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA), were shown to bind ER and induce Vtg production in at least two fish species (Ishibashi *et al.*, 2008; Liu *et al.*, 2007).

The Panel thought that for the acyclic rule to gain more universal acceptance, it would have to undergo further verification. Part of this verification might be addressed by examining test data from other species (*e.g.*, human) and other protocols (*e.g.*, yeast assays). The Panel noted that these data would eventually need a peer reviewed intelligent testing strategy, where adequate coverage of the chemical space is balanced with the need to move away from a chemical by chemical testing scheme. Computational chemistry and other techniques could assist in chemical selection.

The Panel stated that the question of whether all acyclic compounds do not bind to the ER (acyclic rule) is part of the general issues concerning the use of category-based assessments. Panel members generally perceived that regulators are much more likely to accept a positive prediction than a negative one so there would be a natural tendency to be cautious about the general rule that all acyclics are negative ER-binders. The Panel noted that moving from a chemical by chemical evaluation to a category based evaluation will reduce testing, especially animal-based testing. As a consequence, the issue of accepting (or not) a prediction of non-binding for a chemical will be quite important.

Question b.2. Based on studies by Katzenellenbogen *et al.* (2003 and references therein) a working hypothesis in developing the training set was that compounds in the inventories of interest could bind at one site; *i.e.*, the A site or the B site, based on the presence of a hydrogen bond donor or acceptor substituent. The development of the training sets and the resultant ER expert system use a chemical hierarchy based on different binding mechanisms (*i.e.*, A-B binding sites, A binding site only, B binding site only).

Please comment on strengths and limitations of this mechanistic interpretation for selecting chemicals in the training sets and for interpreting the observed binding data.

Panel Response

Strengths

The Panel agreed with the prevailing knowledge that chemicals which bind to the ER form a group of compounds with an important mechanism that can alter processes involved in reproduction. The Panel agreed that the ER-mediated adverse outcome pathway provides strength to the mechanistic interpretation. This pathway clearly demonstrates how ER binding is linked through a series of measurable events at different levels of biological organization to a hazard of regulatory concern (alteration in reproduction). The Panel indicated that as chemicals likely to initiate or hinder the ER-mediated pathway are identified, the pathway will be useful for generating testable hypotheses at various levels of biological organization along the pathway.

The Panel reviewed the four OECD (Q)SAR validation criteria for a mechanistic pathway:

1) It has to have chemistry which has biological relevance.
2) The critical chemical, cellular, and other *in vitro* measurements must have *in vivo* relevance, especially the molecular initiating event.
3) It should have identified events along the pathway which provide means of eliminating chemical from further testing.
4) In the end, the pathway or at least the critical events along the pathway must be simple enough to be manageable, but complex enough to be useful.

They concluded that the two biological endpoints from the two *in vitro* assays used as the basis for this expert system, clearly meet these pathway criteria. One assay was optimized to measure the potential of chemicals to bind rainbow trout estrogen receptors and initiate the ER-mediated pathway, and a second assay measured gene activation through production of Vtg in rainbow trout liver slices to confirm that ER binding translates to an effect further along the pathway.

Based on the literature, the absence of hydrogen bonding group(s), or inappropriate molecular geometry and molecular dynamics can explain the failure of chemicals, especially aromatics, to bind to ER. The Panel agreed that the understanding of the energetic and steric characteristics of the ER binding domain was important to establish the mechanistic basis for defining a chemical structure space(s) associated with ER-binding. The mechanistic interpretation is supported by a clear chemical basis for receptor binding, transparency of the process, and chemistry with specific biological relevance.

The Panel agreed that the mechanistic interpretation of ER binding used by the Agency is consistent with the structural characteristics of the ER binding domain discussed in the literature (*e.g.*, Katzenellenbogen *et al.*, 2003 and references there-in) and informs testable hypotheses of how chemicals bind to the ER. Theories of chemical interaction in the various "sub-pockets" within the ER have been presented by other investigators (*e.g.*, Jacobs *et al.*, 2003, 2004; Vedani *et al.*, 2005; Hartman *et al.*, 2009).

Historically, the ideas of chemical interaction(s) with the ER have been based on "lock and key" information gained from both the ER and steroidal structures. As described in EPA's White Paper, the "lock" focuses on two hydrophilic pockets (*i.e.,* A and B) within a specified distance of each other. The "key" is the complement, *i.e.*, a chemical with two hydrogen-bonding groups within the same specified distance. The Panel also pointed to widely known information that chemicals which contain only one hydrogen-bonding group (*e.g.*, such as alkylphenols, alkyloxyphenols, and parabens) that all mimic the A-ring of E2, are active in human ERα binding assays as well as transcriptional activation assays using bioreporter- and human- ERα. These data describe specific structure-activity relationships in docking models and that potency is related to size and shape functionalities as illustrated by conformational models. These findings are indicative of the contributions in ER binding from the B-and C-rings of E2. The A-site binding results from the rainbow trout estrogen receptor binding assay used in the Agency's expert system are consistent with the literature. This consistency further strengthens the stated mechanistic interpretation.

The Panel noted that ER-mediated gene activation for alkylanilines has been reported; however, the mechanistic interpretation of binding at the B-site has not been specifically described. The structural-activity relationships involving the B-site binding may be extended to other chemicals such as phenones and cyclohexenols and is a further strength of the mechanistic interpretation.

Other strengths noted by panel members include the mechanistic interpretation which allows prediction of a biological response (ER-binding and transcriptional endpoints) from physicochemical features of chemicals, which are relatively simple to obtain. The pathway allows for these *in vitro* responses to be hypothesized to impact higher biological levels of organization (*i.e., in vivo* effects), but empirical confirmation is necessary. The tissue slice assay also allows metabolism to be included for pro-estrogens.

A final strength is the use of mode/mechanism of action information which will allow for the eventual evaluation of mixture scenarios. For example, if multiple estrogenic compounds are present in a system, overall estrogenic activity through ligand binding equivalents may be used to quantify effect.

See also discussion in the Panel's response to charge question 1) b.3.

Limitations

The Panel stated that the major limitation of the expert system was its inability to assess ER binding in other chemical (non-FI and non-AM) inventories. The Panel noted that the "mixed phenols" rules need further development and will involve further testing. The same criticism was stated for the 'mixed organics" rules.

Another limitation of the expert system noted by the Panel is that current knowledge largely addresses ER-binding at the A- and B-sites and not other sites. The Panel stated that they could not rule out anti-estrogenic activity resulting from binding of compounds, particularly acyclic ones, at other sites. The Panel thought that other chemical inventories may not be encompassed by the existing "training set" as the structural binding rules for the expert system were specifically designed to predict relative ER binding affinity for the two inventories in question, *i.e.*, FI and AM pesticides. They indicated that the "training set" used to estimate relative ER-binding affinity had no uncertainty built into it, in other words, it was perfect. Other chemicals (*e.g.*, PFOS/PFOA) would provide uncertainty in predicting relative binding affinities as they might not conform to the rules of this expert system. While much is known about ERα, for example, the receptor is "dynamic and plastic," less is known about the chemical interactions with other ER subtypes. The Panel agreed that additional studies would be necessary to reduce these uncertainties for other ER subtypes and other chemical inventories.

Question b.3. While ER binding can be an initial step in a toxicity pathway leading to adverse reproductive outcomes (see Figure 1 in the White Paper), the ER binding data in the training set, and the associated expert system, were not designed to predict *in vivo* responses. Rather the expert system was designed to predict relative ER binding affinity from a chemical's structure to support the prioritization of food use inert ingredients and antimicrobial pesticides for *in vitro* and *in vivo* Tier 1 screening, which is designed to ascertain if a compound has the potential to interact with the estrogen system.

Please comment on the clarity of the White Paper in describing the differences in (Q)SAR development when the goal is to predict in vitro ER receptor binding from chemical structure vs. when the goal is to predict in vivo reproductive/developmental responses from chemical structure. Please indicate if additional discussion in the White Paper is needed to establish the relevance of ER binding affinity (either measured or predicted) to interpret the potential for in vivo outcomes.

Panel Response

The Panel concurred that the ER-binding adverse outcome pathway illustrated in Figure 1 of the White Paper is transparent and mechanistically sound. Figure 1 provides a good representation of how mode/mechanism of action can be used to

better understand and predict higher order effects (*i.e.,* reproductive effects). The Panel stressed that the White Paper makes it very clear that the goal of the expert system is to provide for the use of "the molecular initiating event as a basis for prioritizing chemicals for further screening with EDSP Tier I assays, which incorporate endpoints at higher levels of biological organization" (see page 12 of the White Paper). The Agency went on to say that "It is important to recognize that while the determination of potential adverse reproductive effects in whole organisms or populations is an important risk assessment issue, it is not a goal of a prioritization application" (page 14 of the White Paper). The Panel had no trouble in understanding that the current expert system only allows prioritization for EDSP Tier I testing. EPA's presenters did a nice job of pointing this out.

The Panel agreed with statements in the White Paper that ER-binding or Vtg induction may not indicate *in vivo* adverse effects. While the prediction of an *in vivo* reproductive/developmental response from chemical structure is very tantalizing, the Panel concurred that ER-binding does not necessarily translate to *in vivo* effects on reproduction. Likewise, lack of ER-binding does not necessarily translate to lack of *in vivo* effects on reproduction. The OECD consultation came to this same conclusion. The Panel emphasized that ER-binding is only one of many key events that might lead to reproductive impairment. DNA activation events that occur downstream from ligand binding, agonist/antagonist or partial agonist/antagonist, are also critical to the manifestation of reproductive effects. *In vitro* assays that give a greater level of information are those such as transactivation assays, and those that have more metabolic competence. Currently there is a lack of systematic experimental results for model chemicals representing the different sub-domains of ER-binding at higher levels of biological organization including gonadal histopathology and developmental assays in zebrafish embryos.

The Panel stated that more data would be needed should the Agency wish to explore the relevance of *in vitro* ER-binding to adverse *in vivo* reproductive effects. The Panel noted that oviparous vertebrates provide an excellent model to allow mechanism/mode of action linkages between ER binding and potential reproductive impairment and population level impacts (Kidd *et al.*, 2007). Panel members recognized that while ER binding is not the only way that chemicals can cause *in vivo* reproductive effects, they agreed that this was an excellent place to start. Concordance analyses between data generated from *in vitro* ER assays and the uterotrophic assays, for example, could provide the basis of a mode of action linkage to *in vivo* effects. Recent and ongoing European Union (EU) projects concerning mixtures and estrogenic responses could provide data to support such a development for fish (e.g., see Brian *et al.*, 2005; Brian *et al.*, 2007; and CREDO Cluster (the cluster of research into endocrine disruption in Europe): http://ec.europa.eu/research/endocrine/projects_clusters_en.html).

As the pathway proceeds from the molecular initiating event (*i.e.,* ER-binding), through cell and tissue level gene transcription and translation (as measured via

the liver slice assay), and continues through organ effects to an adverse outcome observed in the individual or population, the *in vivo* effects can be measured. Among these effects is the conversion of testicular tissue to ova tissue measured morphometrically from histological section in whole fish assays (Zha *et al.*, 2007) and reproductive impairment of fish resulting from feminization of male fish (Seki *et al.*, 2003) especially when exposed during early life stages (Van Aerle *et al.*, 2002). To date, the majority of the documented *in vivo* reproductive impairment effects are for compounds with high or moderate ER-binding affinities. The Panel noted that the USEPA's Office of Research and Development, Mid-Continent Ecology Division, National Health and Environmental Effects Research Laboratory in Duluth, Minnesota has ongoing studies of *in vivo* reproductive impairment effects. Results from these efforts are not reported in the White Paper and have yet to be peer-reviewed and published. However, the Panel believed that these efforts will provide useful information regarding the likelihood of *in vivo* adverse outcomes from exposure to chemicals (*i.e.*, chemicals are associated with either A- or B-site binding, but are not necessarily in the FI or AM pesticide regulatory inventories) with low ER-binding affinities.

Panel members offered some suggestions to improve the discussion of prioritization based on ER-binding affinity in the White Paper.

1) Add a discussion of how the mechanistic/mode of action approach can be used in mixture assessments, and how this information can be partnered with ongoing research at ORD *(i.e.*, genomics/dosimetry).
2) Add some examples of non-receptor mediated estrogenic activity to the White Paper as ER-binding activation may not pick up all estrogenic compounds (*e.g.*, nonylphenols which have several non-ER mechanisms of estrogenic activity).
3) Stress that this model currently only allows prioritization of FI and AM pesticides for EDSP Tier 1 testing. The presentations did a nice job of pointing this out, but the White Paper was not as clear in this respect.

c. Model Domain

The domain of the current ER expert system includes rules to support predictions of ER binding affinity for chemicals in the food use inert ingredient and antimicrobial pesticide inventories.

Please comment on the adequacy of the approach that was used to select chemicals for the training sets in terms of these two inventories.

Panel Response

The Panel agreed that the iterative approach to select chemicals for the trainings sets of these two inventories was adequate and congratulated EPA for conducting this important work. The iterative approach involves strategic selection of

chemicals within chemical groups, assessment of the results in the context of the ER-binding hypotheses, and testing additional chemicals until sufficient information is collected to make a prediction for all the chemicals in the inventory. Chemicals in the training set or assay domain should include the expert system's ER binding domain and be representative of the structures in the regulatory inventories or the chemical domain. Equally important to defining the structural limits of activity is to define the physicochemical limits (*i.e.*, solubility) of the activity. The Panel noted that both of these approaches were used in developing the positive binding applicability domain of the expert system, especially for chemicals with low binding affinity. However, the Panel pointed out that the level of adequacy depends on specific decision rules, because at a specific decision point, the number and representativeness of the training set chemicals may be different. The Panel stated that the scientific community would be interested in assessing compounds outside of the FI and AM inventories.

The Panel concurred that the Agency clearly stated in its White Paper that the intended regulatory use of the ER-binding expert system was as a priority-setting tool. That is, identifying ER-binding chemicals from the FI and AM pesticide inventories that would be placed at a higher priority for EDSP Tier 1 screening. The FI inventory contains 393 discrete chemicals and the AM inventory contains 211 discrete chemicals. The regulatory applicability of the ER-binding expert system to FI and AM inventories is addressed by subdividing the FI and AM chemicals into respective subgroups that have specific chemical attributes that are mechanistically related to the different modes of ER binding.

Some panel members questioned whether the model domain was adequate to identify antagonists of ER-binding both in the binding pocket and outside the binding pocket. They indicated that other data need to be included (see Blair *et al.*, 2000; and Hartman *et al.*, 2009). As an example, the expert system might be inadequate for identifying the pure ER antagonist, ICI 182,780, (an experimental compound to study estrogenic effects), or the partial antagonist Tamoxifen® (a drug used in treatment of estrogen receptor-positive metastatic breast cancer in postmenopausal women). For example, with potent ER antagonists, such as ICI 182, 780, the tail of the molecule binds outside the main binding pocket (Jacobs *et al.*, 2003). In response to a question from the Panel, EPA clarified that they used ICI 182, 780 in the liver slice assay and hoped to provide that information soon.

The Panel appreciated the evolving nature of the development of an expert system as new data and knowledge become available. However, they emphasized that any predictive computational model including the ER-binding expert system should be "frozen" at some point during the development process so that its performance can be appropriately evaluated with additional chemicals not used in the development of the model before regulatory use. The Panel was unclear as to how many discrete chemicals from the FI and AM inventories had been tested in the rainbow trout ER binding assay, and how many of them were used in training and validating the expert system. They suggested that the White Paper would be

more readable and informative if these numbers were explicitly stated in the nodes of the decision tree shown in Figure 9 and other figures of the White Paper. The Panel noted that EPA agreed to incorporate this suggestion in the White Paper.

The Panel suggested that the large body of literature data on ER binding could be used in a meta-data analysis format either for developing or validating the decision rules implemented in the ER-binding expert system. The Agency built the expert system using a single set of data, *e.g.,* the rainbow trout ER-binding data to avoid "noise." Other ER-binding data from different sources and different protocols from the literature could be used to test the confidence of the expert system. The Panel pointed out that meta-analysis will provide more information and confidence in the binding domain of the model.

d. Algorithm

The ER expert system provides predictions for each chemical, with each individual prediction traceable to chemical subgroups, binding mechanism and endpoint databases. In developing the expert system several chemical subgroups were identified as chemicals that contain multiple functional groups.

Question d.1. Please make suggestions for improvements in presenting the expert rules and their underlying rationale, especially with regard to groups with multiple functional groups. The Panel made several suggestions for improvements in presenting the expert rules and their underlying rationale, especially with regard to groups with multiple functional groups (enumerated below). Chemicals with multiple functional groups were termed "mixed phenols," if there was a ring hydroxyl group, or "mixed organics," if there was no hydroxyl group.

Mixed phenols. The Panel recommended developing a systematic testing program based on, for example, Simplified Molecular Input Line Entry Specification (SMILES) or Smiles Aritary Target Specification (SMARTS) strings. A SMILES or SMARTS string could be used as an input into something like a classification tree methodology which could then be used to determine if certain "structures" could be used to predict activity level. Mixed phenols are potential A-site binders because of the presence of the phenolic group; however, the presence of additional substituents can hinder binding activity. As described in the White Paper (page 29), there is insufficient information to establish general rules for these types of chemicals. The Panel stated that the conservative approach employed by EPA of an exact match is appropriate. Clearly, improvement in the expert rules is predicated on additional testing.

Mixed organics. While there is a mechanistic rationale for additional testing for mixed phenols, no such rationale is apparent for mixed organics. The fact that some mixed organics were active on the first reading may discourage further development of structural alert-based expert systems. The Panel thought that a

few mixed organics would act as "skeleton keys" to bind to the ER. However, these chemicals should be weak binders and bind only at concentrations near their cytotoxicity level.

Substructure coding. The Panel indicated that EPA could pick out substructure characteristics that would be important by just looking at the structure, *e.g.*, bisphenol A-like, DDT-like, chemicals with a long fatty acid tail, and phthalate-like chemicals. Chemicals with multiple functional groups would have further coding of the substructures which would be key coders for receptors.

Nature of hydrogen (H) bonds. The Panel suggested that EPA more closely examine chlorinated compounds/pesticides. The nature of the hydrogen bonds (*e.g.*, formed by the hydroxyl group), important for binding in the receptor, could be investigated to determine the effect of adjacent substituents and the strength of hydrogen bonding and distances for hydrogen bonding required for binding.

Decision rules. An expert system consists of a set of decision rules. The order in which they appear in the expert system is important in determining the system's prediction performance and the interpretability of the prediction results of new chemicals. The development of the decision rules and the order in which they are placed in the current ER-binding expert system is a reasonable manual process, mainly based on human experts' knowledge and interpretation of the experimental ER binding data. Three suggestions were made by the Panel to clarify the presentation of the rules.

1) The first suggestion was to use an automatic approach to derive the decision rules. This would be done through the following steps: (1) Describing all training set chemicals in terms of the decision values (descriptors) currently used in the expert system, *e.g.* log Kow and the present/absence of a structural element (cycle, hydroxyl group etc.). (2) Assign each chemical as an ER binder or a non-binder. (3) Use partitioning methodology such as a classification tree (Breiman *et al.*, 1984) to derive the decision rules automatically.

The "automatic" approach will allow for more objective assessment of the performance of the system and the implicit "coding" for chemicals with multiple functional groups. The "manual" and "automatic" versions of the expert system should have a lot of similarity (in terms of the types of rules), confirming to some degree the validity of the current, manually-built expert system. The final implementation of the expert system could be a hybrid of the "manual' and "automatic" versions of the expert system, both in terms of the actual rules and the order in which the rules appear in the decision tree.

2) The second suggestion was to construct the expert system decision tree as chemical classes using log Kow modifier rules. One panel member noted that there seems to be two kinds of "rules" used in the system. Rules I and

II discuss chemical properties that result in low RBA. Rule III describes properties that result in high RBA. The rest of the decision tree is a mixture of these two types of "rules."

3) The third suggestion would be to have decision rules that center on chemical properties. Properties of the sub-classes beyond log Kow would determine RBA activation class. The Panel noted that a model based on these decision rules would provide greater insight about whether a subclass would be considered "ER-binders" or "non ER-binders"

Question d.2. Please also comment on the ability of the expert rules to identify chemicals outside the model domain.

In general, the Panel agreed that the expert system is expected to perform reasonably well for certain chemicals outside the model domain especially in the context of A-site and B-site binding. Panel members commented that the model domain may mean different things for different people and that better definitions are needed in the White Paper. EPA has defined the model domain by chemical classes and log P limits on the class. As a consequence of using the class-wise approach, the model domain is very broad. The Panel remarked that EPA uses only a limited set of characteristics to describe the chemicals. Solubility, as measured by log Kow, does not by itself contain sufficient information to predict the activity level of all chemicals of interest in the training set. The model may perform well to predict the activity of similar chemicals from other inventories, but also may require additional characteristics not used in the current model if the goal is to accurately predict the activity level of all chemicals from these other inventories.

The Panel recommended that a series of definitions be added to the White Paper to improve readability, *e.g.*, model domain, applicability domain. For example, model domain should be clearly defined in the following statement found on p. 13 of the White Paper, "The chemicals in the training set make up the expert system's model domain." One suggestion was that the model domain would be defined better by the descriptions of the molecules.

The Panel encouraged EPA to evaluate the expert system's predictability using large literature data sets of ER binding. Panel members noted that in most of the literature, the chemicals were not tested in concentrations as high as those used in the current White Paper. Inconsistency between the prediction of the expert system and literature data, therefore, could be due to the different limits of detection, and not due to the lack of predictivity of the expert system. The Panel stated that the scientific community will use the expert system when it becomes publicly available for chemicals outside the model domain and beyond those in the two inventories. Therefore, the expert system released to the public should be consistent with as much available data as possible.

The Panel remarked that the scientific and regulatory communities are more reluctant to accept negative predictions (*e.g.*, non ER-binding) than positive predictions (*e.g.*, ER-binding). Therefore, they recommended that the assayed domain (the chemicals tested) includes proper coverage of the inventory being probed. This coverage has to be balanced with the need to move away from a chemical by chemical testing scheme. The best balance is reached by testing a wide variety of chemicals, which represent all the chemical classes or subgroups within the inventory in question and at the same time cover the relevant physico-chemical property ranges of each class or subgroup. The Panel indicated that this strategy has been implemented for the two specific inventories in question (*i.e.*, FI and AM), which forms the basis of EPA's White Paper. For example, 70 alkylaromatic sulfonic acids in the two inventories were represented by six tested chemicals spanning a log Kow range from -0.62 to 5.67. The Panel noted that the Expert System could be used to assign an RBA classification to new chemicals without full structural characterization if the user determines it is sufficiently similar to a chemical used in the training set. Some on the Panel felt that this use has a number of problems, one of which is that it is based on a determination that does not require assessment of structural similarity. The utility of this approach decreases as the chemical molecule becomes more complex.

The challenge, as noted by the Panel, is how this approach is extended to all chemicals. One way of doing this is to implement chemical selection based on the presence of organic functional groups in particular structural environments. This can be accomplished by developing a systematic testing program, based on, for example, use of SMARTS or SMILES strings (see explanation above in 1) d.1).

The Panel remarked that the "acyclic rule" is well-described in the Agency's White Paper for the two regulatory inventories of interest; however, it may be difficult to extend this rule to other chemical inventories because of the likelihood of different structural activity relationships. An integrated testing strategy would be needed for the greatest efficacy. See additional discussion in the Panel's response to question 2.

The Panel agreed that the expert system is a knowledge-based system in which there are seven specific decisions or domains. Each of these domains requires a specific set of rules, each of which should be used separately to judge within or outside the "model domain." Similarly, one panel member remarked that the expert system could be viewed as a hierarchically-ordered set of seven models, resulting in the model domain space as seven "clusters" in multidimensional space. Using this illustration, this panel member suggested the seven chemical class rules could be converted to seven chemical property rules, which would increase the expert system's utility for untested chemicals in other chemical inventories and better describe the "model domain" as limits on chemical properties.

e. Goodness-of-Fit, Robustness, and Predictivity

Consistent with suggestions by the EDSTAC (1998), and typical processes for (Q)SAR development, the expert system rules were established through an iterative process of defining subgroups, gathering empirical data to refine subgroup rules, followed by collection of additional empirical data to cover the structural domain and/or until a consistent pattern of structural rules and activity emerged. The expert rules permit each chemical to be assigned to subgroups and an associated estimated binding affinity value, accompanied by an explanation of the basis for the estimate as well as how the estimate compares to measured data for other members of the same subgroup. The 2009 OECD expert consultation report on the expert system recognized that standard statistical methods such as those used to assess regression model (Q)SARs are not necessarily applicable to expert systems (OECD, 2009). Rather transparency and usefulness as described in the White Paper are more appropriate parameters for assessing the validity of an expert system. The peer consultation report found the current approach, with individual predictions traceable to chemical subgroups, binding mechanism and endpoint databases, to be appropriate although the report noted that if additional information could be made available it would facilitate future peer-review on this issue.

Question e.1. Please comment on the adequacy of information presented in the White Paper to evaluate the scientific rationale of how a chemical is processed through the decision logic; i.e., how a chemical is assigned to a subgroup with an associated binding affinity value; the mechanistic rationale for estimates of binding affinity data, including data for related chemicals; and how it is determined that a chemical is outside of the domain of the expert system.

Panel Response

As a general principle, all panel members agreed that the information in the White Paper was adequate to evaluate the scientific rationale of how a chemical is processed through the decision logic. That is, the Agency addressed, in detail, all of the descriptive elements needed to provide a clear and transparent description of the expert system, *i.e.*, how a chemical is assigned to a subgroup with an associated binding affinity value; the mechanistic rationale for estimates of binding affinity data, including data for related chemicals; and how it is determined that a chemical is outside of the domain of the expert system. The Agency's description of the expert system conforms to the OECD Principles of (Q)SAR Validation. The Panel noted, however, that the Agency's expert system is not a (Q)SAR model by the traditional understanding of regression or discriminant analysis model (Q)SARs. Statistical methods such as those used to assess regression model QSARs are not applicable to the Agency's expert system. Therefore, the description of the expert system in the White Paper reflects this difference and focuses on the transparency and validity of the decision logic.

The Panel indicated that there were a number of very good reasons for providing the details of the scientific rationale to the user. Foremost is that the operation of the expert system is clear and transparent to the user. The detailed explanation of the scientific rationale behind the decision logic, in turn, provides the user with confidence in the expert system.

The Panel addressed the adequacy of the information in the White Paper related to: 1) how a chemical is assigned to a subgroup with an associated binding affinity value, 2) the mechanistic rationale for estimates of binding affinity data, including data for related chemicals, and 3) how it is determined that a chemical is outside of the domain of the expert system.

1) **How a chemical is assigned to a subgroup with an associated binding affinity value**

   The Panel concurred that the considerable amount of information available in the White Paper is adequate for the intended purpose. The presentation of the information in the White Paper, whilst complex, is comprehensive and easy to understand and follow for any person with a reasonable understanding of chemical structure. A user should be able to recognize the structural features underlying the issue and appreciate how and why the conclusions have been made. That is, detailing why a subgroup has been utilized and why a compound has been assigned to a subgroup. The Panel agreed that the educated user would be able to understand the process, which is described in full beginning with page 30 in the White Paper. The use of Figures 9-12 in the White Paper was considered to be particularly useful. The Panel remarked that the information provided is consistent with the current state of the art in this area. One panel member commented that recognized structural features could be used in some fashion to automate the expert system to rapidly screen inventories, based on, for example, use of SMARTS or SMILES strings.

2) **The mechanistic rationale for estimates of binding affinity data, including data for related chemicals**

   The Panel concurred that the considerable amount of information available in the White Paper is adequate for the reader to understand the mechanistic rationale for estimates of binding affinity. The Panel agreed that a complete description of the mechanistic rationale is very important to understand the expert system. A complete description of the rationale provides the non-expert user with all necessary information to grasp how and why the expert system was developed and how it makes predictions. The user will therefore have more confidence in the model. Comfort in the mechanistic rationale for the expert system would be useful to screen other chemicals within/outside the training set. The Panel agreed that the mechanistic rationale stems from a deep understanding of the binding of xenobiotics to the ER. They noted that this

point is discussed in detail in response to question 1) b.2. In particular, they noted the detailed information on several aspects of the mechanism including the differentiation of ER-binding affinity at A-site, B-site or or C-site and how this specifically relates to the concept of chemical grouping and the physico-chemical basis of mechanism.

The Panel recognized that the mechanistic rationale is for ERα and not the entire estrogenic mode of action because other pathways are not included. There are other receptors that may play a role in endocrine disruption. The White Paper provides a good description of the role of ERα, but it was noted that this is one of several receptors that are responsible for endocrine disruption. The role of receptors and non-receptor mediated responses in endocrine mechanisms and modes of action, other than ER, should be stated more clearly in the White Paper.

The Panel thought that EPA provided a strong justification for the mechanistic relationship between binding affinity and hydrophobicity (as parameterized by the logarithm of the octanol-water partition coefficient, log P). Log P is a measure of hydrophobic binding interaction as opposed to binding interaction at the A-, B-, or C-sites which is "electrostatic." The Panel recommended that the context of log P used in the expert system be more explicitly stated as relating to hydrophobic binding interactions to prevent confusion with the context of log P relating to uptake, distribution, and bioavailability of a chemical used in ecotoxicology studies.

The Panel agreed that the White Paper adequately described ER-binding as the molecular initiating event for the ER-binding mediated adverse effects pathway. The description in the White Paper was clear that ER-binding affinity cannot be translated in to adverse *in vivo* effects. The Panel stressed that this point be made very clear in the White Paper and should be re-enforced wherever possible.

The Panel recommended that the EPA should provide the ER-binding and Vtg expression data to support the information behind each of the chemical groups when the expert system is implemented. This will allow the user to see the potency level of the prediction and the distinction between the chemicals with RBA greater or less than 0.0001%. The Panel noted that utilization of the source data with the chemical grouping concept, *i.e.*, within the categories formed by the structural rules, could allow for semi-quantitative estimations of binding potency to be made.

3) **How it is determined that a chemical is outside of the domain of the expert system.**

The Panel concurred that the considerable amount of information available in the White Paper is adequate for explaining how a chemical is determined to be

outside of the domain of the expert system. The Panel agreed that the identification of whether a chemical is outside of the domain of a model is very important for its proper use and application. Before a chemical is determined to be in or outside the domain of the expert system, the model domain must be clearly defined. Here the domain of the model is defined by the single organic chemical structures in the two chemical classes or "bins." Those chemicals outside of the domain are termed "Unknown Binding Potential." The Panel recommended a possible note of caution regarding the phrase "Unknown Binding Potential" in case of possible misinterpretation. A more explicit phrase, such as "Prediction of the binding potential is not possible as the compound is outside of the domain of the model," whilst not as succinct, may be preferable. The Panel commended EPA on the use of solubility cut-offs to define the domain of the ER binding assay implicitly (also see detailed discussion in 1) a.2).

The Panel noted that there are many definitions of domains used in this scientific area and in the White Paper. They recommended that the Agency provide additional definitions and clarification of terms in the White Paper, in particular, with regard to each domain (also see detailed discussion in 1) d.2. For example, the domain(s) of (Q)SAR models can be defined on a number of levels including: the domain of the assay itself, physico-chemical descriptor range, structural fragments, mechanism of action, and metabolic domain. The Panel noted that not all of these domain definitions are necessary for every model including the expert system.

Question e.2. While to date the Agency is not aware of statistical approaches that would provide the means to assess goodness-of-fit or predictivity of expert systems such as the one described here, is the SAP aware of any statistical approaches or data presentations that could be amenable for such evaluations?

Panel Response

The Panel noted that there are a large number of diagnostics and descriptive statistics available for regression-based (*e.g.,* r2, F, etc.) or discriminant (or logistic regression) analysis (*e.g.,* Cooper's statistics for sensitivity and specificity) QSARs. Many, if not all, of these are in the OECD documentation [http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)2] and also supported by the guidance provided by the European Chemicals Agency [http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?vers=20_08_08]. Such types of statistical measures provide estimates of "goodness-of-fit", i.e., how much of the variance in the effect data is described by the descriptor data.

The Panel agreed that the rule-based expert system for predicting the activity of chemicals is not a classical (Q)SAR. Never before has there been an opportunity to develop an expert system for a defined and reasonable number of chemicals

with both a strong mechanistic basis to activity and natural groupings of chemicals. They indicated that this situation is an ideal opportunity to make an expert system inclusive of these defined and restricted inventories (*i.e.,* FI and AM pesticides). This is the first example of a regulatory authority developing a method for the optimization of an ER binding expert system for public domain use. Statistical methods used to measure goodness-of-fit are not applicable in the case of a rule-based expert system. The Panel agreed that the unique nature of the expert system and its development must be borne in mind and some other kind of statistical evaluation is needed.

External validation is recommended in order to obtain a true estimate of the predictivity of a model. This requires predicting the effects for compounds that are not in the original training set, but which have similar structural space and applicability domain for assessing the accuracy of the predictions. Thus, the Panel recognized that typically in model building, one starts with two groups of data, a training set and a validation set. The training set is used to build the model and the validation set is used to test the model. In the expert system presented, all of the training data are used to fit the model and the model has been restructured and optimized so that the ER-binding activity of each tested substance is fit perfectly by the final model. No data are available to test or validate the model. Therefore, the Panel remarked that an assessment of goodness-of-fit or adequacy of model predictivity using the training data provides a very optimistic view of the quality of the final model. The Panel concluded that the types of statistics associated with goodness-of-fit are not applicable to the expert system.

The Panel agreed that some other kind of statistical evaluation of the expert system is needed to provide confidence to the user. They stated that the statistical evaluation should be an assessment of predictivity for compounds not in the training set. The validation data can be used to truly assess the predictivity of the expert system.

For instance, the Panel thought that using misclassification rates, sensitivity, and specificity of the validation data, would provide a good measure of goodness-of-fit for the predictivity of the expert system. The Panel provided an illustration of how this would work. The expert system is essentially a decision tree that takes inputs, *i.e.*, chemical class and chemical properties, and ends with a classification into two ER-binding activity classes (*i.e.*, RBA <0.00001% or RBA > 0.00001%). The expert system is therefore comparable to a multiple logistic regression which is based on binary classification, in this case RBA <0.00001% or RBA > 0.00001%. Therefore, the methods to assess goodness-of-fit of a multiple logistic regression model would be applicable. Many statistics for goodness-of-fit are derived from the difference between the observed value and the predicted value of the model, or the residual value. However, in the case of the expert system, the equivalent would be the misclassification rate.

The Panel agreed that the expert system should be frozen at a point of time before any validation analysis is attempted. This would mean that the expert system is first clearly and unambiguously defined, as in the White Paper, and not updated with new information. Next, predictions with new data could be made for data by the expert system without biasing the outcome, *i.e.,* RBA classification, and a true assessment of system's performance could be achieved.

The Panel recommended that one possibility to investigate the performance of the expert system would be to use data from the literature to assess its predictions. Panel members recognized that there are caveats in using this approach. There are inherent differences in the test methods used and level of binding specified. Despite these concerns, the Panel thought the approach would be illustrative. A particular advantage of using external validation data to test the predictability of the expert system is the ability to determine clearly when a chemical is out of domain. This will prevent studies from being biased by predictions for compounds out of domain.

The Panel agreed that care and effort was required to choose training set data (chemicals) to ensure the expert system was capable of spanning the domains of interest. Similarly, data for the validation set should also be chosen across all domains of interest to challenge the model. For example, Figure 7 in the White Paper shows all of the sulfonic acid dyes that were in the rainbow trout training set and then some representative additional structures in this class. Only a few of the training set chemicals were chosen to be tested for ER-binding activity. The expert system is then able to "predict" RBA for all chemical structures, tested or not. One possibility is a validation data set selected from chemicals in either inventory which had not been tested. These chemicals would be assayed for ER-binding activity and then results would be compared to the expert system's predictions. The Panel commented that there would be obvious cost and resource implications in this approach if many chemicals were tested as part of the validation.

Panel members indicated that a possible source of active ingredient pesticide chemicals for validation purposes might be the European Union list 4 (Link for regulatory status of pesticide active substances in the EU: http://ec.europa.eu/food/plant/protection/evaluation/index_en.htm.) There might be chemicals not tested in the FI and AM pesticide inventories for which there are existing data that could to be used for validation In addition, the Panel suggested comparing the overlay between the FI and AM pesticide inventories and the OECD high production volume chemical (HPVC) list. The HPVC list contains approximately 7000 chemicals and could provide suggestions for chemicals and directions for examining model validation, evaluation of model extrapolation potential and directions for further extension of the model to cover the expanded interest domain. This would help to reduce the number of situations where the model is not able to make a prediction due to the compound being out of the domain.

The Panel agreed that it would be very difficult to define how many compounds would be needed for validation. There are different levels of confidence in each of the seven different models/rules in the expert system. Some models may perform very well with a limited number of chemicals tested; thus they may not need such detailed analysis. Other rules may have lower performance and uncertainty; thus, more chemicals would be required to be tested. Therefore, the Panel specified that the number of chemicals tested to validate the expert system would change according to the rules. There may also be regulatory priorities for different rules; thus, there may be a need to create prioritized rules with the appropriate information regarding uncertainty.

The Panel stressed that the evaluation process of the model is an assessment of the performance of the model, not an extension of the model. As noted above, the model should be frozen while the rules are validated. The validation process could be used to examine the extent of the boundaries of the rules in the expert system.

The difficult task is determining how much the domain can be expanded without losing predictability. For this expert system, different rule domains offer different abilities for expanding its domain. For example, phthalates between log Kow < 1.3 are classified in the group RBA < 0.00001% and those above log Kow >1.9 are classified in the group RBA >0.00001%. Understanding chemical and/or biological mechanisms of ER binding might allow for the extension of the model utility to cover the 1.3<log Kow <1.9 range for this rule. In this expert system, category membership is based on substructure and physico-chemical property cut-offs which were developed and verified with trout ER-binding test data acquired in an iterative process and which were hypotheses driven. The greater the number of chemicals evaluated to form or delineate a category, the greater the confidence in the applicability domain of that category, *i.e.,* the more valid the category. The category can also be validated with data from other test systems, in particular looking for similar relative potency and thresholds. The other test systems include not only the liver slice assay used here in conjunction with the trout ER-binding, but also other ER-related assays such the yeast-based ones which are necessarily ER-mediated (see Schultz *et al.*, 2002, Hamblen *et al.*, 2003; Senseverino *et al.*, 2009) and ER transactivation assays (reviewed in Jacobs *et al.*, 2008).

f. Transparency and Clarity of the Expert System

Question f.1. In its validation principles, OECD recognized the importance of a transparent validation process for the development of (Q)SAR models in order to further enhance their regulatory acceptance.

Please provide any additional comment on how well the White Paper's summary of the expert system conforms to the OECD validation principles and provide

suggestions, as appropriate, to enhance the clarity or transparency of the expert system's development and intended use with regard to the validation principles.

Panel Response

General Comments on Clarity and Transparency

Overall, the Panel commended EPA on their delineation of the basis for the expert system, especially in a context suitable for regulatory risk assessment. Specifically, this includes explicit description of the decision tree's logic for grouping of chemicals and provision of the expert rules on which estimates are based, along with reference to empirically measured data for chemicals in the associated training set. This type of transparency is critical as a basis for acceptance and application of the expert system. which represents a knowledge set from a given point in time. Science marches on and new data will be generated. As this occurs, the Panel recommended that the expert system be modified.

Generally the text flows well in the White Paper. The writing is clear and important information is well-illustrated by relevant figures and tables. The Panel commented on the excellence of the separate document of illustrative case studies.

The Panel thought the clarity of the White Paper could be improved by adding a glossary to define terms such as "model domain." They also suggested that each of the seven "models" which make up the expert system should be discussed separately in the context of their applicability domain, training set, etc. See also suggestions provided by the Panel in response to question 1) f.2.

The Panel thought the transparency could be improved by linking the training sets, especially their molecular structures, to each of the seven model domains. This would be particularly helpful if done in an electronic format linked with automation of the expert system. Transparency would be further enhanced with some form of validation or verification exercise. See discussion in section 1)e. The Panel recommended that this exercise be performed for each of the seven models in the expert system as some are better documented than others. One panel member recommended that standardizing the reporting format such as (Q)SAR Modeling Report Format (QMRF) be considered as a transparent method to record information regarding the rule-based expert system, including results of any validation. The QMRF information is structured according to the OECD (Q)SAR validation principles (see QMRF at http://ecb.jrc.it/qsar/qsar-tools/qrf/QPRF_version_1.1.pdf). A separate QMRF would be used for each of the seven models.

Background on the OECD Principles of (Q)SAR Validation

One panel member provided background on the origin of the OECD Principles of (Q)SAR Validation. A number of principles for assessing the validity of (Q)SARs were originally proposed at an international workshop on the "Regulatory Acceptance of (Q)SARs for Human Health and Environment Endpoints," organized by the International

Council of Chemical Associations (ICCA) and the European Chemical Industry Council (CEFIC) held in Setubal, Portugal, on 4-6 March, 2002. The outcome of the meeting was the so-called "Setubal principles" summarized by Jaworska and colleagues (Jaworska *et al.,* 2003). These Setubal principles state that a (Q)SAR should:

1) be associated with a defined endpoint of regulatory importance,
2) take the form of an unambiguous algorithm,
3) ideally, have a mechanistic basis,
4) be accompanied by a definition of domain of applicability,
5) be associated with a measure of goodness-of–fit,
6) be assessed in terms of its predictive power by using data not used in the development of the model.

However, the workshop stopped short of producing any guidance on how to interpret and apply these principles. The OECD Principles of (Q) SAR Validations were derived from the Setubal principles (OECD, 2004).

Agency White Paper and OECD Principles of (Q)SAR Validation

The Panel concurred that the Agency's expert system conforms to the five OECD Principles of (Q)SAR validation which include demonstration of:

1) a well defined endpoint,
2) mechanistic interpretation of the model,
3) defined domain of model applicability,
4) an unambiguous algorithm, and
5) appropriate measures of goodness-of-fit, robustness, and predictivity.

The Panel stated that these principles should be viewed as a means of evaluating the expert system and not a formal validation. The regulatory domain of the expert system contains the two chemical inventories, FI and AM pesticides. Below is the Panel's explanation of how the EPA's expert system met each of the five OECD validation principles.

1) Defined Endpoint Principle

In the case of software programs and expert systems, as is the case with this White Paper, which are based on multiple models, the Panel stressed the importance of identifying the smallest component of the expert system that functions independently, and to apply the principles to the individual component. The intent of Principle 1, a well-defined endpoint, is to ensure clarity in the endpoint being predicted by the given model, or in this case expert system. Because a given endpoint could be determined by different experimental protocols and under different experimental conditions, the Panel noted the importance of identifying the experimental system that is being modeled. In the scientific sense, a defined endpoint is referred to as a specific effect within a specific tissue/organ under specified conditions.

To meet the defined endpoint principle, Principle 1, the Agency must address the following four questions:

a. Does the model have a clearly defined scientific purpose? Does it make predictions of a clearly defined physicochemical, biological or environmental effect?

b. Does the model have the potential to address, or partially address, a clearly defined regulatory need? Does it make predictions of a specific endpoint associated with a specific test method?

c. Is information given about important experimental conditions that affect the measurement and therefore the prediction, *e.g.*, species, exposure period?

d. Are the units of measurement of the endpoint given?

The Panel answered all four of these questions in the affirmative.

2) Molecular Basis Principle.

The intent of this molecular basis principle, Principle 2, is not to reject models that have no apparent mechanistic basis, but to ensure that some consideration is given to the possibility of a mechanistic association between both the descriptors used in a model and the endpoint being predicted, and to ensure that this association is documented.

To meet the mechanistic basis principle, the Agency must address the following three questions:

a. In the case of a SAR, is there a description of the molecular events that underlie the reactivity of the molecule? In this case, is there a description of how substructural features could form part or all of a receptor binding region?

b. In the case of a QSAR, do the descriptors have a physicochemical interpretation that is consistent with a known mechanism of biological action?

c. Are any literature references cited in support of the purported mechanistic basis of the (Q)SAR?

The Panel answered all three of these questions in the affirmative.

3) Defined Applicability Domain Principle

The need to define an applicability domain, Principle 3, expresses the fact that the expert system is a series of reductionist models, which are associated with limitations in terms of the chemical structures, physicochemical properties, for which the system can generate reliable predictions.

To meet the defined applicability domain principle, the Agency must address three questions:

a. In the case of a SAR, is the substructure associated with any inclusion and/or exclusion rules on its applicability to groups of chemicals?

b. In the case of a SAR, is the substructure associated with rules regarding the modulatory effects of the substructure's molecular environment?

c. In the case of a (Q)SAR, are the descriptor and response variables associated with inclusion and/or exclusion rules that define the variable ranges for which the (Q)SAR is applicable (i.e. makes reliable estimates)?

All three of these questions are answered in the affirmative.

4) An Unambiguous Algorithm Principle

The intent of Principle 4, an unambiguous algorithm, is to ensure transparency in each part of the system which generates predictions of an endpoint from information on chemical structure and/or physicochemical properties. Without this information, the performance of each aspect of the system cannot be independently established and this is likely to be a barrier to regulatory acceptance. The issue of reproducibility of the predictions is covered by this Principle.

To meet the defined algorithm principle, the Agency must address one of the two following questions:

a. In the case of a SAR, is there an explicit description of the substructure, including an explicit identification of its substituents?

b. In the case of a (Q)SAR, is the equation explicitly defined, including definitions of all descriptors used?

The Panel responded that only question "a." is applicable and this question was answered in the affirmative.

5) Appropriate Measures of Goodness-of-fit, Robustness, and Predictivity Principle

Principle 5, appropriate measures of goodness-of-fit, robustness, and predictivity, includes the intent of the original Setubal Principles 5 and 6. The Panel noted that this principle simplifies the overall set of principles, but does not lose the distinction between the internal performance of the system, as represented by goodness-of-fit and robustness, and the external validation of the system, as determined by predictivity.

To meet the internal performance of the system principle, principle five, the Agency must address the following questions:

a. Are full details of the training set given, including details of chemical names, structural formulae, CAS numbers, if available, and data for all descriptor and response variables?

b. If the data used to develop the model is based upon the processing of raw data, e.g., the averaging of replicate values, is there an adequate description of the data processing and are the raw data provided?

c. Is there a specification of the statistical method(s) used to develop the (Q)SAR, including details of any software used?

d. Is the (Q)SAR associated with basic statistics for its goodness-of-fit to the training set?

e. Is the (Q)SAR associated with any statistics based on cross-validation or resampling? If yes, is the number or samples used indicated?

The Panel determined that only questions a. and b. are applicable to this expert system and both are answered in the affirmative.

To meet the external validation of the system, principle five, the Agency must address three questions:

a. Does application of the appropriate statistical method(s) to the training set result in the same (Q)SAR model?

b. Is there any information to indicate that the (Q)SAR has been validated previously, using a test set that is independent of the training set?

c. If an external validation has been performed, is the following information available: 1) the number of test structures, 2) the identities of the test structures, 3) the approach for selecting the test structures, 4) the statistical analysis of the predictive performance of the model, *e.g.*, including sensitivity, specificity, and positive and negative predictivities for classification models, and 5) a comparison of the predictive performance of the model against previously-defined quantitative performance criteria.

Traditionally, "validation" within the OECD principles for (Q)SAR validation is thought of in the context of checking a statistical model for predictivity. The Panel stated that this type of statistical validation is not appropriate for this expert system nor is it attempted in the White Paper. [Note: see section 1) e.2 for a parallel discussion.]

Question f.2. The White Paper and associated presentations at the SAP meeting form the basis of the documentation of the expert system.

Please provide any suggestions for preparing the system documentation that will enhance clarity and understanding for users.

Panel Response

The Panel complimented the Agency on the level of documentation of the ER binding expert system. Panel members recommended some areas for improvement:

1) Reorganization of the White Paper. This specifically related to the information starting on page 30, "A Rule-Based Expert System to Predict ER Binding Affinity." They agreed that there are essentially seven rules (*i.e.*, I-VII in Figure 9) and that discussion to assist in their evaluation according to

the OECD Principles for the Validation of QSARs, e.g. making clear references to the training data used, mechanistic rationale and domain covered etc., should be made for each model domain rather than for the expert system *in toto*. Separate discussion for each of the seven models domain would be more transparent.

2) An Executive Summary to the White Paper would be very useful to the reader. This could include a bulleted list of the rule specific domains, mechanistic rationale and domain to prepare the reader for what follows in the body of the paper.

3) A glossary be added to the White Paper to clarify terms such as "model domain."

4) That each of the seven "models" which make up the expert system should be discussed separately in the context of their applicability domain, training set, etc.

5) The two-sided arrows in Figure 1 of the White Paper be changed to one-way arrows to reflect the current understanding of the flow of responses in the Agency's expert system. Currently Figure 1 has two-sided arrows between all of the "response" boxes in the path from xenobiotics through ER binding to altered protein and eventually to population effects suggesting a reversible path. However, the current expert system only describes or uses the one-way argument from xenobiotic to population effects to justify the use of ER-binding assays. In other words, it is a one-way arrow flow of responses.

6) EPA could amplify the specific criteria by which the number of compounds tested in each of the subgroups were selected as these were only briefly described in the White Paper as being "a wide variety" selected "across the respective log Kow ranges" of each subgroup.

7) The Panel thought the development of illustrative training aids would provide greater clarity and understanding of the expert system, *e.g.,* animation of Figure 9 in the White Paper.

8) The Panel stated that automation of the expert system, including more detailed description of the individual parts, would be critical to increase availability and acceptance of the system. This would include direct access to the databases of supporting training sets and related data. For example, EPA could generate a computer-assisted module for the expert system as it has done with "Oncologic" decision tree which predicts genotoxicity from molecular structure (see: http//www.epa.gov/oppt/newchems/tools/oncologic.htm).

2. *Acyclic Compounds*

Acyclic compounds comprise ~58% of the FI and AM pesticide inventories. As discussed in Question 1c, acyclic compounds were found not to bind to the ER. Generally, the absence of hydrogen bonding groups or inappropriate geometry can explain the failure of these chemicals to bind to ER (e.g., see Katzenellenbogen *et al.*, 2003). Prior to the EPA research described in the SAP review, a diverse set of acyclic structures had not been evaluated for ER binding affinity.

> *Question a. Please comment on the extent to which the finding with acyclic compounds in the FI and AM inventories may be broadly applicable to other acyclic compounds. Suggestions on an approach to empirically and efficiently assess a hypothesis that acyclic compounds will not bind to the ER in other chemical inventories would be welcomed.*

Panel Response

The Panel agreed with the general supposition that acyclic compounds in the FI and AM pesticides inventories do not bind to the ER receptor (α and β) and do not activate the ER-binding mediated pathway. The fact that they are not agonists is broadly supported by the FI and AM pesticides data presented by EPA and by differences with published data for other compounds that are considered ER agonists through interaction with the ER receptor binding site, *e.g.,* Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM, 2006).

The Panel concluded that the findings for acyclic compounds in the FI and AM inventories could be extended to most acyclic compounds in other inventories, but not to all. Panel members pointed out the training set used for the FI and AM pesticides inventories may not be adequate for all acyclic compounds. Acyclic compounds may bind to the ER by different rules. The Panel thought that there may be an automated way to test for fragments that might be able to bind to the ER receptor, *e.g.,* Analog Identification Method (AIM). Some panel members suggested that test compounds could be synthesized by combinatorial methods to aid study of structural binding relationships. The Panel cautioned that special attention should be given to halogenated acyclic compounds.

They pointed out that acyclic compounds may bind to receptors at binding sites other than A and B. Based on the crystal structure of the estrogen receptor ligand binding domain crystallized in the presence of ligands and inhibitors, there are three important binding sites, A, B, C and there is significant plasticity in the binding domains (Katzenellenbogen *et al.,* 2003). They stated that binding can also occur outside the ligand binding domain or other subdomains. This is particularly important for antagonists. Diethylstilbestrol should also be considered.

They indicated that there are several databases which might prove to be useful to inform the expert system. For example, ICCVAM has collated a literature database of both

positive and negative estrogenic and androgenic compounds tested *in vitro* for estrogen (ER) and androgen (AR) receptor binding that could be evaluated using the rules of the expert system (ICCVAM, 2006). That is, one could run the chemicals in the ICCVAM database through the expert system and then compare the predicted results with the experimental results. One panel member pointed out that the European Commission is developing an "Endocrine-Active Compounds Database and Web Portal" which is intended to become a living publicly accessible literature based database covering a wide range of endocrine pathways.

Another example is the Tox21 program. The Tox21 effort is a federal government partnership composed of the National Toxicology Program (NTP), part of the NIH National Institute of Environmental Health Sciences (NIEHS); the NIH Chemical Genomics Center (NCGC), managed by the National Human Genome Research Institute (NHGRI); and the EPA's National Center for Computational Toxicology (NCCT) (see http://www.epa.gov/comptox/articles/comptox_mou.html). Tox21 studies will ultimately test approximately 10,000 chemicals for transcriptional activation of ERα and androgen receptors, and a dozen other nuclear receptors using high throughput assays (Collins *et al.,* 2008). These assays have 15-point dose response curves from 5 nM to 100 μM. So far, approximately 2,500 chemicals have been tested. The data that are being collected implicate other downstream events in the ER-mediated pathway that are also important, *e.g.,* interactions with coactivators. The Panel noted that as the data from the Tox21 screening assays become available, there will be more additional data with which comparisons can be made. This applies not only to the ER but also to the AR, as well as to the dozen or so other nuclear receptors against which the chemical libraries are currently being screened.

The Panel discussed ongoing studies whose results may influence the knowledge basis of the expert system. One panelist reported that the nuclear receptors mentioned in the Tox21 program are being assayed using cell-based commercially available beta-lactamase transcription reporter assays. Preliminary findings indicate some acyclic compounds, in particular, halogenated compounds, are showing up as positive in initial evaluations of the output of the high throughput testing work with the ERα assay. Other studies have shown that compounds such as PFOS and PFOA are estrogenic in other assay systems. The Panel noted that there are many reports of certain metals which can be grouped with acyclic compounds, which show up as estrogenic in certain assay systems (Byme *et al.*, 2009; Denier *et al.*, 2009). Based on the results of current studies, the Panel concluded that acyclic compounds could conceivably have inherent estrogenic activity or perhaps affect the binding of natural ligands for any of the estrogen receptors.

Aside from the cell-based transcription activation assay results, there are other potential effects of acyclic compounds on ER-binding, for example, altering the structure of the ERs and thereby limiting transactivation by ER ligands. In a broad sense, the Panel explained, these acyclic compounds may bind directly to the ER, affect binding of E2 at the ligand binding site, affect structural changes related to receptor activation or coactivator protein interactions, and/or affect other downstream actions related to transcriptional activation.

The Panel commented that if the charge question is strictly restricted to ER-binding and activation in a manner similar to the natural ligand, then EPA's proposal that acyclic compounds be excluded seems reasonable. However, binding to the A and B sites in the ER is only one of many possible ways of chemical interactions that can result in increased or decreased hormonal action. Because of this, the Panel indicated that the expert system should be viewed in the context of the entire EDSP in setting priorities for Tier 1 screening.

> *Question b. Please comment on the extent to which the finding with acyclic compounds in the FI and AM inventories can be applied to other nuclear steroid receptors in general. Suggestions on an approach to empirically and efficiently assess a hypothesis that acyclic compounds will not bind to the androgen receptor would be welcomed.*

The Panel concluded that it would be difficult (if not impossible) to extrapolate finding for acyclic compounds in the FI and AM inventories to other nuclear steroid receptors because of different chemical structural-binding interactions expected for other receptors, *e.g.*, membrane binding receptors, thyroid hormone receptors, and androgen receptors.

The Panel indicated that there are membrane receptors for androgens and progestins (Thomas *et al.*, 2006) that may have possible binding sites for acyclic compounds belonging to the regulatory inventories of interest. Other nuclear receptors, such as androgen receptors and thyroid hormone receptors, while less studied than estrogen receptors, have specific structure limitations in their ligand binding pockets. Crystal structures for androgen receptor ligand binding domains have been studied (Bohl *et al.*, 2007; Pereira de Jésus-Tran *et al.*, 2006) and (Q)SAR models have been built (Lill *et al.*, 2005). Thyroid hormone receptors have been crystallized as well (Nunez *et al.*, 2004), and utilized for (Q)SAR modeling (Vedani *et al.*, 2007). The binding of ligands to these receptors rely on different binding interaction principles, *i.e.*, H-bonding - donation and acceptance capacities. Other receptor binding interactions may rely on Pi-Pi stacking interactions, ion bonding, energetics, shape, globularity, and other determinants for binding (Jacobs, 2005). For example, understanding the required interactions for binding to the AR is harder than it has been for the ER, because there is much less information available. For other nuclear receptors, including AR, there are different conformational changes and different modes of binding and downstream DNA activation or inactiviation that are receptor specific. Even for ERα, some compounds bind to sites outside of the core ligand binding domain such as the antagonist, ICI 182,780. Phenobarbital-like compounds also bind on the outside of the Constitutive Androstanen Receptor (CAR), not in the ligand binding pocket.

The Panel commented that thinking about other nuclear receptors is a valid extension of the work completed by EPA on the ER. They encouraged EPA to study the structural activity relationships for these other receptors with the same logic and thoroughness that was used for soluble ERs, *i.e.*, defining the optimum structure for receptor binding. The

Panel agreed that it would be reasonable to construct an appropriate expert system for other hormone receptors given enough information about structure binding relationships. The Panel thought that it might be interesting to use a wide variety of *in vitro* assays including for example, recombinant yeast assays and transactivation assays transfected with other relevant receptors, to develop the set of rules for other receptors. These assays are available for a variety of receptors including ER agonist/antagonist and AR agonist/antagonist activities. One possible structural motif of interest would be halogenated long chain molecules. Some panel members suggested that test compounds could be synthesized by combinatorial methods to aid the study of structural binding relationships with other steroid receptors.

3. *Prioritization for EDSP Tier 1 Screening*

OECD member countries have long recognized the potential of (Q)SAR for initial assessments for thousands of untested chemicals and to establish priorities for follow up actions. The OECD "Integrated Approaches to Testing and Assessment" framework has encouraged the use of existing knowledge including (Q)SAR to effectively assess and manage large chemical inventories (http://www.oecd.org/dataoecd/45/52/40705314.pdf). In its final report, the EDSTAC (USEPA, 1998) recommended a tiered approach for detecting chemicals with endocrine disrupting potential using a resource-efficient manner that is similar to OECD's Endocrine Disruptor Testing and Assessment Framework (http://www.oecd.org/document/58/0,3343,en_2649_34377_2348794_1_1_1_1,00.html). Like the OECD approach, the framework proposed by the EDSTAC includes use of (Q)SARs and high through put screening assays.

Based on the characteristics of the (Q)SAR-based expert system presented in the White Paper, please comment on the Agency's view that the expert system could be employed to support "sorting and prioritizing" food use inert ingredients and antimicrobial pesticides for EDSP Tier 1 screening.

Panel Response

The Panel was in strong agreement that the expert system is likely to be an important tool to support "sorting and prioritizing" FI and AM pesticides for EDSP Tier 1 screening by the ERα binding mechanism. They supported the intended use of the expert system as a prioritization tool given the likely paucity of data available for food use inerts and antimicrobial pesticides. Considerable and careful effort to cover the structural domains of these two regulatory inventories was invested in development of the expert system. The expert system would therefore constitute a critically important source of relevant information for prioritization of the two regulatory inventories of interest.

The Panel recommended that the Agency continue with its efforts to collect additional data to better understand the relationship of the predicted ERα binding at the cellular and tissue level, e.g., medium and high throughput assays and Tier 1 EDSP testing among species. Additionally, these data would contribute to defensibility of their use in the

expert system. They also recommended that ER-binding of mixed phenols be further studied.

They stressed that the ER binding mechanism is only part of the estrogenic mode of action, and the estrogenic mode of action is one component of the endocrine mode of action. Other aspects of steroid metabolism and action pathways could also be important for prioritization. As an illustration, **Figure 1** highlights several key upstream receptors and P450 enzymes such as aromatase in the steroidogenic pathway. Furthermore, steroid hormone nuclear receptors are only part of the potential endocrine disruptor xenobiotic mechanistic pathway. In addition, there are non-steroidogenic receptors and non-receptor mediated pathways to consider as well (**Figure 2**). For now, the Panel agreed that prioritization is only within the context of relative ERα binding. The "positives" will help decide which chemicals are high priorities for EDSP Tier 1 screening in the next few years.

As one panel member explained, the current expert model orders chemicals into two bins defined by the limit of detection of the associated outcome of interest, ERα binding. Sorting to a certain extent implies ranking highest to lowest based on a scoring method. The "cut point" for declaring priorities versus non-priorities in such a system is arbitrary, in this case, the limit of detection of the specific binding assay. Hence, the Panel noted that there is no need to order priority of the chemicals that test positive, as these will all be a priority for testing, i.e., FI and AM pesticides that will be screened first under the EDSP and those that are below the limit of detection will all be considered a lower priority for testing ERα.

The Panel suggested that sorting and prioritization aspects of the expert system might be enhanced if each of the seven model rules were weighted separately. "Weighting," in this context, means comparing the output of each of the seven model rules to other relevant information on hazard as to whether or not the compound in question may be an endocrine disruptor. This analysis would include all relevant information not captured by the proposed expert system, for example, there is often *in vitro* data from other assays, physical/chemical data, etc., that would inform in this context.

The Panel commented that substructures might also be further classified into ICI 182780-like, PFOS-like, fatty acid-like, etc., to be screened for antagonist properties to key binding sites not thus far included in the ERα binding sites A and B as well as in other related receptor (Q)SAR expert systems. Rules could be developed based on the space of these other receptor binding pockets. This would allow binning into more groups and get closer to a "highest to lowest" priority sorting system.

One panelist discussed whether the ERα binding assay might serve as a replacement or surrogate for the fish short-term reproductive assay and 21-day fish screening assay in the hypothalamic/pituitary/gonadal (HPG) activity Tier I screening assessment. The Panel referred to the conclusions made by the March 2008 SAP on this question (see meeting report; FIFRA SAP, 2008). The March 1998 SAP was very clear that the fish short-term reproductive assay was an "essential part of apical analysis" of HPG. Essentially this

assay integrates across the multiple potential pathways that could produce individual and population effects. The August 2009 SAP affirmed this conclusion.

The Panel highlighted a couple of specific issues for "priority setting."

1) What happens to the chemicals for which ERα binding is not possible?

In a priority setting context, the Panel indicated that there should be a clear understanding of whether compounds for which binding to the ER is not detected and are not considered for further testing or as a low priority for further testing. In the latter case, additional information might prompt reconsideration. As designed, the expert system separates chemicals into two bins defined by the limit of detection of the ER binding assay, *i.e.,* 0.0001%. Sorting is not an issue; therefore, all positives are a priority and all negatives are less of a priority. Chemicals with RBA>0.00001%, *i.e.,* "the positives", would be assigned a higher priority for EDSP Tier 1 testing and chemicals whose RBA< 0.00001%, *i.e.,* "the negatives", would be assigned a lower priority for testing. The Panel indicated that the "negatives" should not be dismissed as those that do not bind to the ER and do not need further testing; rather, they should be characterized as chemicals in which prediction of ERα binding is not possible. The Panel recommended that the term "prediction of ER binding is not possible" be used in the expert system.

2) How will the information from the expert system be integrated with existing information, e.g., likely limited available data from toxicity testing, in a weight of evidence approach, to address traditional criteria such as consistency?

The Panel indicated that the expert system used to predict estrogen receptor binding affinity for FI and AM pesticides is useful for application in a prioritization scheme for endocrine disruptor screening. However, they indicated that the expert system should be viewed in the context of the overall EDSP. That is, use a weight of evidence approach for integrating information coming from many possible sources, including the results of endocrine testing, modeling such as the rule-based expert system for ER-binding, as well as other information relative to the reactivity of the chemical. This would then represent a true screening system, one that is designed to resemble a funnel, with a very wide mouth, rather than a narrow funnel screening, *e.g.*, ER-binding, which is only part of the screening program. Only if substances are negative in this comprehensive, integrated screen would they be declared negative for endocrine disruption and not considered further. Model components of such a meta-screen approach need not be integrated into the model at the outset. The first areas screened might be ERα binding, androgen receptor binding, aromatase activation and thyroid receptor mediation. Other components could be added as they are built from high throughput screening programs, *e.g.*, Tox21. Other information would come from data produced by EPA, in the literature, or from OECD Test Guideline studies.

Panel members commented that the criteria by which information from various sources will be weighted constitutes a critical element of transparency in defensibility

of use of the evolving expert system in a priority setting system. In programs mandated to address priority setting for large inventories, this has been a relatively critical component and requires consideration in the context of criteria such as consistency (see, for example, the principles for combining information and data for cancer/genotoxicity in documentation for health components of the Canadian Domestic Substances List (DSL) program): [http://www.hc-sc.gc.ca/ewh-semt/contaminants/existsub/categor/approach-approche-eng.php and http://www.hc-sc.gc.ca/ewh-semt/contaminants/existsub/categor/publi-comment/index-eng.php].

4. Cross Species Applicability

As discussed in the White Paper, when comparable assay systems are used, e.g., comparing recombinant receptors to recombinant receptors; comparing cytosolic receptors to cytosolic receptors; comparing assays with similar total protein concentrations and thus chemical availability, there is general agreement in measurements of binding affinity across species (fish, rat, human). Thus, the type of assay used appears to explain differences more than species origin of the receptors. To evaluate the applicability of the current expert system based on a trout ER training set for predicting relative binding affinity to human ERα, binding assays using full-length recombinant human ERα and transactivation assays in human T47D cells are in progress with food use inert ingredients and antimicrobial active ingredients. Chemicals selected for human ER testing are based on predictions from the current expert system and designed to cover each chemical group and bracket the log Kow ranges within the group. To date, results show good species concordance with chemical groups that have members that bind to the trout ER also having members that bind to the human ER, although the trend is toward fewer members of a chemical group binding to the human ER than to the rainbow trout ER, *e.g.*, a more restrictive log Kow range for binding within a chemical group for human ER. Therefore rainbow trout ER appears to bind more low affinity chemicals within a group but binds the same type of chemicals as does human ER.

Given what is reported in the literature and similarities between human and rainbow trout ER binding affinity observed thus far in the research described in the White Paper, please comment on the extent to which use of an expert system based primarily on trout ER binding affinity data is a reasonable effects component for prioritizing food use inert ingredients and antimicrobial pesticides for EDSP Tier I screening.

Panel Response

The Panel responded to this question in two different ways.

1) Some commented in a broad sense about the extrapolation of trout cytosolic ER binding data "across species" and were not confined to comparisons between the rainbow trout and mammalian ERs. These comments were meant to consider the ecological implications of screening tests in addition to the human health implications.

2) Others discussed how the trout cytosolic ER assay may fit into prioritization schemes for endocrine disruption screening tests primarily from the viewpoint of human health. In this latter context, the Panel thought that trout cytosolic ER and human ER assays would likely yield comparable results.

The following paragraphs provide further details of the Panel's discussion.

Receptor-mediated interference with the estrogen hormone system by xenobiotics can occur via several types or subtypes of ER in target tissues, including not only ERα, but also ERβ and unrelated estrogen receptors such as the membrane ER and G protein-coupled receptor 30 [GPR30] (Thomas *et al.,* 2005; Revankar *et al.,* 2005). The rainbow trout presents a particularly complex situation, in which at least two isoforms of ERα and two of ERβ are known to be expressed in the liver (Boyce-Derricott *et al.,* 2009). EPA's White Paper and associated presentations provided evidence suggesting that the trout cytosolic ER assay behaves similarly to the recombinant trout ERα assay, and that the latter behaves similarly to recombinant ERα assays across species. The Panel discussed results of recent unpublished studies by European investigators that also provide support to the notion that ERα subtypes have similar patterns of ligand affinities across species. Overall, the Panel concluded that currently available information suggests that the trout cytosolic ER assay can be used to develop reasonable predictions of the RBA of xenobiotics for ERα across species. However, what is unclear at the present time is whether the trout cytosolic ER assay can be used for estimating RBAs of xenobiotics across all ER types and species.

The results of a few available studies suggest that the RBA of ligands, including natural ligands, can vary between the different subtypes of nuclear ERs across species. Whereas ERα and ERβ both appear to bind E2 with similar affinity in mammals (e.g., Kuiper *et al.*, 1997), this is not the case for fish. For example, binding affinity of E2 for ERβ relative to ERα among teleost fishes ranges from similar-or-slightly higher in Atlantic croaker (Hawkins and Thomas, 2004), to greater-than-6-fold higher in gilthead sea bream (Passos *et al.* 2009), and to almost 20-fold higher in channel catfish (Xia *et al.*, 2000; Gale *et al.*, 2004). Conversely, the RBAs of a handful of high and low affinity xenoestrogens tended to be higher for ERα than for ERβ, in channel catfish; and in the case of 4-nonylphenol, its RBA for ERα was almost 100-fold higher than for ERβ (Gale *et al.*, 2004). These observations suggest that the RBAs of natural estrogens and xenobiotics for the different subtypes of nuclear ERs (ERα and ERβ) vary widely across species. Most of the RBAs are defined for soluble estrogen receptors which are found in the soluble cellular extracts. Binding to membrane receptors may be missed by these assays, but they function *in vivo* and would also be captured in the transactivation assay using liver slices.

Therefore, models used to predict the RBA of xenobiotics for ERβ based on data obtained with trout cytosolic ER or ERα assays may yield unreliable results for the purpose of cross-species comparisons. They also noted that membrane and nuclear ERs may have different binding affinities. The Panel suggested that additional research would be needed to understand the extent of taxonomic similarities or differences in the RBAs of xenobiotics for nuclear ERs, particularly ERβ, and for membrane ER.

They specifically discussed whether results from the trout cytosolic ER binding assay should or could be used to prioritize food use inerts or antimicrobial pesticides for EDSP Tier 1 screening in the context of mammalian ER. The Panel thought that if a given chemical yielded a positive result in the trout cytosolic ER binding assay then this chemical would also be likely to yield a positive result in mammalian systems because the trout cytosolic ER assay seems to be less specific than the human ERα assay. However, in extrapolating a negative result from the trout cytosolic ER assay to human systems, there would be some level of uncertainty because some differences in performance have been reported between cytosolic and recombinant ER assays (EPA's White Paper, page 38; and associated presentations). Differences include the fact that the rainbow trout cytosolic ER is less sensitive than the rainbow trout recombinant ERα-ligand binding domain (LBD) assay (Matthews *et al.*, 2000; 2002). The Panel recommended that the Agency pursue studies to calibrate the cytosolic and recombinant ER assays, including β isoforms, to better define the degree of uncertainty.

Panel members were generally in agreement that cross-species extrapolation, *e.g.*, from fish to mammals, can be done as there is good concordance in RBA of cytosolic ERα among species. The Agency is examining the endocrine disrupting potential using a resource-efficient manner that is similar to OECD's Endocrine Disruptor Testing and Assessment Framework. The OECD level 1 testing uses existing data or (Q)SAR systems. Similarly, EPA is considering use of a (Q)SAR expert system as a mechanism to assist in the prioritization of FI and AM pesticides for EDSP Tier 1 screening. They also considered the fact that other endocrine-disruptor screening tests being pursued by the Agency and the OECD include both mammalian and fish systems. They concluded that the trout cytosolic ER assay was adequate for the purposes of using the expert system in prioritization of FI and AM pesticides for EDSP Tier 1 testing.

The Panel also discussed issues concerned with what subtypes of ERs were present in the trout livers and seasonal influences on their tissue expression. They suggested that the Agency pursue some simple assays using trout liver cytosol to better define the ER composition of the preparations and assist in the interpretation of the binding data. For example, one assay would be conducted to determine the relative amounts of ERα and ERβ mRNAs as a measure of their relative expression at the mRNA level. Another assay would be conducted to measure the RBAs of chemicals compared to each other with trout recombinant ERα. However, despite gaps in knowledge and the caveats discussed during the meeting, the Panel concluded that a positive result from the trout cytosolic ER assay would be strong reason to prioritize chemicals (*i.e.*, FI and AM pesticides) for further EDSP Tier 1 testing.

REFERENCES

Blair RM, Fang H, Branham W, Hass B, Dial S, Moland C, Tong W, Shi L, Perkins R, and Sheehan D. (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. Tox. Sci. 54: 138-153.

Breiman L, Friedman J, Stone CJ and Olshen RA. (1984) Classification and Regression Trees. Wadsworth International, Belmont, CA.

Bohl CE, Wu Z, Miller DD, Bell CE, Dalton JT. (2007) Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. J. Biol. Chem. 282:13648-13655.

Boyce-Derricott J, Nagler JJ, Cloud JG. (2009) Regulation of hepatic estrogen receptor isoform mRNA expression in rainbow trout (*Oncorhynchus mykiss*). General and Comparative Endocrinology 161:73-78.

Bian Z, Nilsson S, Gustafsson JA. (2001) Selective estrogen receptor modulators and coronary heart disease. Trends Cardiovasc. Med. 11:196-202.

Brian JV, Harris CA, Scholze M, Backhaus T, Booy P, Lamoree M, Pojana G, Jonkers N, Runnalls T, Bonfà A, Marcomini A, Sumpter JP. (2005) Accurate prediction of the response of freshwater fish to a mixture of estrogenic chemicals. Environ Health Perspect. 113:721-8.

Brian JV, Harris CA, Scholze M, Kortenkamp A, Body P, Lamorce M, Pojana G, Jonkers N, Marcomni A, and Sumpter JP. (2007) Evidence of estrogenic mixtures effects on the reproductive performance of fish. Environ. Sci. Technol. 41: 337-344.

Carruba G. (2007) Estrogen and prostate cancer: an eclipsed truth in an androgen-dominated scenario. J. Cell Biochem. 102:899-911.

Collins, FS, Gray, GM, Bucher, JR. (2008) Transforming Environmental Health Protection. Science 319:906-907.

Crane M, Gross MY, Daginnus K, Deviller G, de Wolf W, Dungey S, Galli C, Gourmelon A, Jacobs M, Matthiessen P, Micheletti C, Nestmann E, Pavan M, Paya-Perez A, Ratte T, Safford B, Sokull-Kluttgen B, Stock F, Stolzenberg HC, Wheeler JR, Willuhn M, Worth AP, Zaldivar Comenges JM. (2009) Thresholds of Toxicological Concern for Endocrine Active Substances in the Aquatic Environment. Integr Environ Assess Manag. Papers In Press, accepted June 26, 2009. DOI: 10.1897/IEAM_2008-092.1.

Denier X, Hill EM, Rotchell J, Minier C. (2009) Estrogenic activity of cadmium, copper and zinc in the yeast estrogen screen. Toxicol In Vitro. 23: 569-573.

Denny JS, Tapper MA, Schmieder PK, Hornung MW, Jensen KM, Ankley GT, Henry TR.

(2005) Comparison of relative binding affinities of endocrine active compounds to fathead minnow and rainbow trout estrogen receptors. Environ Toxicol Chem. 24:2948-2953.

Ellem SJ, Risbridger GP. (2009) The dual, opposing roles of estrogen in the prostate. Ann N Y Acad Sci. 1155:174-186.

Gale WL., Patiño R., and Maule AG. (2004) Interaction of xenobiotics with estrogen receptors α and β and a putative plasma sex hormone binding globulin from channel catfish (*Ictalurus punctatus*). General and Comparative Endocrinology 136:338-345.

Hamblen EL, Cronin MTD and Schultz TW. (2003) Estrogenicity and acute toxicity of selected anilines using a recombinant yeast assay. Chemosphere 52:1173-1181.

Hartman I, Gillies AR, Arora S, Andaya C, Royapet N, Welsh WJ, Wood DW, Zauhar RJ. (2009) Application of screening methods, shape signatures and engineered biosensors in early drug discovery process. Pharm Res. 26:2247-2258. Epub 2009 Jul 22.

Hawkins MB, Thomas P. (2004) The unusual binding properties of the third distinct teleost estrogen receptor subtype ERβa are accompanied by highly conserved amino acid changes in the ligand binding domain. Endocrinology 145:2968-2977.

Interagency Coordination Committee on the Validation of Alternative Methods (ICCVAM). (2006) ICCVAM Evaluation of In Vitro Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays (NIH Publication No: 03-4503), 28 Sept. 2006.

Ishibashi H., Yamauchi R., Matsuoka M., Kim J.-W., Hirano M., Yamaguchi A., Tominaga N., Arizono K. (2008) Fluorotelomer alcohols induce hepatic vitellogenin through activation of the estrogen receptor in male medaka (*Oryzias latipes*). Chemosphere 71: 1853-1859.

Jacobs, MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M, and Burg Bvd. (2008) The use of metabolising systems for in vitro testing of endocrine disruptors. Current Drug Metabolism 9: 796-826.

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

Jacobs MN. (2004) *In silico* tools to aid risk assessment of endocrine disrupting chemicals. Toxicol. 205:43-53.

Jacobs MN, Dickins M, and Lewis DFV. (2003) Homology modelling of the nuclear receptors: human oestrogen receptorβ (hERβ), the human pregnane-X-receptor (PXR), the Ah receptor (AhR) and the constitutive androstane receptor (CAR) ligand binding domains from the human

oestrogen receptor α (hERα) crystal structure, and the human peroxisome proliferator activated receptor α (PPARα) ligand binding domain from the human PPARγ crystal structure. J. Steroid Biochem. Mol. Biol. 84:117-132.

Jaworska JS, Comber M., Auer C. and van Leeuwen CJ. (2003). Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. Environmental Health Perspectives 111:1358-1360.

Katzenellenbogen J., Muthyala R, and 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.

Kidd KA, Blanchfield,PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, and Flick RW. (2007) Collapse of a fish population after exposure to a synthetic estrogen. PNAS 104:8897–8901.

Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA. (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β. Endocrinology 138:863-870.

Leaños-Castañeda O, Van Der Kraak G. (2007) Functional characterization of estrogen receptor subtypes, ERα and ERβ, mediating vitellogenin production in the liver of rainbow trout. Toxicology and Applied Pharmacology 224:116-125.

Lill MA, Winiger F, Vedani A, Ernst B. (2005) Impact of induced fit on ligand binding to the androgen receptor: a multidimensional QSAR study to predict endocrine-disrupting effects of environmental chemicals. J Med Chem. 48:5666-5674.

Liu C, Du Y, Zhou B. (2007) Evaluation of estrogenic activities and mechanism of action of perfluorinated chemicals determined by vitellogenin induction in primary cultured tilapia hepatocytes. Aquatic Toxicology 85:267-277.

Matthews J, Celius T, Halgren R, Zacharewski T. (2000) Differential estrogen receptor binding of estrogenic substances: a species comparison. J. Steroid Biochem. Mol. Biol. 74:223-234.

Matthews JB, Fertuck KC, Celius T, Huang YW, Fong CJ, Zacharewski TR. (2002) Ability of structurally diverse natural products and synthetic chemicals to induce gene expression mediated by estrogen receptors from various species. J Steroid Biochem. Mol. Biol. 82:181-194.

McPherson SJ, Ellem SJ, Patchev V, Fritzemeier KH, Risbridger GP. (2006) The role of Eralpha and ERbeta in the prostate: insights from genetic models and isoform-selective ligands. Ernst Schering Found. Symp Proc.1:131-147.

Nagler JJ, Cavileer T, Sullivan J, Cyr DG, Rexroad C 3rd. (2007) The complete nuclear estrogen receptor family in the rainbow trout: discovery of the novel ERalpha2 and both ERbeta isoforms. Gene. 392:164-173.

Nunes FM, Aparicio R, Santos MA, Portugal RV, Dias SM, Neves FA, Simeoni LA, Baxter JD, Webb P, Polikarpov I. (2004) Crystallization and preliminary X-ray diffraction studies of isoform alpha1 of the human thyroid hormone receptor ligand-binding domain. Acta Crystallogr. D. Biol. Crystallogr. 60:1867-1870.

Organization for Economic Cooperation and Development (OECD 2004) The Report From The Expert Group On (Quantitative) Structure-Activity Relationships [(Q)SARs] On The Principles For The Validation Of (Q)SARs. Series on Testing and Assessment. No. 49. ENV/JM/MONO(2004)24. 17-Dec-2004. http://www.oecd.org/dataoecd/33/37/37849783.pdf.

OECD. (2009) Report of the Expert Consultation to Evaluate an Estrogen Receptor Binding Affinity Model for Hazard Identification. Task Force on Hazard Assessment. ENV/JM/HA/RD(2009)1. 2-March-2009 107pp. (declassified 31-Jul-2009).

Passos AL, Pinto PI, Power DM, Canario AV. (2009) A yeast assay based on the gilthead sea bream (teleost fish) estrogen receptor beta for monitoring estrogen mimics. Ecotoxicology and Environmental Safety 72:1529-1537.

Pereira de Jésus-Tran K, Côté PL, Cantin L, Blanchet J, Labrie F, Breton R. (2006) Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. Protein Sci. 15:987-999.

Revankar CM, Cimino DF, Sklar LA, Arterburn JF, Prossnitz ER. (2005) A Transmembrane intracellular estrogen receptor mediates rapid cell signaling. Science 307:1625-1630.

Sanseverino J, Eldridge ME, Layton AC, Easter JP, Yarbrough JW, Schultz TW and Sayler GS. (2009) Screening of potentially hormonally active chemicals using bioluminescent yeast bioreporters. Toxicological Sciences 107: 122-134.

Schultz TW, Sinks GD and Cronin MTD. (2002) Structure-activity relationships for gene activation oestrogenicity: Evaluation of a diverse set of aromatic compounds. Environmental Toxicology 17:14-23.

Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP). (2008) Report of the March 25-28, 2008 SAP Meeting: Scientific Issues Associated with the Review of the Endocrine Disruptor Screening Program (EDSP) Proposed Tier-1 Screening Battery. http://www.epa.gov/scipoly/sap/meetings/2008/032508_mtg.htm.

Seki, M., Yokota, H., Maeda, M., Tadokoro, H., Kobayashi, K. 2003 Effect of 4-Nonylphenol and 4-*tert*-Octylphenol on the Sex Differentiation in Medaka (*Oryzias latipes*). Environmental Toxicology and Chemistry. 22:1507–1516.

Thomas P, Dressing G, Pang Y, Berg H, Tubbs C, Benninghoff A, Doughty K. (2006) Progestin, estrogen and androgen G-protein coupled receptors in fish gonads. Steroids. 71:310-316.

Thomas P, Pang Y, Filardo EJ, Dong J. (2005) Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. Endocrinology 146:624-632.

United States Environmental Protection Agency (USEPA). (1998) Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report: Chapter 4: Priority Setting. Aug. 1998. http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/finalrpt.htm.

Van Aerle R, Ponds N, Hutchinson TH, Maddix S and Tyler CR. (2002) Window of sensitivity for estrogenic effects of ethinylestradiol in early life-stages of fathead minnow, *Pimephales promelas*. Ecotoxicology 11:423-434.

Vedani A, Zumstein M, Lill MA, Ernst B. (2007) Simulating alpha/beta selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. Chem. Med Chem. 2:78-87.

Vedani A, Dobler M, Lill MA. (2005) Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. J Med Chem. 48:3700-3703.

Wilson ME and Westberry JM. (2009) Regulation of oestrogen receptor gene expression: new insights and novel mechanisms. J Neuroendocrinol. 21:238-242.

Xia Z, Gale WL, Chang X, Langenau D, Patiño R, Maule AG, Densmore LD. (2000) Phylogenetic sequence analysis, recombinant expression, and tissue distribution of a channel catfish estrogen receptor β. General and Comparative Endocrinology 118:139-149.

Zha JM, Wang ZJ, Wang N, and Ingersoll C. (2007) Histological alteration and vitellogenin induction in adult rare minnow (*Gobiocypris rarus*) after exposure to ethinylestradiol and nonylphenol. Chemosphere 66: 488-495.

SELECTED ABBREVIATIONS

AhR: Aryl hydrocarbon Receptor
AIM: Analog Identification Method
AR: Androgen Receptor
ARNT: AhR Nuclear Translocator
CAR: Constitutive Androstane Receptor
CYP: Cytochrome P450
DDT: Dichlorodiphenyltrichloroethane
DNA: Deoxyribonucleic acid
E2: 17β-estradiol
EA: Endocrine Activity
ED: Endocrine Disruptor
ER: Estrogen Receptor
FXR: Farnesoid X Receptor
GPR30: G Protein-coupled Receptor 30
GR: Glucocorticoid Receptor
HPG: Hypothalamic/pituitary/gonadal
HTP: High Throughput
HVC: High Volume Chemical
H: Hydrogen
Kow: Octanol-water partition coefficient
Log P: Log Partition Coefficient, same as Kow when the two phases are octanol and water
LXR: Liver X Receptor
OECD: Organisation for Economic Cooperation and Development
P450: Cytochrome P450
PBPK: Physiologically-based Pharmacokinetic
PFOA: Perfluorooctanoic acid
PFOS: Perfluorooctane sulfonate
PGP: P-glycoprotein
PPAR: Peroxisome Proliferator Activated Receptors
PR: Progesterone Receptor
PXR: Pregnane X Receptor (also termed SXR: Steroid and Xenobiotic Receptor)
QMRF: (Q)SAR Model Report Format
(Q)SAR: Quantitative Structure-Activity Relationship
ROS: Reactive Oxygen Species
RXR: Retinoid X Receptor
SMARTS: Smiles Aritary Target Specification
SMILES: Simplified Molecular Input Line Entry Specification
SULT: Sulphotransferase
TR: Thyroid Receptor
UGT: UDP-dependent glucuronosyl transferase
VDR: Vitamin D Receptor
Vtg: Vitellogenin

Figure 1. The key known receptors and P450's involved in the steroidogenic pathway (Jacobs, 2004) [see p. 68 for a list of Abbreviations]

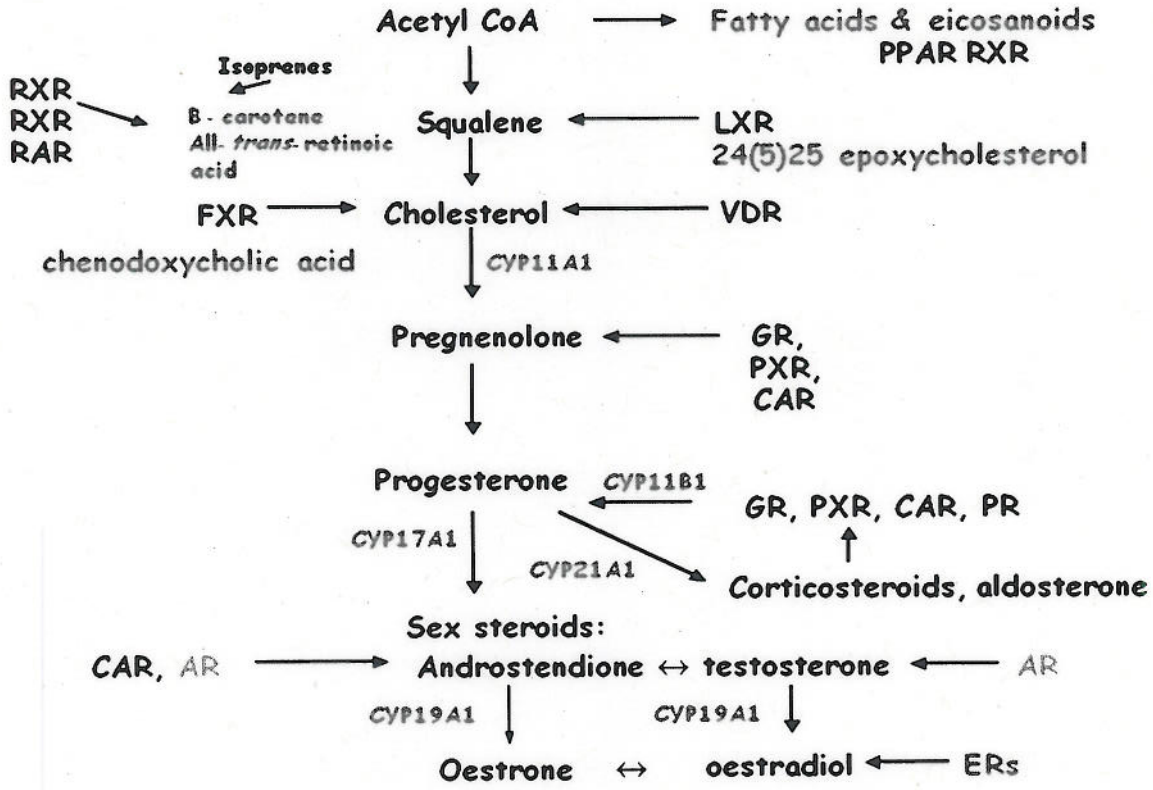

Figure 2. An overview of xenobiotics as modulators of nuclear receptor function (Jacobs, 2004) [see p. 68 for a list of Abbreviations]