



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, DC 20460

OFFICE OF CHEMICAL SAFETY AND POLLUTION PREVENTION

August 21, 2013

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held May 21-23, 2013 on "Endocrine Disruptor Screening Program (EDSP) Tier 1 Screening Assays and Battery Performance"

TO: David Dix, Ph.D.
Acting Director
Office of Science Coordination and Policy

And

Steven Bradbury, Ph.D.
Director
Office of Pesticide Programs

FROM: Fred Jenkins, Jr., Ph.D. *Fred*
Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Laura Bailey, MS *Laura Bailey for*
Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, VA on May 21-23, 2013. This report addresses a set of scientific issues associated with "Endocrine Disruptor Screening Program (EDSP) Tier 1 Screening Assays and Battery Performance."

Enclosure

cc:

Jim Jones
Louise Wise
William Jordan
Margie Fehrenbach
Yu-Ting Guilaran
Robert McNally
Donald Brady
Jack Housenger
Susan Lewis
Lois Rossi
Jay Ellenberger
Richard Keigwin
Mary Manibusan
Jess Rowland
Don Bergfelt
Gregory Akerman
Catherine Aubee
Amy Blankinship
John Liccione
Tom Steeger,
Leslie Touart
Molly Hooven
Dale Kemery
Linda Strauss
Vanessa Vu
OPP Docket

FIFRA Scientific Advisory Panel Members

Daniel Schlenk, Ph.D.
Kenneth Delclos, Ph.D.
Marion F. Ehrich, Ph.D., D.A.B.T, A.T.S
Stephen Klaine, Ph.D.
James McManaman, Ph.D.

FQPA Science Review Board Members

Sherril Green, D.V.M, Ph.D.

Nelson D. Horseman, Ph.D.

Jack Leonard, Ph.D.

Deborah MacLatchy, Ph.D.

Reynaldo Patino, Ph.D.

Aldert H. Piersma, Ph.D.

Katherine F. Roby, Ph.D.

Geoffrey Scott, Ph.D.

Daniel Selvage, Ph.D.

Terry Wayne Schultz, Ph.D.

Donald Tillitt, Ph.D.

SAP Minutes No. 2013-03

**A Set of Scientific Issues Being Considered by the
Environmental Protection Agency Regarding Endocrine
Disruptor Screening Program (EDSP) Tier 1 Screening Assays
and Battery Performance**

**May 21-23, 2013
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA**

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, 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 Fred Jenkins, Jr., Ph.D., SAP Designated Federal Official, via e-mail at jenkins.fred@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented by public commenters.

TABLE OF CONTENTS

PARTICIPANTS 5

INTRODUCTION 8

PUBLIC COMMENTS..... 9

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS 10

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE..... 19

REFERENCES 56

SAP Minutes No. 2013-03

**A Set of Scientific Issues Being Considered by the
Environmental Protection Agency Regarding:
Endocrine Disruptor Screening Program (EDSP) Tier 1
Screening Assays and Battery Performance**

**May 21-23, 2013
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA**



**Daniel Schlenk, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel
Date: August 21, 2013**



**Fred Jenkins, Jr., Ph.D.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: August 21, 2013**

**Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)
Scientific Advisory Panel (SAP)
Endocrine Disruptor Screening Program (EDSP) Tier 1 Screening Assays and
Battery Performance May 21-23, 2013**

PARTICIPANTS

FIFRA SAP Chair

Daniel Schlenk, Ph.D.

Professor of Aquatic Ecotoxicology & Environmental Toxicology
Department of Environmental Sciences
University of California, Riverside
Riverside, CA

Designated Federal Official

Fred Jenkins, Jr., Ph.D.

U.S. Environmental Protection Agency
Office of Science Coordination & Policy
FIFRA Scientific Advisory Panel
EPA East Building, MC 7201M
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone: 202-564-3327
Fax: 202-564-8382
Email: jenkins.fred@epa.gov

FIFRA Scientific Advisory Panel Members

K. Barry Delclos, Ph.D.

Pharmacologist
National Center for Toxicological Research
U.S. Food and Drug Administration
Jefferson, AR

Marion F. Ehrich, Ph.D., D.A.B.T, A.T.S
Co-director, Laboratory for Neurotoxicity Studies
Professor
Pharmacology and Toxicology
Department of Biomedical Sciences & Pathobiology
Virginia-Maryland Regional College of Veterinary Medicine
Blacksburg, VA

Stephen Klaine, Ph.D.
Professor & Director
Environmental Toxicology & Biological Sciences
Clemson Institute of Environmental Toxicology
Pendleton, SC

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

FQPA Science Review Board Members

Sherril Green, D.V.M, Ph.D.
Professor & Chair
Department of Comparative Medicine
Stanford University School of Medicine
Stanford, CA

Nelson D. Horseman, Ph.D.
Professor
Department of Molecular and Cellular Physiology
University of Cincinnati
Cincinnati, OH

Jack Leonard, Ph.D.
Professor
University of Massachusetts School of Medicine
Worcester, MA

Deborah MacLatchy, Ph.D.
Vice-President: Academic & Provost
Professor of Biology
Wilfrid Laurier University
Waterloo, Ontario, Canada

Reynaldo Patino, Ph.D.

Research Scientist/Professor
U.S. Geological Survey
Texas Tech University
Lubbock, TX

Aldert H. Piersma, Ph.D.

Professor
Center for Health Protection
National Institute for Public Health and the Environment
Bilthoven
The Netherlands

Katherine F. Roby, Ph.D.

Associate Professor and Director
Reproductive Endocrinology Laboratory
Women's Health Specialty Center
Division of Reproductive Endocrinology & Infertility of the Department of Obstetrics & Gynecology
University of Kansas Hospital
Kansas City, KS

Geoffrey Scott, Ph.D.

Director
Charleston Center for Coastal Environmental Health & Biomolecular Research
National Oceanic & Atmospheric Administration, National Ocean Services
Charleston, SC

Daniel Selvage, Ph.D.

Associate Professor, Pharmacology
College of Pharmacy
University of New England
Portland, ME

Terry Wayne Schultz, Ph.D.

Emeritus Professor
Department of Comparative Medicine
University of Tennessee
Knoxville, TN

Donald Tillitt, Ph.D.

Branch Chief
Biochemistry and Physiology Branch
Columbia Environmental Research Center, Biological Resources Division
U.S. Department of Interior, Geological Survey
Columbia, MO

INTRODUCTION

On May 21-23, 2013 the FIFRA Scientific Advisory Panel (SAP) met to address scientific issues associated with the “Endocrine Disruptor Screening Program (EDSP) Tier 1 Screening Assays and Battery Performance.” Under the Federal Food, Drug, and Cosmetic Act (FFDCA), section 408(p) and the Safe Drinking Water Act (SDWA), section 1457, the EPA is required to screen all pesticide chemicals (active and inert ingredients) and those drinking water contaminants to which a “substantial population” is exposed for the potential to interact with the endocrine system. As recommended by a Federal Advisory Committee, (Endocrine Disruptor Screening and Testing Advisory Committee, EDSTAC), the EPA Endocrine Disruptor Screening Program (EDSP) established a two tiered screening and testing program to address the potential of chemicals to perturb the estrogen, androgen or thyroid (E, A, or T) systems and elicit adverse human and ecological health outcomes. This FIFRA SAP review focused on a subset of the initial Tier 1 screening data received by the Agency in response to test orders issued for the first list of chemicals in 2009. The SAP provided advice and recommendations to the Agency regarding the performance of the 11 Tier 1 screening assays and the performance of the assays as a battery that was designed to detect the potential of a test chemical to interact with the E, A, or T, hormonal pathways. Opening remarks at the meeting were provided by: David Dix, Ph.D., Acting Director, Office of Science Coordination and Policy; Steven Bradbury, Ph.D., Director, Office of Pesticide Programs; and Mary Manibusan, Director, Exposure Assessment Coordination and Policy Division, Office of Science Coordination and Policy

US EPA presentations were provided by the following staff:

Gregory Akerman, Ph.D
Catherine Aubee, M.P.A
Amy Blankinship, M.S
John Liccione, Ph.D
Tom Steeger, Ph.D,
Leslie Touart, Ph.D.

Public Commenters

Oral Public comments were provided by (provided in the order that they presented)

Steve Levine, Ph.D. of Monsanto on behalf of the Endocrine Policy Forum
Sue Yi, Ph.D. of Syngenta Crop Protection on behalf of the Endocrine Policy Forum
Sue Marty, Ph.D. of the Dow Chemical Company on behalf of the Endocrine Policy Forum
Barb Neal, DABT of Exponent on behalf of the Endocrine Policy Forum
John Brausch of BASF on behalf of the Endocrine Policy Forum
Allen Olmstead, Ph.D. of Bayer CropScience on behalf of the Endocrine Policy Forum
Ellen Mihaich, Ph.D., DABT of the Environmental and Regulatory Resources on behalf of the Endocrine Policy Forum
Lisa Ortego, Ph.D., DABT of Bayer CropScience on behalf of the Endocrine Policy Forum
Christopher Borgert, Ph.D. Applied Pharmacology and Toxicology, Inc. on behalf of the Endocrine Policy Forum
Clare Thorp, Ph.D. on behalf of CropLife America
Scott Slaughter on behalf of the Center for Regulatory Effectiveness
Catherine Willett, Ph.D. on behalf of the Humane Society of the United States
Patricia Bishop, M.S. on behalf of the People for Ethical Treatment of Animals
Colleen Toole, Ph.D. on behalf of CeeTox
Leah Zorrilla, Ph.D. on behalf of Integrated Laboratory Systems
Michael L. Dourson, Ph.D., DABT, ATS on behalf of Toxicology Excellence for Risk Assessment (TERA)

Written Public Comments were provide by:

Will Davies on behalf of LSR Associates Ltd
Richard A. Becker, Ph.D., DABT and Emily V. Tiplado, MA on behalf of the American Chemistry Council

SUMMARY OF PANEL RECOMMENDATIONS

Charge Question 1. *Based on the analysis of the data presented in Section III, please comment on the proficiency of the contributing laboratories to execute each assay in accordance with the test guidelines and achieve the performance criteria.*

Panel Summary

The Panel noted that there are several limitations which hampered their ability to determine how well the laboratories achieved the proficiency criteria. These limitations were based upon the many variations among the laboratories conducting the assays, including differences in their: 1) methodologies, 2) interpretation of the assay results, and 3) adherence to the assay guidelines. The Panel also mentioned that the way in which the data were presented in the Agency's White Paper made it difficult to identify which laboratory tested particular chemicals. Because of this lack of information the Panel was unable to compare the laboratories' proficiencies in conducting the assays.

Concerning the Fish Short Term Reproductive Bioassay (FSTRA) and the Amphibian Metamorphosis Assay (AMA), the Panel remarked that there are major deficiencies among the laboratories in the dosing conformance of these assays. The Panel specifically expressed the concern that if laboratories alter the rate of volume replacement, the total volume of chemical delivered per aquarium per day will be different. This could potentially cause a variation in the effects observed. Thus, the Panel recommended that the parameters of the protocol should assure that the initial target nominal dose is being delivered throughout the experiment.

Lastly, the Panel pointed out two primary specific concerns associated with the Amphibian Metamorphosis Assay. Their first concern was the high rates of tail curvature observed in the controls and exposure groups. As noted by the Panel, this issue was obviously related to the organisms' diet as well as their source of origin. Their second concern was how to use, or not use, the data which demonstrated potential effects from the solvents used to increase the test chemical's solubility. The Panel recommended that the Agency address these concerns in the performance criteria.

Charge Question 2. The performance criteria for each *in vitro* assay are clearly stated in the test guidelines for the ER binding (OCSPP 890.1250), AR binding (OCSPP 890.1150), ER α Transcriptional Activation (OCSPP 890.1300, OECD 455), H295R Steroidogenesis (OCSPP 890.1550) and aromatase human recombinant (OCSPP 890.1200) assays. Although contributing laboratories did not always demonstrate that results were within the specified boundaries of the performance criteria, the majority of the deviations were still close to the performance criteria. In this regard, the EPA concluded that the data were still adequate for use. *Please comment on the EPA's conclusion. Please comment on when a deviation from the recommended performance criteria would render the study unreliable.*

Panel Summary

The Panel was in general agreement with the Agency's determination that, even though there were circumstances in which the performance criteria were not met, these performance criteria deviations were deemed as minor and did not impact the interpretation or reliability of the data. Nevertheless, the Panel's verdict encompassed two caveats/concerns including: 1) a determination based on the limited data set of 21 chemicals, and 2) a limited ability to interpret the data because of the manner in which the data were presented in the White Paper (i.e., the lack of direction and magnitude of the measured values as compared to control values).

The Panel recommended that Tier I guidance be developed to indicate the point at which deviation from the norm is appropriate for an individual run to be considered acceptable. Such guidance should alleviate the inclusion of cases in which there are questions about assay parameters. The Panel noted that this guidance could be in the form of a decision tree that would assist laboratories in deciding when to redo individual runs that have not met the criteria.

The Panel questioned the redundancy of the Estrogen Receptor (ER) binding and ER α Transcriptional Activation (ERTA) assays used in Tier I in the context of the Rainbow Trout Estrogen Receptor (rtER) binding and liver slice gene activation assays used in the EDSP development of the Computational Toxicology Tools. Specifically, the Panel advised the Agency to consider eliminating the Tier I *in vitro* assays related to the ER-pathways based on the premise that "the hormone binding domain of the ER is highly conserved across species." The Panel's recommendation to eliminate the ER-binding and ERTA assays from the battery of Tier 1 assays is further supported by the Agency's interest in using high throughput screening data to expand the structural domain used in the prescreening activities. The Panel recommended that the Agency consider replacing the current Tier I *in vitro* assays related to the AR-pathways (i.e., one based on rat prostate cytosol) with one based on a recombinant cell line. Replacing the current AR-binding assay bypasses addressing the major shortcomings of the protocol, namely the limitations imposed by the preparation, attaining performance criteria, and storage life for the prostate cytosolic preparations. Moreover, replacement of AR-binding assay decreases the numbers of animals required to complete the Tier 1 testing.

Charge Question 3: A positive control is not required for the male and female pubertal assays. For these *in vivo* assays with rats, the coefficient of variation limits are specified in the test guidelines for most endpoints. Submissions from different laboratories sometimes fell short of meeting all the test guideline-recommended Coefficient of Variation (CV) limits for the endpoints evaluated. However, in most cases these shortcomings were considered of minor importance to the overall results, and the EPA concluded that the data are still adequate for endocrine screening. *Please comment on when a deviation from the recommended CV limits would render the study unreliable.*

Panel Summary

The Panel generally concurred that although different laboratories fell short of meeting all the guideline-recommended CV values for endpoints evaluated, the data from the assays were adequate for endocrine screening. Nevertheless, the Panel cautioned that there were many challenges in accepting data that comes short of meeting the performance criterion for CVs. For example, the Panel believed that accepting large CVs, weakens the statistical power of the tests. This can make the assays undependable for the detection of a biologically significant effect.

The Panel agreed with EPA's conclusion that data generated using the male and female pubertal assays are useful. However, the Panel advised that the guidelines be carefully modified and cautioned that significant care be applied in interpreting the results of these assays.

Charge Question 4: The test guidelines for the six *in vivo* assays (Hershberger assay - OCSPP 890.1400, OECD 441; Uterotrophic assay- OCSPP 890.1600, OECD 440; Male Pubertal assay- OCSPP 890.1500; Female Pubertal assay - OCSPP 890.1450; FSTRA - OCSPP 890.1350, OECD 229 and AMA - OCSPP 890.1100) offer some guidance on setting the dose/concentration range when testing for specific effects on the E, A, or T signaling pathways. In some of the *in vivo* assays, overt toxicity was noted based on effects on growth, other sublethal effects, and even mortality at the highest dose/concentration tested. Positive Tier 1 findings indicating the potential for endocrine activity can be difficult to interpret in the presence of overt toxicity.

Please comment on the nature and severity of overt toxicity that would render study results unreliable for the purpose of Tier 1 screening of a chemical for endocrine activity. Please comment on what, if any, additional guidance is needed to minimize overt toxicity while ensuring adequate dose selection for these screening assays. Also, specifically comment on the validity of including treatment concentrations with apparent body weight effects in the analysis for potential endocrine interaction for FSTRA.

Panel Summary

There is much utility in using a broad dose range to evaluate the potential endocrine disrupting effects of test compounds. Nevertheless, overt toxicity seen during testing obscures the interpretation of data from the EDSP Tier I *in vivo* assays. For instance, if effects are observed only within a dose group with overt toxicity and in the only screen that shows a positive effect, it will be complicated to determine whether the compound should be evaluated via Tier 2 testing without a repeat experiment at a lower dose. Thus, the Panel concurred that it is necessary to reduce the test compound to sub-toxic levels to enable an independent evaluation of the target agent on endocrine function.

For specific criteria in defining maximum-tolerated-doses (MTD) in the EDSP assays, the Panel suggested that the common standard of practice/animal welfare benchmark for short-term toxicity studies is a 10% upper limit of body weight loss. This cut off is considered to be adequate to determine maximum-tolerated-doses, and it is a commonly accepted standard by most institutional research animal use and care committees (Chapman et al., 2013). The Panel recommends that this 10% benchmark be used as general guideline only for mammalian assays and not the FSTRA or AMA.

The Panel made several general recommendations regarding dose selection for the *in vivo* assays in the EDSP. When overt toxicity occurs it was suggested that the testing labs employ two additional lower doses below the original MTD used to generate the dataset, thus increasing the probability of using concentrations that do not illicit overt toxicity. If effects on A, E or T endpoints are still observed in the absence of overt toxicity, then there is a higher degree of confidence that the effects on A, E, and T are directly related to the mode of action of the chemical as opposed to a more generalized stress response. If, on the other hand, effects on A, E and T are not observed in the lower dosing regimen, then the original dataset obtained at toxic MTD levels should not be considered to be definitive for direct effects via alterations in hormonal pathways.

Charge Question 5. Spinal curvature, usually manifesting as “bent tail” in *X. laevis* tadpoles, was reported in 15 of 18 AMA studies reviewed thus far. The anomaly appears to be first observed several days after study initiation, and prevalence increases with time. Overall, the prevalence of spinal curvature in these studies ranged from “a few per replicate” to 92% of a given treatment group by test termination. Experimental work by the EPA Office of Research and Development suggests that overfeeding can be a primary cause of spinal curvature in their *Xenopus* test populations; however, spinal curvature remained prevalent (range: 16-92%) in the five industry AMA studies in which feed was reduced by 50% compared to guideline recommendations.

Overall, the incidence of spinal curvature appears to be highly variable. From a qualitative review of the data, there appear to be no consistent differences in the incidence or variability of spinal curvature when studies using guideline versus reduced feeding regimes are compared. *Please comment on whether the presence or prevalence of spinal curvature in test specimens, including controls, compromises the utility or validity of an AMA submission. If so, when does the prevalence of spinal curvature render the study unreliable? What technical guidance may be useful for laboratories in reducing the occurrence of spinal curvature and determining if, or at what point within the study, a study may be compromised because of this phenomenon?*

Panel Summary

The Panel advised that the Agency keep track of the clutch performance (including number and quality of embryos from each spawn) and that the Agency select high quality, healthy breeding pairs. Eliminating breeding pairs that produce bent tail should minimize its occurrence. Specifically minimizing the use of wild-caught breeding stock may be beneficial.

The Panel also recommended that there should be a careful review of water quality conditions under which bent tail tadpoles are kept. Temperature, pH, and nutritional deficiencies (vitamin C and Ca) are reported to cause bent tail in laboratory reared American bullfrogs, Leopard frogs (Marshall et al, 1980; Leibovitz et al, 1992; Martinez et al, 1992), and in several other amphibian species. One panel member noted that, based on personal observation, temperature and pH have appeared to be associated with the condition in *Xenopus tropicalis* and water softness (Ca and K) associated with the condition in *Xenopus laevis*. Over-feeding can alter the water biochemistry and uneaten food and excess waste causes changes in ammonia and pH. These changes are detrimental to the health of the test animals; thus, they could potentially explain the bent tail seen in the laboratories that reported a decrease in the occurrence of the condition when food was decreased. Feed and water sources should be screened for herbicides, biological toxins and pesticides currently known to cause bent tail in *Xenopus laevis*, nominally. The Panel recommended further monitoring the portion of each clutch not used in a given study. It would be potentially useful to report these results. The Panel acknowledged that some information on this concept is in the guidance document, but recommended that it be elaborated further. The Panel further recommended that for acceptance of the AMA test results overall, the clutch should not have a rate of mortality and morbidity > 20%.

Charge Question 6. With the exception of thyroid gross pathology findings (thyroid gland atrophy and hypertrophy) in the AMA, severity grades are generally assigned based on comparison to “normal” *X. laevis* thyroid findings depicted in the guidance or based on the professional opinion of the pathologist conducting the assessment; they are not assigned in comparison to concurrent control findings from a given study. (Please refer to Section III.2.f in the document entitled “Interpreting Amphibian Thyroid Histopathology Diagnoses” and supporting documents, OECD Guidance Document on Amphibian Thyroid Histology No. 82, 2007 and Grim et al., 2009).

- a. *In one study, the pathologist’s report identified a lower incidence and severity of follicular cell hypertrophy when compared to the incidence and severity of this trait in control specimens. Similar trends have been observed in other studies. In this case, the pathologist concluded that the finding was potentially consistent with treatment-related delay of metamorphosis because thyroid follicular cells normally increase in height during tadpole development. Please comment on the validity of this conclusion.*

Panel Summary

Even though thyroid hormone agonistic activity could provide a justification for the pathologist’s observation of a lower incidence and severity of thyrocyte hypertrophy via negative feedback inhibition of the hypothalamic-pituitary-thyroidal (HPT) axis, this circumstance would also have to be complemented by signs of accelerated development in the experimental population. However, since accelerated development did not appear to occur, the Panel concurred with the pathologist’s conclusion that the finding is consistent with a treatment-related retardation of metamorphosis via disruption of the HPT axis. As demonstrated in the public literature, this case may represent an unusual disruption of the axis that is not based on classical anti-thyroid mechanisms such as those observed after exposure to thyroid synthesis inhibitors (goitrogens), but on a failure of the axis to activate normally or fully (Sharma and Patiño 2008).

- b. *What guidance may be given to better distinguish between histological changes in the thyroid associated with the normal progression of metamorphosis and treatment-related effects? Are there certain lesions or diagnoses which may, by their absence or lessened severity as compared to controls, be indicative of treatment-related HPT effects such as delayed metamorphosis?*

Panel Summary

The Panel acknowledged that thyroid activity normally intensifies during metamorphosis, and consequently it is relatively challenging to perform this activity as a control condition. The Panel believed that a marker of normal thyroid activity that does not change during metamorphosis would be ideally useful. The Panel was unaware of any standard histological feature of the thyroid that does not normally change during tadpole development. However, the Panel noted that the T4-immunoreactivity has been demonstrated to concentrate in a ring at the periphery of the colloid and that the intensity of this ring remains fairly constant during metamorphosis (Hu et al., 2006). Thus, the Panel recommended that the Agency explore the T4-immunoreactive ring as a potential marker. They noted, however, that the T4-immunoreactive ring has not been validated for standard testing. They also noted that the T4-immunoreactive ring could possibly provide a benefit to the thyroid assessments in the rat pubertal assays.

Lastly, the Panel recommended that the Agency consider the use of quantitative thyroid activity measurements as opposed to the semiquantitative grading scheme currently used which may be prone to problems with inconsistencies.

Charge Question 7. In 2008, the SAP acknowledged that the *in vivo* assays included in the Tier 1 battery provide both redundancy and complementarity for evaluating interactions with the E, A, or T signaling pathways. The panel also noted that all of the Tier 1 assays and the broad range of endpoints appeared to be necessary to “discriminate positive and negative results”.

- a. *Please comment on the battery performance with respect to the anticipated complementary nature of the more complex, multi-parameter in vivo assays in the context of the observed responses with the case studies. Please comment separately on the E-, A-, and T-related assays and endpoints.*

Panel Summary

The Panel observed that the Tier I assays did not provide information on the magnitude or the direction of the effects. Thus, they noted that it was difficult for them to determine how useful the observed complementary effects will be when applied in the weight-of-evidence (WoE) analysis to strengthen or weaken the case for the E, A, or T activity of these test compounds and determine the need for further testing. Specifically, in regard to the E,A, and T-individual assays and their endpoints, the Panel noted that the prevalence of overt toxicity in each of these assays appeared to confound the interpretation of the results for many of the chemicals tested.

- b. Please comment on the battery performance with respect to the anticipated redundancy across the 11 assays in the context of the observed responses with the case studies. Please comment separately on the E-, A-, and T-related assays and endpoints.

Panel Summary

The Panel concluded that due to the manner in which the data was presented, they could only provide limited input on the performance of the battery. Particularly, without information on the magnitude or direction of change, the Panel determined that it did not seem possible to provide a thorough judgment on how the battery performed.

- c. *The EPA concluded that the battery has performed as anticipated by the 2008 SAP. Please comment on this conclusion.*

Panel Summary

The Panel generally agreed that the battery performed as expected with the subset of chemicals on which data summaries were provided. However, the Panel emphasized that only limited conclusions can be drawn at this time based on the narrow amount of information that has been provided.

Many panel members strongly expressed their belief that positive controls that are well known to interact with E, A, and T pathways should have been used to evaluate the entire battery just as what was done in the validation of each individual assay. The ability of the battery to clearly detect such known compounds would have been useful in providing confidence that the battery functioned as intended and serve as a means for determining which assays might be considered for removal or replacement.

Charge Question 8. The EPA is committed to minimizing animal usage in the screening battery while maintaining the effectiveness of the battery to answer the question of whether a chemical has the “potential” to interact with the endocrine system.

- a. *In 1998, the EDSTAC described the conceptual framework for Tier 1 assays and recommended the strategy to “require the minimal number of screens and tests necessary to make sound decisions, thereby reducing the time needed to make these decisions”, and that the screens should be conducted at a minimal cost necessary to make decisions. Based on the preliminary battery performance evaluation, to what extent can the current Tier 1 battery of 11 assays be modified to reduce animal usage and/or lower cost while adequately ensuring the EPA’s ability to answer the question of “whether a chemical has the potential to interact with the endocrine system?” More specifically, please comment on whether the Uterotrophic and Hershberger assays provide necessary redundancies in the Tier 1 battery based on this preliminary analysis. Please include in your comments what information may be lost and what uncertainties may be introduced by absence of either or both of these assays.*

Panel Summary

While the Uterotrophic or Hershberger assays are limited in what they are able to detect, both of these assays are well-accepted tests of Gonadal hormone action on peripheral tissues. Although the reduction of animal usage and unnecessary testing are big concerns, the Panel thought that the assays should be maintained since they provide for specific testing of the actions of potential endocrine disrupting agents and their metabolites in a mammalian system. Such testing can provide more definitive information regarding the mode of action (MOA) of the test compound than what can be learned from the more comprehensive pubertal assays. At a minimum, a decision to remove these assays from the battery should be postponed until the completion of the Agency’s WoE analysis of these compounds to determine what role these assays play in the decision process.

- b. *Please comment on the scientific criteria the Agency should consider in evaluating necessary redundancies and eliminating assays from the current battery.*

Panel Summary

The Panel advised that the decision to remove an assay from the battery could likely be made after an exhaustive review of all the data generated from the complete set of compounds that have been run through the Tier I screens. As a consequence, an assay should be removed if it consistently produces results that are not clearly interpretable. Concerning the potential of modifying the current battery, the Panel strongly stressed that the Agency should do a more extensive data evaluation in order to best determine the range of possibilities for modification.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

Individual EDSP Tier 1 Assay Performance

Charge Question 1. The results of the EPA's evaluation to determine how well each Tier 1 assay was executed relative to defined performance criteria in the respective test guideline are presented in Section III of the white paper. This evaluation was based on a subset of pesticide active ingredients that represent a broad range of physical-chemical properties (e.g., water solubility, octanol-water partition coefficients, volatility), as well as a range of chemical classes and pesticidal activities. This preliminary analysis generally demonstrates that the laboratories executing the assays can meet the performance criteria defined in the test guidelines and the results provide useful information to determine whether a chemical has the potential interact with the endocrine system. During this initial analysis of individual study data, however, EPA identified a few issues with some assays. The following questions seek comment from the FIFRA SAP on assay (laboratory) performance and advice on how to address technical deviations and other issues encountered with specific assays. *Based on the analysis of the data presented in Section III, please comment on the proficiency of the contributing laboratories to execute each assay in accordance with the test guidelines and achieve the performance criteria.*

Panel Response

Established guidelines for each test in the Tier 1 battery define test parameters, criteria for assessing testing proficiency, and criteria for acceptable limits for various positive and negative controls to be included in most tests. Thus, general proficiency of the testing laboratories can be monitored by several end-points.

Based on the information presented in Section III of the White Paper, a general assessment indicates that the testing laboratories demonstrate proficiency in carrying out the battery of tests. However, there were variations across laboratories in their ability to achieve proficiency criteria, raising some concerns.

The overall proficiency of the contributing laboratories to execute each assay in accordance with the test guidelines and achieve the performance criteria was variable according to each *in vitro* and *in vivo* test. These differences in proficiency across testing laboratories were primarily attributable to slight variations in methodology, interpretation, or follow-through of testing guidance. For example, when solubility issues were encountered one laboratory did not adhere to the guidance to reset with the recommended half-log increment dilution; instead, they omitted the higher log dose and added an additional lower dose at the end of the dose series. Utilization of the mid-log approach at the upper end of the concentration range would better characterize the response curve and provide more confidence in the interpretation of the results. Additional variation may also be introduced due to an individual laboratory's conditions; such as an inability to maintain water temperature requirements to the $\pm 1^{\circ}\text{C}$ in the fish tests. Further, variation within a given laboratory may occur based on the proficiency of

different analysts performing the test and thus, individual performance criteria for each analyst should be documented and recorded over time.

Given the manner in which the data are presented in Section III, it is difficult to determine what laboratory tested what chemical. In addition, within the description of a given test, there are places in the text where the number of testing laboratories and the numbers of chemicals tested do not agree, further complicating the assessment of laboratory proficiency. Further, it appears that each test chemical was assessed by only one laboratory. Having more than one laboratory test the set of chemicals, or even a subset of chemicals, would provide an additional monitor for proficiency. Based on this lack of multi-laboratory testing for chemicals, the Panel was only able to comment on the proficiency for an individual laboratory to perform the battery.

In the evaluation of results from individual assays in Section III, especially with regard to the *in vitro* assays, multiple statements were made indicating the testing laboratories did not provide complete proficiency data with reports submitted to the Agency. The exclusion of proficiency data from some datasets weakens the ability of the Panel to carefully assess the proficiency in these instances. Examples include results presented for the Steroidogenesis and Aromatase assays. The test guidance should clearly indicate the requirement of the testing laboratory to include the complete proficiency data set with submission of the test report to the Agency. If this is not clearly stated, the guidance should be amended.

Similar to the data description for the *in vitro* tests, in evaluating the summary of the *in vivo* assays it was not clear how many laboratories participated in each test and thus conclusions about multiple laboratory proficiency cannot be made. In addition, Section III generally provided limited data for analysis. The lack of data providing direction and magnitude for test endpoints including controls, proficiency measures, and test chemicals limit the Panel's ability to carefully assess laboratory proficiency.

In regard to the aquatic tests data, again, the laboratories conducting each of the tests were not respectively identified. Consequently, variation in laboratory proficiency cannot be judged. As with the other Tier 1 tests, the aquatic tests had variation around performance criteria and, similar to the other tests, clear definition of acceptable limits of deviation around the performance criteria are needed. It is understood that the acceptable deviation may differ for each test and endpoint measure.

In many cases, including both the *in vitro* and *in vivo* tests, the established performance criteria were not met. As described in Section III, deviation from the performance criteria was most often considered to be 'slight' or 'minor'. The Agency concluded these instances of slight deviation from the established criteria did not affect interpretation of study outcome. The Panel was in agreement with the conclusion by the Agency and also believed these slight variations did not impact interpretation of the data. However, describing the deviation around performance criteria as "slight" or "minor" should be avoided in favor of quantitative definitions for the variation. An example might be defining minor and major non-conformities and then define quantitatively what these two

ranges of deviation from norm include, such as : (1) based upon a certain percentile deviation from each performance criteria (e.g., minor < 10% different and major > 10% different); or (2) a statistical based estimation (such as +/- 1 or 2 standard deviations). Specific criteria should be established to aid in the identification of unreliable test runs and the interpretation of successful runs. In addition, more specifically in regard to CVs reported in *in vivo* studies, criteria defining acceptable performance CVs may need to be revisited to account for the variations existing in control populations.

In addition to further defining non-conformities, it was considered that a decision-tree approach would allow for clear assessment of when test criteria have or have not been met and when a given test can be accepted or rejected, or should be rerun. For example, when negative controls included in a test run are read as positive, as indicated in at least one instance in the data summary provided, the run should likely be rejected. In addition, guidance on how to utilize the data in the presence of cytotoxicity and toxicity should be included.

Although the majority of the deviation from the proficiency criteria was not considered to impact the interpretation of the data, there were some deviations that may be cause for concern. In regard to the FSTRA and the AMA, there are major deficiencies in dosing conformance. If this was due to analytical limitations, absorption or biodegradation, these processes need to be characterized. There was some concern that if laboratories alter the rate of volume replacement, the total volume of chemical delivered per aquarium per day will be different and this will be a source of variation, in effect. The parameters of the protocol should assure that the initial target nominal dose is being delivered. An additional measurement in aquaria not containing test animals is recommended, although this will add additional cost. As mentioned by EPA, there are routine stock concentration measurement requirements as well as daily flow rate determinations. This combination of measurements should be reported and could be used to predict the estimated nominal concentration being delivered throughout each bioassay. In addition, this would at least provide an accurate estimate of the delivered nominal dose, without having to measure it. Also, it is important to try to use only one stock for dosing purposes. A second concern with the FSTRA relates to the cumulative effects of several non-conformities, including low dissolved oxygen levels, elevated exposure temperatures, and elevated levels of chlorine. The fish will respond to these changes with increased ventilation rate that may damage the fish gills and may also result in increased uptake of the chemical being tested. An additional concern is that the accumulation of multiple non-conformities may increase the potential for false positives. These three mentioned non-conformities for performance criteria are easily corrected and participating laboratories should have taken steps to better assure compliance with these criteria. The Agency may consider evaluating and identifying combinations of cumulative non-conformities which may rise to a level of non-conformity requiring re-testing.

There were two primary concerns related to the performance of the AMA. The first concern relates to the high rates of tail curvatures observed in controls and exposure groups. This high incidence raises the question as to the health of the animals being used in this bioassay. There are obvious issues with this abnormality being tied to diet as well

as the source of the test organisms. EPA should examine the underlying health of frogs post metamorphosis since the tail is reabsorbed during the developmental process. If this abnormality does not affect growth, health, and survival later in life, this issue may not be as significant as it would be if the animals were affected later in life. Also, maintaining accurate records of the source of frogs is a critical performance criterion that should be included. The second concern is related to the potential effects of carriers used to increase chemical solubility. The performance criteria presented clearly demonstrated that it is possible to discern this effect with the current assay. The question is how to use or not use the data when a carrier effect is observed. EPA should formally address this in the performance criteria, if they have not already done so. For example, if carrier effects are observed, an end point analysis between carrier control and chemical should be conducted. This should be clearly informed within the guidance.

Charge Question 2. The performance criteria for each *in vitro* assay are clearly stated in the test guidelines for the ER binding (OCSPP 890.1250), AR binding (OCSPP 890.1150), ER α Transcriptional Activation (OCSPP 890.1300, OECD 455), H295R Steroidogenesis (OCSPP 890.1550) and Aromatase Human Recombinant (OCSPP 890.1200) assays. Although contributing laboratories did not always demonstrate that results were within the specified boundaries of the performance criteria, the majority of the deviations were still close to the performance criteria. In this regard, the EPA concluded that the data were still adequate for use. *Please comment on the EPA's conclusion. Please comment on when a deviation from the recommended performance criteria would render the study unreliable.*

Panel Response

Charge Question 2 is specifically related to evaluating how well each Tier 1 assay test guideline can be consistently executed based on defined performance criteria and to highlighting any issues associated with interpretation of responses within each assay.

As noted in the Endocrine Disruptor Screening Program (EDSP) White Paper, the EDSP Tier 1 screening consists of a battery of eleven assays (see White Paper Table 1 "Current EDSP Tier 1 Screening Assays"). However, only the five *in vitro* assays are addressed in Question 2 which include the: 1) Estrogen Receptor (ER) binding assay, 2) ER Transcriptional Activation assay (ERTA), 3) Androgen Receptor (AR) binding assay, 4) Aromatase assay, and 5) Steroidogenesis assay.

Tier 1 screening consists of a battery of less complex *in vitro* and *in vivo* assays designed to effectively and efficiently detect the potential of a chemical to interact with E, A, or T hormonal pathways. Tier I assays are also intended to determine, based on a WoE analysis, whether a chemical is a candidate for Tier 2 testing. In contrast, Tier 2 testing involves more complex *in vivo* studies to clarify the potential interaction with the endocrine or non-endocrine systems and to establish dose-response relationships that may be needed for risk assessment.

As such, Tier 1 screening lies between: 1) prescreening or prioritization with *in silico* tools, which identify chemicals with the potential to be hazardous via select endocrine disruption mechanisms, and 2) more resource-intensive testing for risk assessment. Generally, Tier 1 assays are: 1) shorter duration and lower cost, 2) designed to minimize false negative, 3) report the potential to interact in a yes/no format, and 4) use limited dose regimes, focused on the higher end.

Each of the Tier 1 assays, which form the basis for this SAP was previously validated with a known set of reference compounds for endocrine screening specific to either the E, A, or T signaling pathways through a process of test method development, pre-validation, and inter-laboratory validation (US EPA, 2009). These efforts culminated in an evaluation by a 2008 FIFRA SAP (i.e. March 25-28, 2008 Session of the FIFRA SAP entitled “Review of the Endocrine Disruptor Screening Program (EDSP) Proposed Tier 1 Screening Battery”), which supported using the eleven assays to initiate Tier 1 screening (US EPA, 2008).

The Panel understood that one aspect of the current SAP focuses on current assay performance analysis, specifically the proficiency of the contributing laboratories to execute each Tier 1 assay in accordance with the test guidelines. This includes the proper execution of the performance criteria and the generation of the expected results for those assays utilizing positive controls. Additionally, the Panel understood that while Tier 1 screening is linked to a WoE analysis, WoE analysis is not part of this SAP.

The Panel congratulated the EPA on providing a summary of the results for the 21 compounds brought forth as part of this SAP for the eleven Tier I assays (see section III “Evaluation of Results from Individual Assays” of the White Paper). The Panel also appreciated the public commenters' input, which in many cases, the Panel found to be highly informative.

The Panel acknowledged that the 21 test compounds evaluated as part of this SAP represent the fifty active and two inert ingredients that have been screened in the Tier 1 battery as an intermediate step toward the 67 pesticides ordered for testing by the Agency in 2009 in follow up to the recommendation by the 2008 SAP. While the Agency stated in the White Paper that the 21 chemicals are representative of the 52 chemical set selected for Tier 1 testing, especially in regard to physico-chemical properties and modes of toxic action, the ability of the Panel to address this and other charge questions is impacted by this limited data set. Moreover, because of a variety of issues, not all 21 compounds were assessed in each *in vitro* assay associated with the Tier 1 battery.

The Panel recognized that while the assays *in toto* were designed to identify whether a chemical has the potential to interact with estrogen, androgen or thyroid (E, A and T) hormone systems, the *in vitro* assays, however, focus primarily on the E and A systems due to the characteristics of their endpoints. Furthermore, the Panel understood that not all *in vitro* and *in vivo* data are of equal value, and the *in vitro* data may be or may not be informative to results from the *in vivo* assays. Specifically, the Tier 1 *in vitro* assays are

intended to provide some mechanistic data for single known pathways, while the *in vivo* Tier 1 assays are more amenable to detecting multiple mechanisms of action.

As noted in Table 1 and in the White Paper, the *in vitro* assays in the current EDSP Tier 1 screening battery address only the E and A systems by quantifying the key endpoints of the most likely mode of action (e.g., receptor binding as well as steroidogenesis).

Table 1. Endocrine Effect Assessment by the Tier 1 Assays

					<i>Steroid</i>	<i>Synthesis</i>		
	<i>E</i>	<i>E-</i>	<i>A</i>	<i>A-</i>	<i>A</i>	<i>E</i>	<i>HPG</i>	<i>HPT</i>
<i>In vitro</i>								
<i>ER Binding</i>	X	X						
<i>ER Transcriptional Activation</i>	X							
<i>AR Binding</i>			X	X				
<i>Steroidogenesis (H295R)</i>					X	X		
<i>Aromatase (Recombinant)</i>						X		
<i>In vivo</i>								
<i>Uterotrophic</i>	X							
<i>Hershberger</i>			X	X				
<i>Pubertal male</i>			X	X	X		X	X
<i>Pubertal female</i>	X	X				X	X	X
<i>Fish Reproductive Screen</i>	X	X	X	X	X	X	X	
<i>Amphibian Metamorphosis</i>								X

Specific to the estrogen system are four *in vitro* assays. The ER binding assay OCSPP 890.1250 is a rat uterine cytosol assay that has the potential to detect ER α or ER β binding but cannot distinguish between estrogen agonists and antagonists. The ERTA assay OCSPP 890.1300 is a human cervical tumor cell (Hela 9903) assay that has the potential to detect estrogen agonists through ER α binding and subsequent gene transcription. The aromatase assay OCSPP 890.1200 is a human microsome assay that has the potential to detect the inhibition of the enzyme aromatase that converts androgen to estrogen. The Steroidogenesis assay OCSPP 8980.1550, is a human cell line H295R assay that has the potential to detect a change in production (i.e., increase or decrease) of testosterone and estradiol.

There are three *in vitro* assays specific to the androgen system. The AR Binding assay OCSPP 890.1150 is a rat prostate cytosol assay that has the potential to detect androgen receptor binding but cannot distinguish between androgen agonists and antagonists. The previously noted Aromatase Assay has the potential to detect the inhibition of conversion of androgen to estrogen and the previously noted Steroidogenesis assay has the potential to detect an increase or decrease in production of testosterone and estradiol.

The Panel recognized that there are no *in vitro* assays associated with the T pathway in the Tier I battery. Moreover, the Panel perceived that, based on the outcome of FIFRA SAPs in August 2009 and January 2013 related to the EDSP, the Agency has a much more experience with screening the ER mediated pathway than the AR mediated pathway.

The Panel acknowledged that an earlier FIFRA SAP (USEPA, 2008) reviewed the combination of EDSP Tier 1 screening assays of the current battery and concluded that the battery of assays considered the different key events (e.g., receptor binding, cell signaling, organ response) along the most common pathways for E, A, and T disruption. In addition, the degree of redundancy (i.e., concordance of endpoints across the assays that are intended to evaluate an interaction within a particular pathway, E, A or T) and complementarity (i.e., concordance of endpoints within an assay) of the assays and endpoints in the Tier I battery were likely to minimize the potential for "false negatives" and "false positives" and did not result in unnecessary redundancies for a mode of action. While the 2008 review was based on individual assay validation results with selected reference chemicals known to interact with the E, A, and T hormonal pathways, the current review focuses on the application of the entire battery of eleven assays with chemicals not necessarily known to interact with the endocrine system.

In answering Charge Question 2, the Panel has elected to address each of the five *in vitro* assays separately and then provide other comments of a more general nature.

ER-Binding Assay

As noted in the White Paper, 19-norethindrone was used by all laboratories as a replacement for the recommended weak positive control ligand, norethynodrel. Since the guideline does not contain performance criteria for 19-norethindrone the Panel recommended that the Agency establish performance criteria for 19-norethindrone.

The Panel agreed that ER-binding assay performance was by and large, consistent across all 18 test compounds and that, in general, issues were with individual runs. In addition, while a number of the evaluated assays did not meet all of the performance criteria, overall, the laboratories' performance of the ER binding assays was acceptable.

One Panel member noted that the most problematic aspect of this assay is that it uses a highly complicated and labor-intensive protocol involving extracted uterine cytosol. This assay was conducted by different labs with varying success. It is recommended that the Agency standardize this parameter of the assay if they are to continue to use this assay. The Panel believed the results and conclusions for this assay are supported in a WoE by the Agency's work with the Rainbow Trout ER-binding assay.

ER α Transcriptional Activation (ERTA) Assay

As noted earlier the ERTA is used to determine if a chemical interacts with the endocrine system by functioning as an ER α ligand and activating transcription of an estrogen-dependent reporter gene. This protocol is validated for the detection of estrogen agonists.

As noted in the White Paper the response curve parameters for the weak estrogen agonist (17 α -estradiol) were met for only 5 of the 19 chemicals. Since the out-of-range values were very close to the guideline ranges for the majority of the outliers, the Panel recommended the Agency look into revising the guideline ranges for this control concomitant with reviewing whether the issues are with laboratory performance (minimizing the need to revise the guidelines and focusing more on laboratory compliance with the guidelines).

The Panel generally agreed that the majority of the deviations in the ERTA can be considered minor. However, some panel members raised concern about the cytotoxicity observed with corticosterone (Laboratory 1) and 17 α -methyltestosterone (Laboratories 1, 4 and 5) as they impact interpretation of the results.

Androgen Receptor (AR) Binding Assay

As noted earlier, the AR binding assay consists of a saturation binding experiment and a competitive binding experiment using rat prostate cytosol as such it is analogous to the ER-binding assay. Specifically, the assay measures the ability of radiolabeled [3H]-R1881 to interact with the AR in the presence of increasing concentrations of a test chemical. While the assay cannot distinguish between AR agonists and AR antagonists,

the fact that the AR is conserved across species means that a chemical capable of binding in this assay is highly likely to affect different species.

The Panel agreed with the Agency in their findings that, overall, the laboratories performance of the AR binding assays was generally acceptable and the data for the 17 tested chemicals were reliable. The Panel agreed that these results demonstrate that the AR binding assay as performed can distinguish between chemicals that bind or do not bind to the AR *in vitro*. The robustness of this assay is demonstrated by the result that the control compounds tested appropriately in the assay, even if the assay preparations did not fully meet all the assay criteria.

Steroidogenesis Assay

As noted, the Steriodogenesis assay is an *in vitro* screen intended to identify chemicals that affect the steroidogenic pathway beginning with the sequence of reactions involved in the production of testosterone and estradiol/estrone occurring after gonadotropin hormone receptors are activated. As such, the assay is not intended to identify chemicals that affect steroidogenesis due to alterations in the hypothalamus or pituitary gland. The assay endpoints measure changes in estradiol and testosterone levels.

The Panel agreed with the Agency that the laboratories' performance of the assay was mainly consistent across all 18 test compounds, and the performance criteria were by and large generally met for all compounds. In the majority of cases where the performance criteria were not met, the values only slightly exceeded the expected values and these slight deviations did not impact the interpretation or reliability of the studies. The Agency reported in the White Paper that proficiency testing for the ERTA and Steroidogenesis *in vitro* assays was often omitted from the laboratory reports. The Panel suggested that if these proficiency compounds were routinely run and reported, the Agency would have greater confidence in the results of these tests, even if all of the performance criteria were not met. The Panel agreed that the overall results demonstrate that the Steroidogenesis assay as performed can distinguish between chemicals that alter or do not alter testosterone and/or estrogen levels *in vitro*.

Aromatase Assay

The Aromatase assay determines if a chemical could affect the endocrine system by inhibiting the catalytic activity of the aromatase enzyme, which is responsible for the conversion of androgens to estrogens. Specifically, the assay uses recombinant human microsomes containing aromatase (CYP19) and cytochrome P450 reductase, and measures the release of tritiated water during the conversion of 3H-androstenedione to estrone. As such, it measures competitive inhibition of aromatase by the tested chemical.

The Panel concurred with the Agency that, overall, the Aromatase assay as performed by the testing laboratories was able to distinguish between inhibitors and non-inhibitors of activity and the performance of the Aromatase assay was mainly consistent across all 18 test compounds. Moreover, the Panel agreed with the Agency that the performance

criteria for the assay were generally met in each study and the deviations noted were minor. Specifically, the failure, by small amounts, to stay within performance criteria ranges did not adversely impact interpretation of the results. The Agency should consider requiring assay proficiency data when drawing conclusions on the results of an assay as it pertains to the Tier 1 screen.

General Comments

As noted in the White Paper, the laboratory performance of each Tier 1 assay was evaluated across 15 to 21 chemicals. Although there were situations where performance criteria were not met, the Agency considers those deviations to be “minor” and they did not impact the interpretation or reliability of the data. As noted above, in summary the Panel was in basic agreement with the Agency’s findings. However, most panel members agreed that their enthusiasm for this endorsement is tempered by two important factors: 1) limiting the reported chemicals to 21 rather than all 52 chemicals tested and 2) the manner in which the data was presented in the White Paper (i.e., the lack of direction and magnitudes of the measured values as compared to control values).

The Panel noted that the number of test chemicals reported, 21 or less, is small. The Agency should look at adjusting the performance criteria for the individual assays to better reflect the experience gained since 2008. Specifically, the Panel recommended that Tier I guidance be developed that indicates how much deviation from the norm is appropriate for an individual run to be considered acceptable. This type of guidance should minimize cases where there are questions about assay parameters. The Panel further noted that this guidance could take the form of a decision tree that would assist labs in deciding when to redo individual runs that have not met the criteria.

Additionally, the Panel recommended that in the new guidance the Agency should refrain from using terms such as “slight” and rather report allowable percent difference from performance criteria. The Panel felt that small deviations (i.e., < 10%) are likely to be insignificant. Generally, the Panel felt that deviation from the recommended performance criteria that would render a study unreliable are likely to vary between specific assays. Moreover, the Panel generally thought that, since data for less than 21 chemicals are reported for most of the *in vitro* assays, in some cases it may be too early to set more specific reliable criteria. However, after the data for all 52 chemicals are evaluated this task should be easier.

The Panel understood that the protocols used in the Tier 1 assays represented the state of the science in 2008 when the initial Tier 1 testing order was issued. The Panel agreed with the Agency’s decision to not alter the protocol until the first round of testing was completed. While, this SAP is based on data for only 21 of the 52 chemicals being tested in this order, the Panel recommended that the Agency make all data available on the remaining 31 chemicals as soon as possible. The Panel felt strongly that having data sets for all Tier 1 tested chemicals will have a positive impact on WoE deliberations.

With that said, the Panel believed that there is sufficient evidence that a reevaluation of selected test protocols may be in order. Based on both the report from the January 2013 FIFRA SAP on the EDSP prioritization scheme and the public comments for the present SAP, the Panel questioned the redundancy of the ER-binding and ERTA assays used in Tier I in the context of the rainbow trout rER binding and liver slice gene activation assays used in the EDSP development of the Computational Toxicology Tools. Specifically, the Panel recommended that the Agency look at eliminating the Tier I *in vitro* assays related to the ER-pathways in light of the fact that “the hormone binding domain of the ER is highly conserved across species.”

The recommendation for the Agency to eliminate the ER-binding and ERTA assays from the battery of Tier 1 assays is reinforced by the Agency’s interest in using high throughput screening data to expand the structural domain used in the prescreening activities. Moreover, eliminating the binding assay eliminates the need to address the major shortcomings of the protocol, namely the limitations imposed by the preparation, attaining performance criteria and storage life for the uterine cytosolic preparations. Moreover, elimination of the ER-binding assay will reduce the numbers of animals required to complete the Tier 1 testing.

The Panel recommended that the Agency look at replacing the current Tier I *in vitro* assays related to the AR-pathways (i.e., one based on rat prostate cytosol) with one based on a recombinant cell line. Replacing the current AR-binding assay bypasses addressing the major shortcomings of the protocol, namely the limitations imposed by the preparation, attaining performance criteria, and storage life for the prostate cytosolic preparations. Moreover, replacement of AR-binding assay will reduce the numbers of animals required to complete the Tier 1 testing.

Charge Question 3: A positive control is not required for the male and female pubertal assays. For these *in vivo* assays with rats, the coefficient of variation limits are specified in the test guidelines for most endpoints. Submissions from different laboratories sometimes fell short of meeting all the test guideline-recommended Coefficient of Variation (CV) limits for the endpoints evaluated. However, in most cases these shortcomings were considered of minor importance to the overall results, and the EPA concluded that the data are still adequate for endocrine screening. *Please comment on when a deviation from the recommended CV limits would render the study unreliable.*

Panel Response

The male and female pubertal assays use rats undergoing puberty to test for potential endocrine disrupting effects by various compounds designated by the EPA for initial Tier I testing. These assays have the potential to yield valuable information, as this period is highly dependent on normal endocrine function for the successful progression of the HPG axis to sexual maturity. Moreover, the assays have an advantage over others in the EDSP in that they utilize a relatively realistic route of administration (oral gavage) to determine the potential endocrine disrupting effects of test chemicals (see White Paper for description). The Panel noted that the assays also use a significant number of animals in

order to ensure proper study populations for both the preliminary range of finding studies and for the assays themselves.

Regarding the acceptability of the data for use, in general, the Panel agreed that, although different laboratories sometimes fell short of meeting all the guideline-recommended CV for endpoints evaluated, the data from the assays were adequate for endocrine screening. However, it was noted that there are many difficulties with accepting the evidence from experiments that fall short of meeting the performance criterion for CVs, and that additional guidance is needed to clearly establish criteria defining acceptable / unacceptable tests. Ultimately, the acceptable level of variance in the screening assay needs to be sufficiently broad to capture the range of biological variability, but not so large as to allow for careless execution of the assays.

Specifically, regarding the CVs for the different endpoints measured in the assays, the Panel concurred in general with the EPA that most of those that were marginally outside of the criteria in the guidelines likely did not impact the interpretation of the data. However, it was cautioned that higher CVs in the test controls will tend to increase the likelihood of committing a type II error (that is, failing to detect a true difference, a false negative), and that decisions regarding the value of these assays are difficult to make without considering larger issues, some of which relate directly to considerations of the “weight of evidence” determinations that will be the subject of a future Panel. It was also noted that in Tier 1 screening, if a bias were to be acceptable, one might suggest that the bias would be to accept type I errors over type II errors. While doing so would likely increase the number of compounds moving forward to Tier 2, it would minimize the chances of failing to move compounds that are, in fact, endocrine disruptors. The application of redundant testing in the Tier 1 battery should be sufficient to take care of the potential type I problem, to the extent that it exists. On the other hand, there is the potential for piling multiple type II errors on top of one another if the Tier 1 battery admits multiple assays that do not meet the performance criteria.

Additionally, the Panel believed that by accepting large CVs, the extent of variability weakens the statistical power of the test and can make assays unreliable for the detection of biologically significant effects. In the event that other assays (Uterotrophic assay or Hershberger assay) that evaluate estrogenic or androgenic activity are removed from the battery, perhaps the inclusion of positive controls, or at least periodic laboratory proficiency assays with controls, should be considered, although this is not a favorable option from the standpoint of animal use reduction or cost. It was also noted that performance consistency within and between testing laboratories is important, and that this can be achieved by increasing communication among the laboratories performing the tests and the EPA. Potential laboratory-specific confounding factors contributing to high CVs that were mentioned include diet, genetics, daily care protocols, and stress responses. Further, while Sprague-Dawley rats were used in most of the pubertal assays reported, it was not clear that these rats were from the same supplier, and factors such as substrain or diet could contribute to variability in means and variances among laboratories. Consideration for comparison to laboratory-specific historical CVs for the measured endpoints may reduce deviation from the guidance criteria and provide a more

meaningful comparison for determination of the significance of chemical exposure. Moreover, assays for female rats likely exceeded CV levels due to the effects of the estrous cycle, which causes large cyclic changes in many of the endpoints being measured, particularly uterine and ovarian weights. Determination of estrous cycling was mentioned in the guidelines as a parameter to be measured, but was not addressed in the White Paper provided by the EPA.

One panel member expressed the opinion that in order to accurately identify potential endocrine disruptors rather than agents that alter the parameters tested via other pathways such as stress, potential interactions between the hypothalamic-pituitary-adrenal (HPA), hypothalamic-pituitary-thyroidal (HPT), and hypothalamic-pituitary-gonadal (HPG) axes need to be taken into account when interpreting the data from each assay. In particular, when a CV for a test chemical is high, low-level chronic stress can alter HPG axis development and function. Specifically, it was noted that the female pubertal assay is a useful tool that has the potential to indicate if an agent disrupts HPG function at a variety of levels. It is also likely the most difficult assay to interpret, as many findings could be due to factors not related to the test chemicals, including experiment-induced stress. This is especially of concern given the frequency of overt toxicity reported in the White Paper. Potential problems from stress and the chronic, low level activation of the HPA or 'stress' axis by some of these chemicals include decreased or altered GnRH secretion at the level of the hypothalamus, decreased or altered LH or FSH secretion from the pituitary, and increased estradiol production at the level of the ovaries. In the studies where doses of test chemicals were high enough to elicit overt toxicity, HPA axis activation nearly certainly would have played a role in the results; even in experiments where overt toxicity did not occur, the likelihood that the chemicals used provided enough of a homeostatic challenge to either acutely or chronically alter HPA function is high. Because relatively large CVs are allowable in this assay in order to account for the effects of the estrous cycle on parameters like estrogen secretion and ovarian/uterine weight, effects of HPA activation on these and related measures have the potential to be masked (leading to false negatives), particularly when secondary histological analysis is not performed. For instance, chronic stress has been shown to increase a number of assay endpoints influenced by estrogen, including uterine weight (Guinn, 1996).

In order to take into account stress parameters in the female pubertal assay, it was suggested that the EPA consider monitoring HPA axis parameters other than simple overt toxicity and adrenal weight in order to rule out potential interplay between this axis and the HPG axis. Since measures of the stress hormones corticosterone or adrenocorticotropin hormone (ACTH) at sacrifice are unlikely to provide useful information regarding stress in the experimental animals over the course of the assay, these measures could be taken during the range-finding phase during which doses are decided. Alternatively, a non-invasive approach would be to measure corticosteroid levels in the fecal matter of test animals in order to identify whether a generalized stress response is a confounding factor in the responses elicited by various test chemicals (Cavigelli et al., 2005). (References for a test kit that can be used for this is: <http://www.enzolifesciences.com/ADI-900-097/corticosterone-elisa-kit/>).

In addition to suggesting that the EPA consider taking measures of HPA axis activation in the female pubertal assay in order to account for the potential confounding effects of stress on HPG axis function, ways to decrease variability in results due to the estrous cycle were also suggested. These effects could be accounted for experimentally while still assaying rats at the same age by synchronizing the estrous cycles of the rats using GnRH analogue injections in the days prior to the end of the assay. Doing this would allow the rats to be examined at the same age and on the same day of the cycle, plus effectively test the responsiveness of the HPG axis at the levels of the pituitary (GnRH challenge) and ovaries (LH and FSH challenge), assuming these parameters aren't disrupted by test chemicals. Given that the 42 – 43 day period is 'standard' but still rather arbitrary, this approach could potentially yield important information currently lost due to estrous cycle variations. Alternatively, since the parameters being investigated are reproductive, instead of using absolute age as the deciding factor for taking tissue samples, the EPA could use "reproductive age" (i.e., the day of first estrus of each animal) as the comparison point.

In specific comments regarding the male pubertal assay, one panel member stated that as occurred in the female pubertal assay, a fairly large number of test chemical CVs were out of range, but accepted as interpretable by the EPA. The panel member reiterated that, due to the nature of these studies, there was a chance that the effects of HPA axis activation on the parameters being tested could have contributed to the high CVs seen in the assays. It was noted that even low levels of chronic stress can significantly inhibit testosterone secretion, and that this was a potentially confounding factor because most of the measures taken are a direct reflection of the actions of this hormone. Further, such an effect would unlikely be seen in single measures of testosterone, as the secretion of this hormone fluctuates widely. As such, it was suggested that HPA axis activation needed to be ruled out as the cause for effects seen, and that to do this would require the measurement of endpoints related to this axis in a similar manner as suggested for the female pubertal assay. Finally, it was suggested that in order to more accurately test the integrity of the HPG axis and the responsiveness of Leydig cells in the testes to luteinizing hormone receptor stimulation and the synthesis and secretion of androgens, the EPA should consider using an human chorionic gonadotropin (hCG) challenge to stimulate testosterone. In addition to testing testicular responsiveness to LH receptor activation, an hCG challenge would give the Agency a potentially more stable measure of testosterone secretion, and given the short-term nature of the test, would be highly unlikely to influence parameters being measured other than testosterone secretion. A reference for the use of hCG to stimulate testosterone secretion can be found here: <http://www.ncbi.nlm.nih.gov/pubmed/14684600>.

In conclusion, the Panel generally agreed with the EPA that the data generated using the male and female pubertal assays are useful, but advocated for careful modification of the guidelines, and cautioned that significant care needs to be taken when interpreting results from the assays.

Charge Question 4: The test guidelines for the six in vivo assays (Hershberger assay - OCSPP 890.1400, OECD 441; Uterotrophic assay- OCSPP 890.1600, OECD 440; Male Pubertal assay- OCSPP 890.1500; Female Pubertal assay - OCSPP 890.1450; FSTRA - OCSPP 890.1350, OECD 229 and AMA - OCSPP 890.1100) offer some guidance on setting the dose/concentration range when testing for specific effects on the E, A, or T signaling pathways. In some of the in vivo assays, overt toxicity was noted based on effects on growth, other sublethal effects, and even mortality at the highest dose/concentration tested. Positive Tier 1 findings indicating the potential for endocrine activity can be difficult to interpret in the presence of overt toxicity.

Please comment on the nature and severity of overt toxicity that would render study results unreliable for the purpose of Tier 1 screening of a chemical for endocrine activity. Please comment on what, if any, additional guidance is needed to minimize overt toxicity while ensuring adequate dose selection for these screening assays. Also, specifically comment on the validity of including treatment concentrations with apparent body weight effects in the analysis for potential endocrine interaction for FSTRA?

Panel Response

The use of a range of doses when investigating the potential environmental endocrine disrupting effects of test compounds is obviously of great importance. However, the overt toxicity seen during testing complicates the interpretation of data from the in vivo assays employed by the EPA for the EDSP. Several difficulties are encountered when overtly toxic doses are used in the screening program. For instance, toxic compounds impact the basic cell machinery essential for endocrine function: energy utilization, metabolic precursors, genome stability, etc. Further, energy disruptors (mitochondrial targets) alter energy dynamics. This can lead to well characterized changes in signal transduction pathways (i.e. Protein Kinase B (AKT), Autophagy-related (ATGs), Mammalian target of rapamycin (mTOR), or c-Jun amino-terminal kinase (JNKs)) that directly impact hormone responses that are required for transcriptional activity and for cell progression through the cell cycle. Because of these and other physiological changes associated with overtly toxic doses, there is a high potential for confounding effects to be elicited by test substances, and these can render findings potentially unreliable. If effects are observed only in a dose group with overt toxicity and in the only screen that shows a positive effect, it will be difficult to decide whether the compound should undergo Tier 2 testing without a repeat experiment at a lower dose. As such, the Panel in general agreed that efforts to reduce the test compound doses to sub-toxic levels is required to allow the impact of the target agent on endocrine function to be evaluated independent of its general impact of organismal well-being.

Suggested criteria to define the Maximum-Tolerated-Doses (MTD)

In suggesting criteria for defining MTD in the EDSP assays, the Panel noted that the common standard of practice/animal welfare benchmark for short-term toxicity studies is a 10% upper limit of body weight loss. This cutoff is deemed to be sufficient to determine maximum-tolerated doses, and is a common standard accepted by most Institutional Research Animal Use and Care Committees (NC3R², 2013). However, in

discussing this parameter further panel members recommended the 10% weight loss benchmark be used as a general guideline for the mammalian assays, but not the FSTRA or AMA (see below). As an alternative, it was suggested the EPA better define MTD estimates derived from range finding studies by increasing the number and dose spacing in these bioassays. One way to do this would be to evaluate the 95% Confidence Limit for the MTD and if this variance is high (>20%), it would trigger additional range finding studies to better define the MTD. If these determinations are made for several compounds, it may be possible to compare the “better defined” MTD estimate with the original maximum MTD estimate statistics, such as the lower 95% Confidence Limit of that estimate. If these two values are close, it may be possible to simply use the lower 95% Confidence Limit in lieu of having to conduct additional range finding tests to define the MTD.

Suggestions regarding test dose selection and reducing overt toxicity:

The Panel made several general recommendations regarding dose selection for the in vivo assays in the EDSP. When overt toxicity is present it was suggested that the testing labs employ two additional lower doses below the original MTD used to generate the dataset, using concentrations that do not illicit overt toxicity. If effects on A, E or T endpoints are still observed in the absence of overt toxicity, then there is a higher degree of confidence that the effects on A, E, and T are directly related to the mode of action of the chemical as opposed to a more generalized stress response. If, on the other hand, effects on A, E and T are not observed in the lower dosing regimen, then the original dataset obtained at toxic MTD levels should not be considered to be definitive for direct effects via alterations in hormonal pathways. However, the effects of a generalized stress response altering A, E, and T should be noted in the event that more information about this effect of overt toxicity can be further used in future risk assessments.

An alternative method for reducing overt toxicity proposed by the Panel included the following steps: 1) reduce the initial highest test concentration; 2) add another concentration group (i.e., four test concentrations); 3) consider an approach similar to the *in vitro* binding assays in which solubility becomes a problem and run test concentrations in half-log unit steps as opposed to whole-log units with additional concentration groups; 4) change the criteria for evaluation such that concentration-related changes are not required to indicate potential endocrine-related effects (i.e., effects at a single concentration indicate an effect); or 5) accept the overt toxicity, loss/obscurity of information, and reduced power of the test for detecting of endocrine-related effects.

Finally, it was suggested by a panel member that the EPA use toxicology data that already exists from the previous testing of chemicals initially chosen for the EDSP. Briefly, concentrations of pesticides and other chemicals (xenobiotic may be a better choice for the generic “chemicals”) that are overtly toxic can already be subject to regulation based on non-endocrine disrupting criteria, so there is no value added when these compounds are retested for endocrine disrupting activities at those concentrations. It would, therefore, be important for the Agency to analyze how its Tier 1 testing can encompass dose/concentration ranges that allow for maintaining the necessary

physiological integrity of the systems, while providing the sensitivity needed for Tier 1 screening.

Mammal specific comments:

For the mammalian in vivo assays, overt toxicity will include the activation of the HPA axis, which in turn affects all endocrine systems, including those that are the focus of the EDSP. Many of the potentially confounding stress axis-related factors in the mammalian studies are discussed in the Panel's reply to Charge Question #3, which addresses the variability found in the data obtained in the Male and Female Pubertal Assays. These are also true with the Hershberger and Uterotrophic Assays, although with these two protocols, the potential confounding direct and indirect gonadal effects of stress-related parameters are not in play. Also in the Panel's response to Charge Question #3 are recommendations for testing markers of HPA stress-response axis activation such as corticosteroid production, which could be assayed in the range finding studies. In terms of dose selection for the mammalian assays, it was suggested that screening dosage should be chosen as the highest concentration of the test compound that does not alter body weight, or general physiological tone beyond expected changes. Alternatively, if doses at or near the MTD are used to set the high dose, the Panel suggested that at least 3 doses, or more, be used in the screening assay to increase the likelihood that a dose without overt toxicity is tested. In this regard, the MTD should be not relied upon in every case in dose setting. If exposure data were available, dose setting could be based on a margin of exposure approach. Other existing toxicity data, even if not using the same administration protocol, should also be considered. Alternatively, when overt toxicity is observed, it is essential to run effects studies at lower doses that do not elicit overt toxicity. The general approach of lowering the overall dose response curve to lower log doses or half-log doses around the mid-point of the exposure range and then excluding the higher, toxic doses seems like a reasonable compromise to address this issue. This approach avoids biasing the data analysis with a dose that may only have more resistant and less affected individuals. Finally, it was noted that initiation of oral gavage dosing is expected to cause a transient drop in total body weight. Thus, examination of historical data from each of the laboratories performing the assays will likely provide insight as to the extent of transient change in weight with dosing and thus, inform as to an anticipated weight loss unrelated to overt toxicity.

FSTRA-specific comments

In comments specific to the FSTRA, the Panel agreed with the Agency that the FSTRA test protocol is robust with study metrics (endpoints) at multiple levels of biological organization. This is a real strength of the test and greatly enhances the utility of FSTRA testing as an integral part of the EDSP Tier 1 Panel. However, the complexity of the protocol and the variety of expertise required to conduct and evaluate all of the endpoints renders this a challenging assay for most laboratories to conduct, while meeting all of the criteria. That said, evaluation of individual endpoints and their relevance toward the determination of the potential of a xenobiotic to interact with endocrine function is greatly enhanced in the FSTRA due to the fact that consideration of other test metrics can

be evaluated simultaneously. At several times in the EPA report it was correctly stated that the validity of a test was not lost when performance criteria for a single endpoint was not met. Consideration of the results within the context of other endpoints is required and is a major strength of the FSTRA.

Regarding dose selection for the FSTRA, it was noted that the greatest of the three test concentrations is one-third of the 96-hr LC50 for the compound in fathead minnow. The 96-hr LC50, by definition will cause 50% lethality to the test organism in 96-hr. Depending on the slope of the dose-response curve, a concentration that is one-third of the 96-hr LC50 would likely be associated with some degree of mortality within a four-day period. Thus, it seems likely that there would be some degree of lethality when fish were exposed to this same concentration for 21-days. Moreover, some other overt toxic effects of this concentration (one-third 96-hr LC50) during the course of a 21-day exposure that is more than five-times longer in duration is highly likely. Therefore, it is not surprising at all that survival was affected in one-third (7 of 21) of the assays, weight reductions were apparent in nearly half (10 of 21) of the FSTRA evaluated, and over one-third of the test chemicals resulted in sub-lethal clinical signs of toxicity. Overt toxicity, including reduced survival and body weight loss can have a confounding impact on the results of the FSTRA at the exposure concentrations that express the toxicity. It may not be possible to discern if changes in endocrine-related endpoints (e.g., GSI) are due to direct effects on endocrine functions or secondary effects of toxicity. There are certainly cases where a clear endocrine-related effect of a chemical could be evident, even in the presence of clinical signs of overt toxicity (e.g., ovo-testes, secondary sexual traits of the other sex, etc.); yet all too often, those exposure groups that exhibit an overt toxicity response will be lost to the final dataset. The end result is an assay with fewer test concentration data, and ultimately a test with reduced power of detection of endocrine-related effects. Results would be rendered unusable when: 1) high levels of actual mortality (> 10-25%) occur at a given exposure treatment; 2) overt toxicity effects commonly occurring throughout a large portion of the exposed individuals (> 50%) within a given treatment; or 3) there are multiple evident stress effects (altered behavior, reduced feeding, etc.) that will alter physiology and may affect hormonal pathways.

Specifically regarding body weight in the FSTRA, the Panel noted that these determinations are easy to measure and may be directly related to altered behavior and reduced feeding rates that may be contributory in causing overt toxicity. Body weight effects are directly measurable and are not subjective as opposed to observations of behavior that by their very nature are more subjective. The percentage of feeding response is also more directly quantified (presence (feeding)/absence (not feeding) criteria) and are not subjective. In the fish bioassay, only 1 compound of the 21 (< 5%) compounds tested resulted in significantly reduced length and weight determinations. Changes in body weight co-occurred with other sub-lethal effects in 5/21 compounds tested (23%) as significant differences in male or female body weight were observed at doses where other sub-lethal effects occurred. Changes in body weight also occurred in some compounds in the absence of other sub-lethal effect in 4 out of the 21 compounds tested (19%). Thus changes in body weight in the fish bioassay were not always found to co-exist with overt toxicity effects. It was also noted that changes in body weight of

females were problematic and would lead to exclusion of some data at higher doses and thus, reliance on data at 10 and 100 times lower than the higher dose. Females are more apt to see changes in body weight due to the energy diverted into egg production when compared to males. Despite these complications, tracking of body weight is still an important endpoint for test compliance guidelines as it is a generalized measure of fish health. Given that length and weight changes were noted in the FSTRA it may be useful to use a condition index approach (Bolger and Conolly, 1989) that integrates weight and length (e.g., $CI = \text{Weight}/\text{Length}^3$) as a more integrated measure of response changes in the presence of overt toxicity .

Finally regarding the FSTRA, it was noted that overt toxicity may affect the potential to discern endocrine disrupting chemical (EDC) effects. It is the most sensitive and most affected individuals that are highly stressed, thus compromising interpretation of EDC effects using the different bioassays in the Tier 1 Screening Battery. If there is actual mortality, this further complicates results, as test end points in the Tier 1 battery cannot be measured in what may be the most sensitive portion of the population of individuals exposed.

AMA-specific comments

Regarding the AMA, the Panel noted that 10% weight loss is not recommended as a parameter to set MTD, as tadpoles may grow considerably during the 21-day test period, and under certain conditions they can exhibit wide individual variability in development and growth rates. Growth and development during late premetamorphosis and prometamorphosis in *Xenopus laevis* are concurrent but not necessarily associated events. For example, tadpole development can be inhibited by a number of well-known goitrogens without significant effects on growth (Degitz et al., 2005); or growth can be accelerated by compounds such as triclosan without effects on development (Fort et al., 2007); or, for a more complicated example, tadpole growth and certain aspects of development such as hindlimb length can be inhibited by environmentally relevant concentrations of cadmium without detectable effects on endpoints of thyroid activity such as epithelial cell height (Sharma and Patiño, 2010). Although it is important to measure and differentiate effects on growth from effects on development, given the known variety in the specific responses of growth and development to different chemicals there is no *a priori* reason to believe that overt toxicity as defined by effects on growth would render the assay results unreliable for the purpose of Tier I screening. Nonetheless, the overarching goal of the EDSP and the Tier 1 battery is to identify chemicals/xenobiotics that impact the endocrine system in the absence of overt toxicity. As such, all efforts should be made to eliminate toxicity as a confounding response when screening of endocrine disruptors.

A relevant issue in regards to growth measurements in the AMA is the method for normalizing hindlimb length (HLL) for tadpoles of different size. The Agency recommends normalizing HLL based on snout-vent length (SVL); namely, by dividing HLL by SVL. This method, however, assumes that the relationship between the two variables is isometric. Generally speaking, an isometric scale relationship can be defined

as one where the slope of the regression line on a double-log plot is equal to 1. However, if the slope is different than one, higher or lower, this normalization method could lead to errors depending on the magnitude of the deviation from a slope of 1 and the size range of individuals in the population. Most size variables do not scale isometrically and it is likely that neither do HLL and SVL or HLL and any other size scale such as body mass or whole body length. An alternative method to correct for tadpole size that would eliminate the potential problem of allometric scaling would be to use an appropriate regression method for HLL and SVL and calculate the residuals, and to use these residuals for statistical analyses instead of raw values or ratios. The Agency was advised to consult with a biometrician or morphometrician to find appropriate ways to address this issue.

Charge Question 5. Spinal curvature, usually manifesting as “bent tail” in *X. laevis* tadpoles, was reported in 15 of 18 AMA studies reviewed thus far. The anomaly appears to be first observed several days after study initiation, and prevalence increases with time. Overall, the prevalence of spinal curvature in these studies ranged from “a few per replicate” to 92% of a given treatment group by test termination. Experimental work by the EPA Office of Research and Development suggests that overfeeding can be a primary cause of spinal curvature in their *Xenopus* test populations; however, spinal curvature remained prevalent (range: 16-92%) in the five industry AMA studies in which feed was reduced by 50% compared to guideline recommendations. Overall, the incidence of spinal curvature appears to be highly variable. From a qualitative review of the data, there appear to be no consistent differences in the incidence or variability of spinal curvature when studies using guideline versus reduced feeding regimes are compared. *Please comment on whether the presence or prevalence of spinal curvature in test specimens, including controls, compromises the utility or validity of an AMA submission. If so, when does the prevalence of spinal curvature render the study unreliable? What technical guidance may be useful for laboratories in reducing the occurrence of spinal curvature and determining if, or at what point within the study, a study may be compromised because of this phenomenon?*

Panel Response

Summary Recommendations for Technical Guidance:

The Panel recognized that the AMA is a screening assay in the EDSP Tier 1 battery intended to empirically identify substances that may interfere with the normal function of the HPT axis. The AMA is a general vertebrate model to the extent that it is based on the conserved structures and functions of the HPT axis. As such it is important as an *ex vivo* model for thyroid-dependent process which respond via the HPT axis.

The Panel further recognized the AMA is related to altered hypothalamic-pituitary function, anti-thyroid activity and thyromimetic activity. In the case of altered hypothalamic-pituitary function, the AMA is complimentary to the rat pubertal assays and the fish reproductive screen. In the case of altered anti-thyroid activity, the AMA is complimentary to the rat pubertal assays. In the case of thyromimetic activity, the AMA has no complimentary assays within the Tier 1 battery.

The general experimental design was described by Fort et al. (2007). Briefly, Nieuwkoop-Faber (NF) stage 51 African clawed frog (*Xenopus laevis*) tadpoles, are exposed in separate treatment groups, to a minimum of three concentrations of a test chemical or a negative control (clean water) for 21 days. There are four replicate tanks within each test treatment. Larval density at test initiation is 20 tadpoles per test tank (replicate) for all treatment groups (80 larvae per treatment). The observational endpoints include: 1) hind limb length (HLL); 2) snout-to-vent length (SVL); 3) NF developmental stage; 4) body weight; 5) thyroid histopathology, and 6) daily observations of mortality and clinical signs.

As noted in the White Paper the occurrence of “bent tail” or “spinal curvature” in developing frogs during the AMA test appears to be very common place in most of the studies reported; specifically 83% (15 of the 18) compounds tested reported this effect and the other studies did not indicate whether bent tail was observed in those tests. The bent tail syndrome was present in both controls and treatment in almost all tests and was unrelated to the mode-of-toxic-action of the chemical tested.

As further noted in the White paper, within the AMA Test Guidelines, attempts have been made to reduce feeding in order to control development within the frogs to produce appropriate life stages at the end of the test. EPA has investigated this reduction in feeding and in particular the reduction in iodide within the diet as a result, which may in part play a role in the occurrence of bent tail. Results reported in Table 32, (see pages 120-121 in the White Paper) indicate three sources of food were used including Sera Micron (Compounds B, C, E, F, H, I, J, N, O, P, S, U, and V), *Xenopus* Express (Compounds A, G, K, and Q), and Nasco Frog Brittle (Compound M) and all food sources were used in studies with high levels of reported bent tail. Levels of iodide were only reported for Sera Micron (range 40-54 ug/g) and Nasco Frog Brittle (48 ug/g). The levels of iodide in these two feed stocks are very similar. Water levels of iodide reported were also quite similar in most tests ranging from 3.6 - < 50 ug/L for all studies and ranging from 3.6 - < 10 ug/L in 16/18 studies. Therefore, differences in iodide in food and water do not seem to explain this syndrome *per se*. An examination of the data presented in the White Paper suggests that the response seems to be somewhat laboratory dependent suggesting the importance of husbandry conditions within different laboratories. Thus, the Panel specifically recommended that it will be important to:

1. Keep track of clutch performance (number and quality of embryos from each spawn) and select high quality, healthy breeding pairs. Eliminating breeding pairs that produce bent tail should minimize its occurrence. Minimizing the use of wild-caught breeding stock may be beneficial.
2. Closely review the water quality conditions under which the bent tail tad poles were kept. Temperature, pH and nutritional deficiencies (vitamin C and Ca) are reported to cause bent tail in laboratory reared American bullfrogs and the Leopard frog (Marshall et al, 1990; Leibovitz et al, 1992; Martinez et al, 1992), and in several other amphibian species. Temperature and pH have been associated with the condition in *Xenopus tropicalis* and water softness (Ca and K) associated

with the condition in *Xenopus laevis*. Over-feeding can alter the water biochemistry and uneaten food and excess waste causes changes in ammonia and pH that are detrimental to the health of the animals and could potentially explain the bent tail seen in the laboratories which reported decreased occurrence of the condition with a decrease in feed.

3. Feed and water sources should be screened for herbicides, biological toxins and pesticides currently known to cause bent tail in *Xenopus laevis*, nominally. Biological toxins would include both potential Harmful Algal Bloom (HAB) and microbial toxins.
4. The Panel recommended further monitoring the portion of each clutch not used in a given study. It would be potentially useful if these results were reported. The Panel understood that some information on this concept is in the guidance document, but it needs to be elaborated further.
5. The Panel further recommended that for acceptance of the AMA test results overall, the clutch should not have a rate of mortality and morbidity > 20%.

The Panel understood that the Fort protocol represented the state of the science in 2008 when the initial Tier 1 testing order was issued. The Panel agreed with the Agency in not altering the protocol until the first round of testing was completed. While, this SAP is based on only 21 of the 52 chemicals being tested in this order, the Panel believed there is sufficient evidence with the AMA that a reevaluation of the test protocol may be in order.

The Panel expressed the opinion that without an evaluation of WoE, the AMA must currently be considered important to the Tier 1 screening battery, but the prevalence and variability in percent of clutch expressing this tail malformation does affect confidence in the utility of the assay. However, the Panel recommended not changing the experimental design of the actual study outcomes until all 52 chemicals are evaluated.

Since SVL is a key measurement in the AMA, the Panel felt the “bent tail” issue has the potential to impact both the performance criteria for acceptance of individual test results and the scientific confidence in the utility and validity of the AMA. Therefore, the Panel recommended that the Agency determine the cause and possibly mode of action for the bent tail syndrome so guidance can be given on how to eliminate or at least reduce this morbidity response to < 10% of the populations.

The Panel understood that the assay typically begins with stage 51 larvae that do not show tail curvature. However, curvature is typically not obvious until the larvae are further into prometamorphosis (i.e., the window of susceptibility in the AMA is stage 51-55). Therefore, laboratories cannot simply select the effect out of the test. However, the Panel thought that there may be signs in the stage 51 embryo that can be used to do a better job of selecting against this response.

Evidence presented in the White Paper and in the oral presentations at the SAP does not allow for a definitive understanding of the bent tail syndrome (i.e., there is no histopathology or biochemical data presented for the syndrome). The Panel understood that from a developmental point of view, the bent tail syndrome does not seem to have an effect on the actual resorption of the tail, in most cases. An examination of the AMA data suggests that the occurrence of bent tail seems to be laboratory dependent. The Panel believed the potential causes for bent tail are: 1) genetics (as described above), 2) nutrient deficiency, 3) poor egg quality, and/or 4) poor water quality (possibly due to overfeeding). There was a high probability that one or more of these factors may all have a role in the development of bent tail. The Panel believed that a pathological examination of the tadpoles may reveal information that may be helpful.

The Panel benefited from having members that have extensive experience with *Xenopus* husbandry. There is antidotal information that leads to suggesting the syndrome is related to the ionic imbalance, poor water quality or a combination of factors that are affected by both the composition of the water and the feeding scheme.

One Panel member has extensive experience in performing the Frog Embryo Teratogenesis assay: *Xenopus* (FETAX). While the assay is different in its protocol from the AMA, the test species is the same. Unpublished work by this panel member found that softer water (water with less Ca and K) caused a greater incidence of lateral tail curvature and other general “toxic” effects such as edema and abnormal gut coiling. Moreover, experience by some panel members has shown that by keeping track of clutch performance (number and quality of embryos from each spawn) high quality breeding pairs could be identified. Breeding these paired frogs on a regular schedule, regardless of the need for embryos, improved clutch quality.

The Panel noted that in fish, nutritional deficiencies (micronutrients, such as vitamin C) can lead to developmental curvatures of the spine. Also, excess of certain nutrients (e.g., selenium and copper) can also cause this lesion. Lastly, low dissolved oxygen can lead to spinal curvatures in fish embryos.

While not directly asked in Charge Question 5, the Panel strongly noted that the issues of bent tail (Charge Question 5), and thyroid histology (Charge Question 6) and their interactions with other AMA endpoints (i.e., HHL, SVL, final NF stage and body weight) leads to a recommendation that the Agency look into changing the experimental design of the AMA study. It is further recommended that the Agency specifically look into developing a stage dependent rather than a time dependent protocol.

Since neither the White Paper nor the Agency’s oral presentation described the histopathology of the “bent tail” lesion or presented photomicrographs of the condition, the Panel found it difficult to provide meaningful guidance. Those members of the Panel with experience with *Xenopus* testing and issues associated with tail malformations generally realized that lateral tail flexion is typically considered a non-specific effect, which is more of a muscular phenomena than a connective tissue or skeletal phenomena.

In contrast, if curvature of the spine in the AMA test is a dorsal to ventral kinking which occurs primarily in the anterior region of the tail, the Panel was concerned that the lesion is in the notochord and related to connective tissue elements. The latter anatomical pathological description is very similar to that noted for an osteolathyrogenic response where the alteration in elastin and collagen formation leads to a weak notochord which kinks or bulges in response to swimming movement (Schultz et al., 1985).

Schultz et al. (1985) examined the osteolathyrogenic effects of semicarbazide using the FETAX. Osteolathyrisms is a connective tissue lesion associated with decreased intermolecular bonding in collagen and elastin. It is manifested in early embryos as gross anatomical changes in the long axis of the animal and kinking of the tail. Histopathology examination, including electron microscopy, reveals that gross effects were produced by changes in connective tissue fibers of the notochordal sheath, specifically a marked reduction in elastic fibers and a concomitant disorganization of collagen fibers.

Lysyl oxidase (*Lox*) is a copper-dependent amine oxidase that catalyzes the cross-linking of collagen and elastin fibers. Geach and Dale. (2005) characterized *Xenopus laevis* cDNAs for *Lox*, *Loxl-1* and *Loxl-3*, and showed that they are expressed during early embryonic development. Using RT-PCR they detected maternal transcripts for *Xloxl-1*, but levels remained low until tailbud stages. Transcripts for *Xlox* and *Xloxl-3* were not detected until early neurulae, although transcripts for *Xlox* remained at low levels until tailbud stages. Whole mount *in situ* hybridization showed that transcripts for *Xloxl-1* and *Xloxl-3* are localized in the notochord, while transcripts for *Xlox* are found in the notochord, somites, and head. *X. laevis* *Lox*-like enzymes were inhibited by incubating embryos, from cleavage stages to tadpole stages, in β -aminopropionitrile, a specific inhibitor of the catalytic domain. The resulting embryos appeared to differentiate normally but suffered from poor collagen fiber formation. Defects included kinks in the notochord. These results suggest that *Lox*-related enzymes are required for the proper formation of the collagen and elastin fibers during *X. laevis* development.

The Panel suggested that other factors to consider in examining the bent tail syndrome may be to include examining other ingredients in food, such as anti-oxidants. For example, Ellenberg (2000) found that TetraMin fish food contained high levels of the antioxidant ethoxyquin which increased the production of P Glycoproteins (PGP) and Multi-Drug Resistant (MDR) Proteins in estuarine fish, *Fundulus heteroclitus*. This effect was very pronounced as fish from pristine habitats had elevations in PGP and MDR to levels comparable to fish which were collected from highly polluted sites (e.g., EPA Superfund Sites). Elevated PGP and MDR in the fish has an effect on overall energy metabolism, diverting energy away from growth and development, which might render fish more susceptible to disease. Also this increase in PGP and MDR will also affect uptake/depuration rate kinetics for some compounds and may result in possible underestimation of toxicity since these proteins are generally associated with reduced drug uptake. In the case of the frogs and bent tail, examining other contents of the food may be warranted to systematically eliminate these as a contributing factor in "bent tail". Of particular concern are inert food ingredients that may increase the costs of

maintenance metabolism and possibly increase susceptibility to disease, if bent tail is found to be associated with a disease process as more is learned about it.

The Panel noted that frogs are not the only toxicity test species to have altered development or abnormalities. For example, the Eastern Oyster, *Crassostrea virginica*, which has been extensively used in toxicity testing, may commonly be infected with the protistan parasite *Perkinsus marinus*, which may produce annual mortalities in excess of 40% in adult oysters. Aquatic toxicologist simply avoided periods of high temperature and salinity during the summer months (July-September) and confined toxicity testing to the other times of the year (October-June) so that high control mortality rates could be avoided. Similarly, crustaceans such as grass shrimp, used extensively in marine testing of pesticides, may have diseases (e.g., black spot) which are parasitic in origin that have been well studied and have not compromised toxicity test results. In other words, there have been instances with other toxicity testing where confounding factors have been addressed and tests have been adapted accordingly, so as to minimize impacts of the confounding factor.

As noted above one factor in the bent tail syndrome may be the source or suppliers of frogs as there appeared to be some differences in the rates of this syndrome observed in comparisons of the different supplier. One of the panel members mentioned the need to possibly conduct some molecular sequencing to discern if there is an underlying molecular basis for this defect, which could enhance our knowledge of this problem.

Given that the tail reabsorption as the frog matures is a critical factor in this assay, occurrence of bent tail may affect overall health, growth, development, reproduction and survival over the life time of the organism. Thus, the phenotype may not be a highly confounding factor. If it does affect any of these variables, then the significance of bent tail will have to be further examined. In terms of its affect on E, A, and T outcome measurements within this portion of the life cycle of the frog is the most important thing for the Agency to focus on. Resolving this along with determining the overall health effects of this syndrome will be important.

This is a vertebral/muscular deformity, which has been reported with certain pesticides, which raises the significance of concern regarding this syndrome. The fact that this occurs in all control and exposure treatments clearly indicates this is not mode-of-action related to compounds being tested; however, if a compound were to illicit this effect, with a high rate of effects in controls, the ability to discriminate this effect will be reduced and possibly obscured. This is a concern. It may be prudent to test a known pesticide such as endosulfan, or another chemical known to illicit tail deformities in fish, and examine whether the assay can discern this effect in the AMA, assuming it would exert this effect in frogs. This approach would at least allow analysis of the degree of impediment this syndrome may cause and help possibly develop a positive control step for addressing bent tail. This would be an aid in discerning when a test is unreliable.

Charge Question 6. With the exception of thyroid gross pathology findings (thyroid gland atrophy and hypertrophy) in the AMA, severity grades are generally assigned based on comparison to “normal” *X. laevis* thyroid findings depicted in the guidance or based on the professional opinion of the pathologist conducting the assessment; they are not assigned in comparison to concurrent control findings from a given study. (Please refer to Section III.2.f in the document entitled “Interpreting Amphibian Thyroid Histopathology Diagnoses” and supporting documents, OECD Guidance Document on Amphibian Thyroid Histology No. 82, 2007 and Grim et al., 2009).

- a. *In one study, the pathologist’s report identified a lower incidence and severity of follicular cell hypertrophy when compared to the incidence and severity of this trait in control specimens. Similar trends have been observed in other studies. In this case, the pathologist concluded that the finding was potentially consistent with treatment-related delay of metamorphosis because thyroid follicular cells normally increase in height during tadpole development. Please comment on the validity of this conclusion.*

Panel Response

The AMA test guideline states that the assay should begin with Nieuwkoop-Faber (NF) stage 51 African clawed frog (*Xenopus laevis*) tadpoles and that the median NF stage in control tadpoles at test termination (21 days) should be ≥ 57 with a difference between the 10th and 90th percentile of ≤ 4 NF stages. It is therefore important to avoid or minimize the occurrence of NF stage ≥ 60 in control tadpoles at test termination.

An assessment of HPT axis activity at test termination is at the core of the AMA test guideline. Because thyrocyte height is directly dependent on TSH stimulation and is relatively easy to measure, an evaluation of thyrocyte appearance is a primary component of the test guideline. The test guideline calls for the use of a standard scheme to grade thyroid “lesion severity” that is based on normal changes that occur during metamorphosis in *X. laevis* (Grim et al. 2009). The lack of quantitative endpoints of thyroid activity, however, was one of the concerns expressed in some of the written statements provided by public commenters. As explained in more detail in the paragraphs that follow, this concern was shared by the Panel.

Although thyroid hormone agonistic activity could explain the pathologist’s observation of a lower incidence and severity of thyrocyte hypertrophy via negative feedback inhibition of the HPT axis. However, accelerated development in the experimental population would also have resulted in similar histological findings. Because accelerated development appears not to have occurred, the Panel agreed with the pathologist’s conclusion that the finding is consistent with a treatment-related retardation of metamorphosis via disruption of the HPT axis. This case may represent an unusual disruption of the axis that is not based on classical anti-thyroid mechanisms such as those observed after exposure to thyroid synthesis inhibitors (goitrogens), but on a failure of the axis to activate normally or fully. A similar finding by the laboratory of a panel member

was reported in *Xenopus* tadpoles exposed to a high concentration of cadmium where metamorphosis and development were suppressed along with the apparent failure of the HPT axis to activate (Sharma and Patiño 2008). Possible interpretations of these findings (pathologist's finding; Sharma and Patiño 2008) include a reduced ability of thyrocytes to respond to TSH stimulation (thyroid lesion) and/or a reduced ability of the pituitary to produce TSH (hypothalamus/pituitary lesion). A better understanding of the pathologist's finding would require not only consideration of the actual effects on tadpole development recorded during the study (HLL or NF stage), but also clarification of the mechanisms responsible for the failure of the HPT axis to activate properly, if in fact this is what happened. It may be appropriate to consider these questions in more detail during the assessment of Tier 2 assay performances.

- b. *What guidance may be given to better distinguish between histological changes in the thyroid associated with the normal progression of metamorphosis and treatment-related effects? Are there certain lesions or diagnoses which may, by their absence or lessened severity as compared to controls, be indicative of treatment-related HPT effects such as delayed metamorphosis?*

Because thyroid activity normally increases during metamorphosis, the assessment of treatment effects is relatively difficult to perform as the control condition is, literally, a moving target. To partially compensate for this situation, the AMA test guideline recommends the use of stage-matched individuals for histological analyses at study termination. Specifically, the guideline states that 5 tadpoles per treatment replicate (20 total) should be matched to the median stage of pooled controls tadpoles, whenever possible. However, comments provided by the Endocrine Policy Forum indicate that studies conducted in response to the test orders reported variation in the median stage of development of control tadpoles at test termination and raised concern about the inconsistency among studies with respect to the developmental stage at which treatment effects are being evaluated. It would appear that a marker of normal thyroid activity that does not change during metamorphosis would minimize concerns associated with Charge Question 6 and, therefore, be preferable over changing reference markers. Although the Panel was unaware of any standard histological feature of the thyroid that does not normally change during tadpole development, there is one immunohistochemical feature of the *Xenopus* thyroid that may be independent of development.

A study of *X. laevis* reported that T4-immunoreactivity is concentrated in a ring at the periphery of the colloid and that the intensity of this ring remains fairly constant during metamorphosis (Hu et al. 2006). Moreover, changes in the intensity of this ring following exposure to a goitrogen (perchlorate) showed better sensitivity to detect anti-thyroid activity than changes in standard histological features such as thyrocyte height. The T4-immunoreactive ring has not been validated for standard testing and its value for situations like the one presented in Charge Question 6, which may represent failure of the HPT axis to activate, is also uncertain. Nevertheless, the availability of development-independent markers of thyroid activity would offer clear advantages over the current scheme. The Panel noted that there are additional assays that may also be useful, such as

iodide uptake assays. The Panel encouraged the Agency to consider the possibility of future refinements to the AMA test guideline by incorporating some of these assays if they can be validated and confirmed to be development-independent. The Panel recommended that the T4-immunoreactive ring marker, if properly validated, could potentially also benefit thyroid assessments in the rat pubertal assays.

The Panel offered an additional comment of relevance to HPT activity assessments in the AMA. Although the current AMA test guideline relies on a semiquantitative grading scheme to assess thyroid effects, rank variables typically do not yield the same statistical power that continuous variables do. A second limitation of grading or ranking schemes is that they are subject to technician biases and therefore may yield inconsistent results among studies as well as among laboratories. Therefore, the Agency was advised to consider the use of quantitative thyroid activity measurements to further standardize the AMA test guideline. There are a number of computer-assisted image analysis procedures that could be used for quantitative measurements of thyrocyte height, T4-immunoreactive ring intensity, and possibly other histological features in a fairly automated manner. Software packages for image analyses are publicly available at no cost to users.

Charge Question 7. In 2008, the SAP acknowledged that the *in vivo* assays included in the Tier 1 battery provide both redundancy and complementarity for evaluating interactions with the E, A, or T signaling pathways. The panel also noted that all of the Tier 1 assays and the broad range of endpoints appeared to be necessary to “*discriminate positive and negative results*”.

- a. *Please comment on the battery performance with respect to the anticipated complementary nature of the more complex, multi-parameter in vivo assays in the context of the observed responses with the case studies. Please comment separately on the E-, A-, and T-related assays and endpoints.*

Panel Response

General Comments

A total of 18 chemicals for E, 17 chemicals for A and 16 chemicals for T were evaluated in this battery of assays. For the E and A pathways, there were apparent complementary responses with two or more endpoints responding in the pubertal assays and in the FSTRA. In some case studies, the male and female pubertal assays and/or the FSTRA identified endpoint responses along with endpoint responses in the relevant *in vitro* assays for both the E and A pathways. As anticipated from the battery design, there was less complementarity evident for the T pathway effects.

One panel member noted that to increase the chances that complementary results are provided by the male and female pubertal assays, the interplay between the HPA, HPT, and HPG axes need to be taken into account. In particular, accounting for HPA activation, for instance by assaying for corticosteroids in rat feces, could be an important

addition to the batteries. While this question focused on the *in vivo* assays, it was also pointed out that the ER and AR binding assays had many or all of the attendant problems associated with *in vivo* assays in that they use rat tissues to provide receptors. It was suggested that more complementary results would likely be provided by alternative assays using recombinant receptors; this suggestion was made repeatedly during the meeting. It was noted that there were examples in which bioassay concordance was not observed between the *in vivo* and *in vitro* tests. Overall, findings suggest the importance of multi-parameter and multi-taxa among *in vivo* assays as a component of the battery to evaluate E, A, and T alterations.

There was a general feeling expressed among the Panel that since the present exercise does not provide information on the magnitude or the direction of the effects, it is difficult to say how useful the observed complementary effects will be in the WoE analysis to strengthen or weaken the case for the E, A, or T activity of these test compounds and determine the need for further testing. The fact that multiple endpoints were reported to respond is encouraging.

Estrogen Pathway

The current Tier 1 battery contains 3 *in vivo* assays that are designed to detect estrogenic activity. The endpoints and reported responses measured in each were as follows: (1) Uterotrophic (U) - The uterotrophic assay includes only one required endpoint, uterine weight, but is a well-established and reliable assay for the detection of estrogenic activity. There were no significant effects reported for this assay (0/17). Since there is only one end point, no complementary analysis could be conducted. (2) Female Pubertal Rat (fPR) - The Female pubertal assay includes vaginal opening, time to first estrus, reproductive organ weights, and histology of reproductive organs as complementary endpoints. The guideline also includes monitoring the estrous cycle, but this was not discussed and the contribution of this endpoint to the Female pubertal assay is not clear. Significant effects in at least one or more of these end points were observed for 7/17 chemicals tested (42%). Three tested chemicals gave a response for only one endpoint (3/17 = 17.6%). Complementary responses were observed as follows: two responsive end points (4/17 chemicals tested = 23.5%), three responsive endpoints (2/17 chemicals tested = 11.8%), and four responsive endpoints (0/17 chemicals tested = 0%). (3) Fish Short Term Reproduction Assay (FSTRA) - The FSTRA assay potentially has seven complementary endpoints (fecundity, fertilization, secondary sex characteristics, gonadosomatic index, gonad histopathology, plasma concentrations of vitellogenin in both sexes, and an optional endpoint of plasma sex hormones). The optional endpoint of plasma concentrations of sex steroids was stated by EPA to be useful and potentially should be added as a requirement. Significant effects in at least one or more of these end points were observed for 16/17 chemicals tested (95%). One tested chemical gave a response for only one endpoint (1/17 chemicals tested = 5.9%). Complementary responses among the seven different test endpoints were observed as follows: two responsive end points (3/17 chemicals tested = 17.6%), three responsive endpoints (7/17 chemicals tested = 41.2%); four responsive endpoints (3/17 chemicals tested = 17.6%); five responsive endpoints (3/17 chemicals tested = 17.6%); six responsive endpoints (2/17 chemicals tested =

11.8%); and seven endpoints (0/17 chemicals tested = 0%). Overt toxicity confounded the interpretation of results for some (or many) of the chemicals tested and to avoid this serious problem, potential remedies in the dose selection procedure were discussed in responses to earlier Charge Questions.

The 2008 SAP indicated that the Tier 1 assays have good coverage for the detection of estrogenic responses mediated through Estrogen Receptor 1 (ESR1), and relatively less ability to detect anti-estrogenic effects or effects that might be mediated through other potential mechanisms Estrogen Receptor 2 (ESR2, membrane receptors, etc.). This situation remains the same. The 2008 SAP encouraged pursuit of assays for such potential mechanisms, and pursuit of the modification of the uterotrophic assay to improve the ability to detect antiestrogens. It is unclear at this time whether such efforts are warranted, although at least one panel member believed that the extension of the uterotrophic assay for detection of antiestrogenic activity would have added value.

Androgen Pathway

Three *in vivo* assays in the battery including the Hershberger assay, the Male pubertal assay and the FSTRA are able to detect effects on the androgen pathway. The endpoints and reported responses measured in each were as follows:

(1) Hershberger - The Hershberger assay measures changes (increase or decrease) in weights of androgen-dependent tissues and is a well-established and reliable assay. Complementary responses among different androgen-dependent tissues were observed for 7/17 chemicals tested (41.2%). (2) Male pubertal assay (MPR) - The male pubertal assay evaluates sexual development (prepubertal separation), androgen-dependent organ weights and histology, and single point (sacrifice) serum testosterone. Significant effects in at least one or more of these end points were observed for 10/17 chemicals tested (58.8%). Four of the 17 chemicals tested (23.5%) gave responses in only one endpoint. Complementarity responses among the four endpoints were observed as follows: two responsive end points (4/17 chemicals tested = 23.5%), three responsive endpoints (2/17 chemicals tested = 11.8%), and four responsive endpoints (0/17 chemicals tested = 0%). (3) Fish Short Term Reproduction Assay (FSTRA) – The FSTRA contains seven potential androgen-sensitive endpoints (fecundity, fertilization success, secondary sex characteristics, gonad-somatic index, histopathology of androgen-dependent reproductive tissues, plasma vitellogenin, and the optional measurement of plasma testosterone). Significant effects in at least one or more of these end points were observed for 15 of the 17 chemicals tested (88.2%). One of the chemicals gave a response for only one endpoint (1/17 = 5.9%). Complementary responses among the seven different test endpoints were observed as follows: two responsive end points (4/17 chemicals tested = 23.5%); three responsive endpoints (8/17 chemicals tested = 47%); four responsive endpoints (1/17 chemicals tested = 5.9%); five responsive endpoints (4/17 chemicals tested = 23.5%); six responsive endpoints (1/17 chemicals tested = 5.9%); and seven responsive endpoints (0/17 chemicals tested = 0%). As was the case for estrogen pathway endpoints, overt toxicity appeared to confound interpretation of the results for some (many) chemicals tested.

Thyroid Hormone Pathway

Three *in vivo* bioassays included endpoints to evaluate the thyroid hormone pathway (Male and Female pubertal assays and the AMA). Of the 16 chemicals evaluated, there were significant effects on T noted in 7 of 16 (43.8%) chemicals tested in the male pubertal assay, 9 of 19 (56.3%) chemicals in the Female pubertal assay and 7 of 16 chemicals (43.8%) in the AMA. The Female and Male puberty assays include measurements of T4, TSH, and thyroid gland weight, and thyroid gland histopathology as endpoints to evaluate the HPT axis. A variety of chemicals affected one or more of these endpoints with no consistent pattern and with no endpoint clearly being the most affected. In some cases these were sex-specific. The AMA assay contains evaluations of tadpole development (NF stage and normalized hind limb length) and thyroid histology. EPA has provided a clear decision logic for evaluating potential thyroid axis activity. In 4 cases of the positive AMA assays, both thyroid endpoints were altered while in 3 cases there was only one altered. Overt toxicity may confound the interpretation of the effects in several cases.

- b. *Please comment on the battery performance with respect to the anticipated redundancy across the 11 assays in the context of the observed responses with the case studies. Please comment separately on the E-, A-, and T-related assays and endpoints.*

Panel Response

The Panel noted, given the manner in which the data are presented, only limited statements on battery performance can be provided. Without knowledge of the magnitude or direction of the change, it does not seem possible to thoroughly judge how the battery has performed.

Estrogen Pathway

There were no clear ER binders and no test chemicals active in the uterotrophic assay. On the other hand, 4 of the 21 test chemicals gave positive results in the ERTA. It seems somewhat surprising that these compounds did not show activity in one or both of those assays. One panel member noted that the variability in the ER transcriptional activation assay evident in one of the public commenters' presentations suggested that this may in part explain the lack of concordance between the ER binding and ER transcriptional activation assays presented in the report. As shown in the Agency presentations, there were no compounds that showed only *in vitro* effects, although in 9 of the 10 cases in which both *in vitro* and *in vivo* effects were seen, there was a positive in only one *in vivo* assay, with the FSTRA being the positive in 8 of those 9 cases. There were some potential issues of overt toxicity in several of the positive *in vivo* studies, so that it appeared that there was only one compound (Q) that showed clear positives in two *in vivo* assays. In short, there is redundancy across assays for the estrogen pathway endpoints, but it remains to be determined how clear the interpretation of these

redundancies will be. There seem to be a high proportion of equivocal results in the aromatase assay and it is not clear how important that assay will be in the overall evaluations.

The results indicated that at least one of the three E bioassays (Uterotrophic, Female pubertal and FSTRA) detected an effect in 17/17 chemicals tested (100%). As noted above, positive results were not observed in the Uterotrophic assays, but only in the Female pubertal and FSTRA tests. Thus, there were no positive results (0%) in all 3 bioassays for the 17 compounds tested. Positive results in two of the three bioassays were observed for 5 of the 17 (29.4%) chemicals tested in the Female pubertal assay and FSTRA. Overt toxicity may confound the interpretation of some of these results. In terms of redundancy between *in vitro* and *in vivo* E tests, highest rates of redundancy were observed for the steroidogenesis assay and the FSTRA, as concordance was observed for 7/17 (41.2%) of the chemicals tested. Redundancy was also observed between the *in vitro* aromatase assay and FSTRA, as concordance was observed for 5/17 (29.4%) of the chemicals tested.

Androgen Pathway

The data presented clearly indicates that there is redundancy across the *in vitro* and *in vivo* assays designed to detect effects on the androgen pathway in the sense that 13 of the 17 compounds presented as having been tested in the six androgen pathway assays appear to show effects in multiple assays. As with the estrogen pathways discussed above, the value of these redundancies will be better determined after the WoE evaluation.

Results indicated that at least one of the three A bioassays (Hershberger, male pubertal, and FSTRA) detected an effect in 17/17 chemicals tested (100%). Positive results in all 3 bioassays were observed in 3/17 chemicals tested (17.6%). Positive results in two of the three bioassays were observed in 1/17 (5.9%) of the chemicals tested for Hershberger and male pubertal assays; 3/17 (17.6 %) of the chemicals for Hershberger and FSTRA; and 5/17 (29.4%) of the chemicals for Male pubertal and FSTRA tests. Redundancy was observed in at least 2 or more of the 3 assays for 12/17 (70.6%) chemicals tested, with Male pubertal and FSTRA accounting for 8/12 (67%) of the redundancy, Hershberger and FSTRA accounting for 3/12 (25%) of the redundancy, and Hershberger and male pubertal accounting for 1/12 (8%) of the redundancy. Overt toxicity may have confounded some of these results.

Thyroid Hormone Pathway

As indicated in the 2008 SAP report, in the current White Paper and the EPA presentations, there is less redundancy in the assays for effects on the thyroid axis. The 2008 SAP suggested exploring additional endpoints, such as thyroid hormone-sensitive gene expression and transcriptional activation assays, but it is not clear if additional endpoints to strengthen the evaluation of this pathway are being considered. Three of the 16 chemicals tested in this battery showed some thyroid effect in all three thyroid axis assays, three showed effects in both rat assays but not the frog assay, and three showed effects in the frog but not the rat. Four chemicals showed effects in only one sex in the

rat assays. The Panel agreed with the EPA conclusion that all three assays are necessary to evaluate this axis. Dose selection and overt toxicity were issues associated with performance of these *in vivo* T pathway bioassays.

There were responses in all three *in vivo* T assays for 3 of 16 (19%) chemicals (Chemicals C, F, and G) tested. For the AMA, thyroid histopathology and developmental alterations were both affected in 2 of the 3 cases. In the male and female pubertal assays, T4 was affected in both sexes for all 3 compounds. Thyroid histology was affected in both the female rat and AMA for one of the 3 compounds (G).

There were responses in only the male and female pubertal assays, but not in the AMA, for 3 of 16 (19%) tested chemicals (Chemicals E, K, and P). Thyroid histopathology, TSH, and T4 were responsive for these 3 compounds, but with different patterns based on chemical and sex. There were responses in only the AMA for 3 of 16 (19%) chemicals (Chemicals H, M, and V) tested, with both thyroid histology and development affected in 2 cases and development alone affected in the third case. Potential solvent interference in the tadpole development assessment was noted in 2 of the cases. There were responses in only the AMA and female pubertal assays for 1 of the 16 (6.4%) tested chemicals (Chemical B), with thyroid histology affected in both assays. There were responses in only one bioassay for 3 of the 16 (19%) chemicals tested (Chemicals A, J and S). There were negative responses in all three T pathway bioassays for 3 of 16 (19%) chemicals tested (Chemicals I, N, and T).

Overall concordance for the T Pathway was:

Positive Responses all three T Pathway Bioassays = 19% (i.e. of the chemicals tested)

Negative Responses all three T Pathway Bioassays = 19%

Positive Responses in both rat pubertal assays but not AMA = 19%

Positive Responses in only the AMA = 19%

Positive Response in the female pubertal and AMA = 6.7%

Positive Response in only one of the 3 bioassays = 19%

- c. *The EPA concluded that the battery has performed as anticipated by the 2008 SAP. Please comment on this conclusion.*

The 2008 SAP concluded that the Tier 1 suite of assays should be able to detect agents that alter E, A, and T pathways. Although there was general agreement that the battery appears to have performed as expected with the subset of chemicals on which data summaries were provided, there can be limited conclusions drawn at this time given the amount of information that has been provided. The assays were able to be performed by several laboratories and showed changes in the assay critical endpoints that will be used to determine if a compound should be required to undergo further testing to establish definitively the ability of the compound to alter the E, A, or T pathways. Without information on the direction, magnitude and consistency of the observed effects, it is difficult to fully assess how the assays performed as a group. No compounds were clearly positive in the estrogen receptor-binding assay or in the uterotrophic assay, but several chemicals were positive in the ER transcriptional activation assay and also in the pubertal

female and/or FSTRA. It might be expected that positives in the transcription assay would give positives in the ER binding and/or uterotrophic assays. With respect to the *in vitro* assays, one point that has been made repeatedly is that there have been significant scientific advancement in *in vitro* assay technology that may allow for improved testing, and the EPA should take this into account. In any case, the battery of assays did show responses in one or more assays and/or endpoints for E, A, and T pathways. As mentioned above and in previous charge questions, overt toxicity overlapping potential hormonal effects is a concern in several of the *in vivo* assays, particularly the FSTRA and the pubertal assays. Thus, one might question whether the degree of selectivity envisioned by the 2008 SAP was achieved. The WoE analysis may address this.

Several panel members indicated strongly that the use of positive controls that are well known to interact with E, A, and T pathways should have been used to evaluate the entire battery as was done in the validation of each individual assay. The ability of the battery to clearly detect such known compounds would have been helpful in providing confidence that the battery functioned as intended and serve as a means for determining which assays might be considered for removal or replacement.

Charge Question 8. The EPA is committed to minimizing animal usage in the screening battery while maintaining the effectiveness of the battery to answer the question of whether a chemical has the “potential” to interact with the endocrine system.

- a. *In 1998, the EDSTAC described the conceptual framework for Tier 1 assays and recommended the strategy to “require the minimal number of screens and tests necessary to make sound decisions, thereby reducing the time needed to make these decisions”, and that the screens should be conducted at a minimal cost necessary to make decisions. Based on the preliminary battery performance evaluation, to what extent can the current Tier 1 battery of 11 assays be modified to reduce animal usage and/or lower cost while adequately ensuring the EPA’s ability to answer the question of “whether a chemical has the potential to interact with the endocrine system?” More specifically, please comment on whether the Uterotrophic and Hershberger assays provide necessary redundancies in the Tier 1 battery based on this preliminary analysis. Please include in your comments what information may be lost and what uncertainties may be introduced by absence of either or both of these assays.*

Panel Response

Based on the data presented, the majority of the Panel believed that a definitive decision to discard the Uterotrophic assay or Hershberger assays from the battery cannot be made at this time. While the tests are limited in what they can detect, and are considered by some as “*in vivo* test tube assays,” agents with estrogen agonist or anti-androgenic activities have been of particular concern up to this point. Both the Hershberger and uterotrophic assays are well-accepted tests of gonadal hormone action on peripheral tissues. These assays have the ability to very specifically detect androgenic- or

estrogenic- receptor mediated effects of test compounds using *in vivo* models that (when gavage is used) involve important parameters such as a realistic method of test substance administration and metabolism. When expanded they also have the ability to test for anti-androgenic and anti-estrogenic effects. They are simple to run and have easy to measure endpoints. In comparison to the male and female pubertal assays, they have the advantage of testing the potential androgenic or estrogenic effects of various compounds using model systems where possible indirect, non-HPG related effects (i.e. stress response parameters) on gonadal androgen or estrogen secretion are not directly related to the effects on HPG function. While reduction of animal usage and unnecessary testing are major concerns, the majority of the panel members believed that these assays should be retained as they allow for specifically testing the actions of potential endocrine disrupting agents and their metabolites in a mammalian system, and may provide more definitive information than that which may be gleaned from the more comprehensive Pubertal assays.

At the very least, a decision to remove these assays from the battery should await the completion of the WoE analysis on these compounds to determine what role these assays play in the decision process. Several members of the Panel indicated strongly that the WoE analysis should include all of the compounds on which data have been collected. In the case of the Hershberger assay, for example, the Panel observed from the data presented that a response in the Hershberger assay was accompanied by a response in the male pubertal or the FSTRA assay, but it was not indicated if these effects were in the same direction. In the case of the Uterotrophic assay, it is not clear at this point if the negative response for the 21 chemicals evaluated in the set of chemicals under review could be factored in to a decision on a compound that showed an effect in the FSTRA assay only in the presence of overt toxicity. Even if they are removed from the battery as absolute requirements, they should be maintained as options if the other assays do not provide clear results. While, as indicated above, reduction of animal use and cost are important, it is also important that the screening assays serve their intended function to limit the number of compounds that go to Tier 2 while minimizing false negatives. A potential problem indicated by the Tier 1 results from the first set of 21 compounds is the potential confounding of endocrine effects with overt toxicity in the FSTRA and to a lesser extent, the pubertal assays.

The advantages of moving away from the current ER and AR binding assays that use rat tissue cytosols were discussed at multiple points during the meeting. One advantage of cell line based assays would be to reduce animal usage, another would be increased robustness and reproducibility. The Panel recommended that the Agency give serious consideration to replacing the binding assays.

- b. *Please comment on the scientific criteria the Agency should consider in evaluating necessary redundancies and eliminating assays from the current battery.*

Panel Response

A decision to remove a battery from the assay could potentially be made after a thorough review of all data from the complete set of compounds that have been run through the Tier 1 screens. If an assay consistently produces results that are not clearly interpretable, removal could be considered. Likewise, if assays produce data that always only confirm the results of other assays and are not helpful in resolving questionable data, they could be considered for removal.

With regard to evaluating potential modifications to the current battery, the need for more extensive data evaluation was stressed. The need to increase the database, chemical and endocrine space, reveal compound identities and compare with existing data on mechanisms of action, and consider any potency data available was stressed by one panel member. Several panel members recommend a reanalysis of the data using the entire data set of 52 chemicals to provide an assessment of complementary and redundancy relationships. This more robust data set would also be amenable to analysis to determine if a particular assay or assays could be eliminated without increasing the probability of finding false negatives. Further, after completion of this assessment it would also be useful to consider splitting the Tier 1 assay battery into two sub-tiers using the *in vivo* assays as a prescreen followed by targeted *in vitro* tests to explore mechanisms.

References

Bolger, T. and Connolly, P.L. 1989. The selection of suitable indices for measurement and analysis of fish condition. *Journal of Fish Biology* 34(2): 171-182.

Cavigelli, S A, Monfort, S L, Whitney, T K, Mechref, Y S, Novotny, M and McClintock, M K, 2005. Frequent serial fecal corticoid measures from rats reflect circadian and ovarian corticosterone rhythms *Journal of Endocrinology* 184, 153–163; Available at:
<http://psychology.uchicago.edu/people/faculty/cacioppo/jtcreprints/frequentserialfecal2005.pdf>

Chapman K, Sewell F, Allais L, Delongas J-L, Donald E, Festag M, Kervyn S, Ockert D, Nogués V, Palmer H, Popovic M, Roosen W, Schoenmakers A, Somers K, Start C, Stei P, Robinson, S 2013. A global pharmaceutical company initiative: An evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. *Regulatory Toxicology and Pharmacology*.
<http://dx.doi.org/10.1016/j.yrtph.2013.04.003>

Ellenberg, W. 2000. Functional expression of the multidrug (xenobiotic) transporter (Pgp) in tissues of the killifish from selected urban and industrial sites in estuaries of South Carolina. Medical Univ. of South Carolina, Marine Biomedicine and Environmental Science Program. PhD Dissertation: 192pp.

Fort, D.J., Degitz S., Tietge, J., and Touart, L.W. 2007. The hypothalamic-pituitary-thyroid (HPT) axis in frogs and its role in frog development and reproduction. *Critical Reviews in Toxicology* 37: 117-161.

Geach, T. J. and Dale, L. 2005. Members of the lysyl oxidase family are expressed during the development of the frog *Xenopus laevis*. *Differentiation* 73(8): 414–424.

Grim KC, Wolfe M, Braunbeck T, Iguchi T, Ohta Y, Tooi O, Touart L, Wolfe DC, Tietge J. 2009. Thyroid histopathology assessments for the amphibian metamorphosis assay to detect thyroid-active compounds. *Toxicologic Pathology* 37:415-424.

Guinn A.G. 1996. Effect of chronic stress on estradiol action in the uterus of ovariectomized rats. 66(2):169-74.

Hu F, Sharma B, Mukhi S, Patino R, and Carr J.A. 2006. The colloidal thyroxine (T4) ring as a novel biomarker of perchlorate exposure in the African clawed frog *Xenopus laevis*. *Toxicological Sciences* 93: 268–277.

Marshall GA, Amborski RL, Culley DD. 1980. Calcium and pH requirements in the culture of bullfrog (*Rana catesbeiana*) larvae. *Journal of the World Aquaculture Society* 11: 445-453.

Martinez I, Alvarez R, Herraez I, Herraez P. 1992. Skeletal malformations in hatchery reared *Rana perezi* tadpoles. *Anatomical Records* 233(2): 314-320.
osteolathrogenic effects of semicarbazide. *Toxicology* 36: 185-198.

National Center for the Replacement Refinement and Reduction of Animals in Research (NC3R²), 2013., Industry evidence sets animal welfare benchmark for short-term toxicity studies 10% upper limit of bodyweight loss sufficient to determine maximum-tolerated dose; Available online at: <http://www.nc3rs.org.uk/news.asp?id=1932>

Schultz, T.W., Dumont, J.N. and Epler, R.G. 1985. The embryotoxic and Leibovitz HE, Culley DD, Geaghan JP. 1982. Effects of vitamin C and sodium benzoate on survival, growth and skeletal deformities of intensively culture bullfrog larvae (*Rana catesbeiana*) reared at two pH levels. *Journal of the World Aquaculture Society* 13: 322-328.

Sharma B, Patiño R 2008. Exposure of *Xenopus laevis* tadpoles to cadmium reveals concentration-dependent bimodal effects on growth and monotonic effects on development and thyroid gland activity. *Toxicological Sciences* 105:51-58.

USEPA 2008. U.S. Environmental Protection Agency. Scientific Advisory Panel meeting. Review the Endocrine Disruptor Screening Program (EDSP) Proposed Tier-1 Screening Battery. March 25-28, 2008. Arlington, VA.
http://www.epa.gov/scipoly/sap/meetings/2008/032508_mtg.htm#frn

USEPA 2009. U.S. Environmental Protection Agency. Federal Register Notice (FRN), volume 74, page 54422; October 21, 2009