ENDOCRINE DISRUPTOR SCREENING PROGRAM

Weight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening to Identify the Need for Tier 2 Testing

Office of Chemical Safety and Pollution Prevention
US Environmental Protection Agency
Washington DC 20460

September 14, 2011

PREFACE

The Agency submitted a draft of the Weight-of-Evidence (WoE) document for public review and comment as described in a Federal Register Notice issued November 4, 2010 (75 FR 67963). Submitted public comments were compiled and grouped according to the commonality among individual submissions so that they could be more readily and fully considered by EPA during revision of the WoE document. The WoE approach that has been revised and described herein is expected to provide general guidance to EPA staff and managers who will be reviewing data submitted in response to Orders for Tier 1 screening that began October 29, 2009 under the Endocrine Disruptor Screening Program (EDSP). Additionally, outside parties submitting data may be interested to know how the results from Tier 1 screening are being evaluated. This paper provides general guidance and is not binding on either EPA or any outside parties. The use of language such as “will,” “is,” “may,” “can,” or “should” in this paper does not connote any requirement for either EPA or any outside parties. As such, EPA may depart from the guidance where circumstances warrant and without prior notice.

Application of WoE analysis is an integrative and interpretive process routinely used by EPA to evaluate health (USEPA 1991; 1996; 2002a; 2005) and ecological (USEPA, 1998) toxicity in a manner that takes into account all relevant scientific and technical information. The principles and criteria for weighing and integrating different lines of evidence articulated in existing EPA documents are considered generally applicable to evaluating data from the EDSP Tier 1 battery.

It should be recognized that significant advances in both computational and molecular-based technologies are enabling a more rapid identification of markers for evaluating toxicity pathways since EPA began work on developing and implementing the EDSP in 1998. In 2007, the National Research Council Report ‘‘Toxicity Testing in the 21st Century: A Vision and a Strategy’’ (NRC, 2007) acknowledged these advances and recommended that the Agency develop a strategy to use modern *in silico,* computational models and molecular-based *in vitro* high-throughput screening assays to

increase the efficiency of, and reduce and ultimately replace reliance on, whole-animal toxicity testing. Currently, there are ongoing efforts within and outside the Agency to use endocrine screening as a prototype for applying these contemporary methods as proposed by the NRC. A key objective of the work is to improve the speed, reliability, cost effectiveness, and mechanistic specificity of the EDSP. In acknowledging this ongoing research, it should be stressed that the Agency's risk assessment guidance documents are typically viewed as "living documents" that is, they are open to periodic updates and revisions to reflect advances in the science and technology. Although the general principles and criteria articulated in this document for using a WoE approach apply to any study type, this policy is open to periodic updates to incorporate important new scientific and technical knowledge as it becomes available.

TABLE OF CONTENTS

Abbreviations 5
1. PURPOSE AND SCOPE OF DOCUMENT 6
2. ENDOCRINE DISRUPTOR SCREENING PROGRAM (EDSP) OVERVIEW 7
2.1. EDSP Tier 1 Battery of Screening Assays 8
2.1.1. Assays for Detecting the Effect of Chemicals on the Estrogen Hormonal Pathway 10
2.1.2. Assays for Detecting the Effect of Chemicals on the Androgen Hormonal Pathway 11
2.1.3. Assays for Detecting the Effect of Chemicals on the Steroidogenic Pathway 13
2.1.4. Assays for Detecting the Effect of Chemicals on the HPG Axis 13
2.1.5. Assays for Detecting the Effect of Chemicals on the HPT Axis 14
2.2. EDSP Tier 2 Testing 15
3. SOURCES OF SCIENTIFIC AND TECHNICAL INFORMATION 15
3.1. Test Guidelines — EDSP Tier 1 Screening Studies 16
3.2. Scientifically Relevant Information 17
3.2.1. Test Guidelines — Health and Ecological Effects Studies 17
3.2.2. Published or Publically Available Peer-reviewed Studies 21
4. QUALITY OF SCIENTIFIC AND TECHNICAL INFORMATION 22
4.1. General Assessment Factors (GAF) 23
4.1.1. Soundness 24
4.1.2. Applicability and Utility 24
4.1.3. Clarity and Completeness 25
4.1.4. Uncertainty and Variability 25
4.1.5. Evaluation and Review 26
4.2. Standard Evaluation Procedure (SEP) and Data Evaluation Record (DER) 26
5. WEIGHT-OF-EVIDENCE APPROACH 26
5.1. Assembling and Evaluating the Individual Studies 29
5.2. Integrating the Different Lines of Evidence 35
5.3. Weight-of-Evidence Narrative/Characterization 40
5.4. EDSP Tier 2 Testing Recommendations 41
6. SUMMARY 44
7. REFERENCES 45

Abbreviations

Abbreviation	Terminology
A	Androgen (hormonal pathway)
AR	Androgen Receptor
DER	Data Evaluation Record
EDSTAC	Endocrine Disruptor Screening and Testing Advisory Committee
EDSP	Endocrine Disruptor Screening Program
E	Estrogen (hormonal pathway)
ER	Estrogen Receptor
FIFRA	Federal Insecticide, Fungicide, Rodenticide Act
FFDCA	Federal Food, Drug, and Cosmetic Act
FQPA	Food Quality Protection Act
GAF	General Assessment Factors
GLP	Good Laboratory Practices
HPG	Hypothalamic-Pituitary-Gonadal Axis
HPT	Hypothalamic-Pituitary-Thyroidal Axis
MoA	Mode of Action
OCSPP	Office of Chemical Safety Pollution and Prevention
OECD	Organization for Economic Co-Operation and Development
PND	Postnatal Day
SAB	Science Advisory Board
SAP	Scientific Advisory Panel
SEP	Standard Evaluation Procedure
T	Thyroid (hormonal pathway)
WoE	Weight-of-Evidence

1. PURPOSE AND SCOPE OF DOCUMENT

This guidance document provides basic principles and criteria for using a weight-of-evidence (WoE) approach for evaluation and interpretation of EDSP Tier 1 screening, which includes Tier 1 assay results and other information to identify candidate chemicals for Tier 2 testing. General guidance is also provided on the considerations that will inform the tests and information that may be needed for Tier 2 testing.

The purpose of the EDSP Tier 1 battery of screening assays is to identify chemicals that have the potential to interact with the estrogen, androgen, or thyroid (E, A, or T) hormonal pathways. Currently, the battery consists of 11 assays that have been developed and validated through a collaborative effort involving EPA program and research offices and published as harmonized test guidelines by the Office of Chemical Safety Pollution and Prevention (OCSPP 890 Guideline Series, Table 1). EPA intends to evaluate the results of the Tier 1 screening assays using a WoE approach to determine whether or not a chemical has the potential to interact with E, A, or T hormonal pathways and to assess the need for Tier 2 testing. The purpose of Tier 2 testing is to further characterize the effects on E, A, or T identified through Tier 1 screening by using Tier 2 *in vivo* studies that establish dose-response relationships for any potential adverse effects for risk assessment.

- EPA refers to the WoE approach as "…*a collective evaluation of all pertinent information so that the full impact of biological plausibility and coherence is adequately considered.*" (USEPA, 1999).
- In its recommendations to EPA, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) referred to the WoE approach as "…*a process by which trained professionals judge the strengths and weaknesses of a collection of information to render an overall conclusion that may not be evident from consideration of the individual data*" (EDSTAC, 1998).

The WoE approach for Tier 1 screening is discussed in Section 5 which is preceded by introductory and supportive information, including a brief historical overview of the

EDSP two-tiered screening and testing paradigm (Section 2), sources of scientific and technical information (Section 3), and general guidance for determining the quality and relevance of scientific and technical information (Section 4).

2. ENDOCRINE DISRUPTOR SCREENING PROGRAM (EDSP) OVERVIEW

A detailed history of the program can be found at the EDSP website (http://epa.gov/endo/) and in other documents or websites referenced in this document.

In 1996, amendments to the Federal Food, Drug, and Cosmetic Act (FFDCA) required EPA to:

> *"…develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator may designate" [21 U.S.C. 346a(p)].* (http://www.epa.gov/pesticides/regulating/laws/fqpa/)

Pursuant to the Administrator's discretionary authority, EPA adopted a two-tiered screening and testing strategy and expanded the EDSP to include the androgen and thyroid hormonal pathways and ecological effects:

- In 1998, subsequent to the EDSTAC recommendations (EDSTAC, 1998), EPA notified the public of a proposed EDSP as described in a Federal Register Notice issued December 28, 1998 (63 FR 71542). EPA submitted the proposal for review by the Agency's Science Advisory Board (SAB) and FIFRA Scientific Advisory Panel (SAP). A final report of the joint peer review is available (SAB/SAP, 1999).
- In 2008, after an extensive validation process (USEPA, 2007), including peer review of individual assays, EPA notified the public of the EDSP proposed Tier 1 battery of screening assays in a Federal Register Notice issued January 24, 2008 (73 FR 4216). EPA submitted the proposed battery for peer review by FIFRA SAP. A final report of the peer review is available (SAP, 2008).

- In 2009, EPA notified the public of the current EDSP Tier 1 screening battery (Table 1) and availability of harmonized test guidelines (OCSPP 890 Guideline Series) for each of the assays in a Federal Register Notice issued October 21, 2009 (74 FR 54416).
- In 2009, after public review and comment, a final list of 67 chemicals and schedule for issuing Test Orders for Tier 1 screening was made available in a Federal Register Notice issued October 21, 2009 (74 FR 54422).

2.1. EDSP Tier 1 Battery of Screening Assays

The current EDSP Tier 1 battery consists of 11 diverse yet complementary *in vitro* and *in vivo* screening assays as recommended by the FIFRA SAP (SAP, 2008) and is indicated in Table 1. The battery of assays was designed to be conducted as a whole to maximize sensitivity and reliability for determining the potential of a chemical to interact with the E, A, or T hormonal pathways (EDSTAC, 1998). Various factors contributed to selecting the Tier 1 screen that generally included the potential of the assays to evaluate:

- E, A, or T hormonal pathway effects in different taxa,
- estrogen- and androgen-mediated effects via receptor binding (agonism and antagonism),
- estrogen-mediated gene transactivation,
- enzyme inhibition involving the reproductive steroidogenesis pathway,
- interactions with gonadal estrogen and androgen production that may alter feedback mechanisms involving the hypothalamic-pituitary-gonadal (HPG) axis,
- androgen and estrogen influenced endpoints within an assay that are complementary among the assays, and
- interactions with thyroid hormone production or function and associated alterations in feedback relationships involving the hypothalamic-pituitary-thyroid (HPT) axis.

The robustness of the Tier 1 battery is based on the strengths of each individual assay and the complementary endpoints within the battery. Thus, *"…the value of each*

individual assay cannot be considered in isolation from other assays in the battery, as they have been combined in a manner such that limitations of one assay are complemented by the strengths of another" (EDSTAC, 1998).

Table 1. The EDSP Tier 1 screening assays encompass key endpoints within a MoA (*e.g.,* receptor binding) and along endocrine pathways (*e.g.,* steroidogenesis, effects on hypothalamic-pituitary-gonadal and –thyroid axes) through which a chemical has the potential to interact with the estrogen, androgen, or thyroid (E, A, or T) hormonal pathways.*

Screening Assay	Test Guideline	Receptor Binding				Steroidogenesis		HPG Axis	HPT Axis
		E	Anti-E	A	Anti-A	E	A		
In vitro									
ER Binding (Rat uterine cytosol)	OCSPP 890.1250	■	■						
ERα Transcriptional Activation (Human cell line HeLa-9903)	OCSPP 890.1300 OECD 455	■							
AR Binding (Rat prostate cytosol)	OCSPP 890.1150			■	■				
Steroidogenesis (Human Cell Line H295R)	OCSPP 890.1550					■	■		
Aromatase (Human target tissue or cell-line microsomes)	OCSPP 890.1200					■			
In vivo									
Uterotrophic (Rat)	OCSPP 890.1600 OECD 440	■							
Hershberger (Rat)	OCSPP 890.1400 OECD 441			■	■		■[1]		
Pubertal Male (Rat)	OCSPP 890.1500			■	■		■	■	■
Pubertal Female (Rat)	OCSPP 890.1450	■	■			■		■	■
Fish Short-term Reproduction	OCSPP 890.1350 OECD 229	■	■	■	■	■	■	■	
Amphibian Metamorphosis (Frog)	OCSPP 890.1100 OECD 231								■

*Complementary endpoints across assays are indicated (solid black box) within each column.
[1]5α-reductase inhibition only.

In addition to the test guidelines, a detailed characterization of each Tier 1 screening assay, including its development, validation, strengths, and limitations, can be found in EPA Integrated Summary Reports or OECD Final Reports for individual assays at the EDSP website (http://epa.gov/endo/pubs/assayvalidation/index.htm). However, for the purposes of this document, an overview of the distinctive characteristics of each assay and their complementary endpoints within the battery of assays is provided in the next sections.

2.1.1. Assays for Detecting the Effect of Chemicals on the Estrogen Hormonal Pathway

Five Tier 1 assays are capable of detecting chemicals with estrogenic and anti-estrogenic activity (Table 1):

1. ER binding (rat uterine cytosol),
2. ERα transcriptional activation (human cell line HeLa-9903),
3. Uterotrophic (rat),
4. Pubertal female (rat), and
5. Fish short-term reproduction.

The *in vitro* ER binding assay examines the potential of a chemical to bind estrogen receptors (*i.e.,* ERα or ERβ) isolated from the cytosol of excised rat uterine tissue. However, binding alone cannot distinguish whether the chemical is an estrogen agonist or antagonist. The *in vitro* ER transcriptional activation assay examines the potential of an estrogen agonist to activate ER- (*i.e.,* ERα) mediated gene transcription in cells derived from a human cervical tumor.

In vivo assays used to evaluate the estrogen pathway involve different routes of exposure to a chemical such as subcutaneous injection (uterotrophic), oral gavage (pubertal), and water (fish). The uterotrophic assay is conducted using adult ovariectomized or sexually immature intact female rats and has the potential to detect estrogen agonist activity based on:

- an increase in uterine weight, and
- optional histology of uterus and vagina.

The pubertal female assay is conducted in rats post-weaning and has the potential to detect both estrogen agonist and antagonist activity based on multiple endpoints:

- sexual developmental characteristics (age at vaginal opening and estrous cyclicity, length and percent of animals cycling), and
- weight and histology of reproductive organs (ovaries with oviducts and uterus with fluid, and pituitary gland).

The fish short-term reproduction assay is conducted using mature male and female fathead minnows and has the potential to detect both estrogen agonist and antagonist activity based on multiple endpoints:

- behavior,
- fecundity,
- fertilization success,
- secondary sex characteristics (number and size of nuptial tubercles and dorsal nape pad in males),
- survival,
- body weight and length,
- gonadal size (gonado-somatic index) and histopathology,
- assay of plasma concentrations of vitellogenin in females and males, and
- optional assay of plasma concentrations of estradiol in females.

2.1.2. Assays for Detecting the Effect of Chemicals on the Androgen Hormonal Pathway

Four Tier 1 assays are capable of detecting an androgenic or anti-androgenic effect of a chemical (Table 1):

1. AR binding (rat prostate cytosol),
2. Hershberger (rat),
3. Pubertal male (rat), and

4. Fish short-term reproduction.

The *in vitro* AR binding assay examines the potential of a chemical to bind to the androgen receptor isolated from the cytosol of rat prostate tissue. However, binding alone cannot distinguish whether the chemical is an androgen agonist or antagonist.

In vivo assays used to evaluate the androgen pathway involve different routes of exposure such as subcutaneous injection or oral gavage (Hershberger), oral gavage (pubertal), and water (fish). The Hershberger assay is conducted using castrated male peripubertal rats and has the potential to detect androgen agonist and antagonist activity as well as 5α-reductase inhibitors (*i.e.*, inhibition of the conversion of exogenous testosterone to dihydrotestosterone) based on multiple endpoints involving changes (increase or decrease) in weight of androgen-dependent organs or tissues:

- ventral prostate,
- seminal vesicle plus coagulating gland with fluid,
- levator ani plus bulbocavernous muscle complex,
- paired Cowper's gland, and
- glans penis.

The pubertal male assay is conducted in rats post-weaning and has the potential to detect both androgen agonist and antagonist activity based on multiple endpoints:

- sexual developmental characteristic (age at preputial separation),
- weight and histology of reproductive organs or tissues (testes, epididymides, ventral and dorsolateral prostate, seminal vesicle plus coagulating gland with fluid, levator ani plus bulbocavernous muscle complex, and pituitary gland), and
- assay of total serum concentrations of testosterone.

The fish short-term reproduction assay is conducted using mature male and female fathead minnows and has the potential to detect both androgen agonist and antagonist activity based on multiple endpoints as already indicated in Section 2.1.1, including optional assay of plasma concentrations of testosterone in males.

2.1.3. Assays for Detecting the Effect of Chemicals on the Steroidogenic Pathway

Six Tier 1 assays are capable of detecting disruption in the steroidogenic pathway (Table 1):

1. Steroidogenesis (human cell line H295R),
2. Aromatase (human target tissue or cell-line microsomes),
3. Hershberger (rat),
4. Pubertal male (rat),
5. Pubertal female (rat), and
6. Fish short-term reproduction.

The *in vitro* steroidogenesis assay uses a human cell line (H295R) to examine the potential of a chemical to interact with the steroidogenic pathway based on the change (increase or decrease) in production of testosterone and estradiol. The *in vitro* aromatase assay uses human microsomes from various target tissues or cell lines to detect the inhibition of aromatase activity and the conversion of androgen to estrogen. The assay is not well suited to detect induction of aromatase activity.

In regard to the *in vivo* assays and corresponding endpoints presented in preceding sections, an apparent effect on steroidogenesis (*i.e.,* gonadal production of estrogen or androgen) may be observed based on changes in respective endpoints even though there are no apparent corroborating steroidogenic effects observed *in vitro*. It is possible that a chemical can disrupt steroidogenesis indirectly by acting directly on gonadotropin synthesis or secretion or by eliciting feedback responses along the HPG axis. For example, exposure to a chemical with androgen or estrogen activity may cause reductions in endogenous androgen or estrogen concentrations as a result of negative feedback along the HPG axis. The implication for an effect at the hypothalamic or pituitary levels is further explored in the next section.

2.1.4. Assays for Detecting the Effect of Chemicals on the HPG Axis

In general, regulation of the HPG axis involves a complex array of positive and negative feedback mechanisms. Gonadal estrogen or androgens bind to corresponding

receptors in the hypothalamus or pituitary to regulate gonadotropic-releasing hormone and gonadotropic hormones, respectively, which in turn regulate ovarian and testicular steroidogenesis. The current Tier 1 screening battery does not have a specific *in vitro* assay to detect chemicals with the potential to affect hypothalamic or pituitary regulation of gonadal estrogen or androgen hormone production but does include three *in vivo* assays that have the potential to detect these effects on the HPG axis (Table 1):

1. Pubertal female (rat),
2. Pubertal male (rat), and
3. Fish short-term reproduction.

It is possible to examine the results of the *in vivo* pubertal and fish short-term reproduction assays in the context of the Tier 1 screening to infer whether or not a chemical is having an effect involving hypothalamic-pituitary regulation of the reproductive axis. A simple example may be that vaginal opening or preputial separation are delayed in the pubertal female or male assays, respectively, but the combined results of the *in vitro* ER- or AR-binding and steroidogenesis and aromatase assays are negative. Although such results may not appear to support a direct effect of a chemical on steroidogenesis, they may suggest the possibility of an indirect effect on steroidogenesis as a result of an effect at the hypothalamic or pituitary levels.

2.1.5. Assays for Detecting the Effect of Chemicals on the HPT Axis

In general, regulation of the HPT axis is comparable to the HPG axis except that the feedback relationship involves thyroid hormones (*e.g.*, thyroxine). Thyroid hormones feedback to the hypothalamus or pituitary to regulate thyrotropin-releasing hormone and thyroid stimulating hormone (TSH), respectively, which, in turn, regulate hormone production by the thyroid gland. The current Tier 1 screening battery does not have a specific *in vitro* assay to detect chemicals with the potential to affect hypothalamic or pituitary regulation of thyroid hormone production but does include three *in vivo* assays that have the potential to detect these effects on the HPT axis (Table 1):

1. Pubertal female (rat),
2. Pubertal male (rat), and

3. Amphibian metamorphosis (frog).

The pubertal assays have been presented in preceding sections but, in regard to their potential to detect chemicals that may interact with the thyroid hormonal pathway, endpoints include:

- thyroid gland weight and histology and
- assay of serum concentrations of pituitary TSH and thyroid gland thyroxine (T4).

The amphibian metamorphosis assay uses tadpoles to evaluate the effect of chemical exposure through water on the thyroid pathway based primarily on changes in developmental endpoints associated with metamorphosis and thyroid gland histology over days 7 to 21:

- development stage,
- histology of thyroid gland,
- hind limb length, and
- whole body length (snout to vent).

2.2. EDSP Tier 2 Testing

Tier 2 testing is expected to involve more comprehensive studies across taxa (*e.g.,* mammalian, birds, amphibians, fish, and invertebrates) to quantify dose-response relationships in a larger context of toxicity and potential adversity that may involve endocrine as well as non-endocrine systems (*e.g.,* neurological, immunological, hepatic, renal) for human health and ecological risk assessments. A complete discussion of Tier 2 testing is available in the EDSTAC report (EDSTAC, 1998), in a Federal Register Notice issued December 28, 1998 (63 FR 71542), and on the EDSP website (http://epa.gov/endo/).

3. SOURCES OF SCIENTIFIC AND TECHNICAL INFORMATION

The EDSP Tier 1 battery of screening assays was designed to evaluate the potential of a chemical to interact with the E, A, or T hormonal pathways and, therefore, the

collective results from the Tier 1 screening is generally expected to be the main source of scientific and technical information considered in the WoE evaluation. However, other sources of scientific and technical information may also be considered in the evaluation, such as information that is submitted as relevant to Tier 1 screening. Such information could come from any number of sources, including studies conducted by pesticide registrants or chemical companies and from published or publically available peer-reviewed studies.

3.1. Test Guidelines — EDSP Tier 1 Screening Studies

Data generated in response to test orders for EDSP Tier 1 screening come from OCSPP Test Guideline Series 890 indicated in Table 1. These screening assays have undergone an extensive validation process, individual peer review, and independent review by the FIFRA SAP (SAP, 2008). In addition, the test guidelines have been prepared with a level of detail to provide clear guidance to the user, with recommendations on how to conduct each assay and interpret results. Standard Evaluation Procedures (SEP) have also been developed for all screening assays to aid in the evaluation of results (Section 4.2). Subsequent to the availability of the test guidelines and in response to concerns raised by outside parties, EPA prepared a document that provides additional guidance on technical conduct and expectations for applying performance criteria for each of the screening assays (USEPA, 2011a). Furthermore, EPA has developed an EDSP website that provides resources such as assay information, including SEPs and Data Evaluation Record (DER) templates, test order response, and status tracking (http://epa.gov/endo/pubs/toresources/index.htm).

Thus, EPA has provided multiple levels of guidance that are available to Agency staff and managers as well as to outside parties to aid in conducting the Tier 1 screening assays and evaluating results. No one assay or endpoint in or among assays is intended to be interpreted in isolation as emphasized in the WoE approach presented in Section 5. The Tier 1 screening was designed to be conducted as a whole to provide complementary information to support a WoE evaluation that may include other scientifically relevant information to determine the need for Tier 2 testing.

3.2. Scientifically Relevant Information

Information that is submitted voluntarily and is applicable to support or clarify an EPA action is generally referred to as "other" scientifically relevant information. Sources of relevant scientific and technical information may include results from EPA or OECD equivalent test guideline studies and information from published or publically available peer-reviewed studies. Regardless of the source, the information is evaluated for quality and relevance, taking into account the Agency's Information Quality Guidelines (USEPA, 2002b) before use in an EPA action.

EPA began issuing test orders for EDSP Tier 1 screening in the fall of 2009. In the initial responses, test order recipients indicated their intentions to comply with the order and often submitted existing scientific information to be considered by the Agency in lieu of the Tier 1 assays (USEPA, 2009a). EPA considered whether or not the submitted information could fulfill the test order requirements for one or more of the Tier 1 assays and informed the test order recipients accordingly.

To comply with the test orders, recipients must submit the results of EDSP Tier 1 screening. The submission may also include other scientifically relevant information. Sources of relevant scientific and technical information may consist of information that was previously submitted in the initial response to test orders or new or additional information. EPA will consider the additional information submitted and, based on the quality and relevance of that information, will consider it along with the results of the Tier 1 screening assays in a WoE analysis to determine whether or not a chemical has the potential to interact with the E, A, or T hormonal pathways.

3.2.1. Test Guidelines — Health and Ecological Effects Studies

EPA regulations in 40 CFR Part 158, subparts F and G define the toxicological data requirements for health (870 Guideline Series) and ecological (850 Guideline Series) effects, respectively. Data generated in response to FIFRA requirements come from studies using the aforementioned series of peer reviewed test guidelines or OECD

equivalents. These studies are subject to Good Laboratory Practices (GLP) regulations (40 CFR Part 160 and Part 792) to ensure consistency, reproducibility, and integrity of the data. Many of the traditional toxicity studies include apical endpoints that can be affected through multiple modes of action (MoA) involving both endocrine and non-endocrine systems. Hence, certain EPA or OECD equivalent test guideline studies may provide relevant scientific and technical information for consideration along with the results of Tier 1 screening in a WoE analysis to determine whether or not a chemical has the potential to interact with the E, A, or T hormonal pathways.

In the EDSTAC report to EPA (EDSTAC, 1998), specific reference was made to the mammalian two-generation reproductive toxicity study:

> "…*potential hormonal effects can be detected through behavioral changes, ability to become pregnant, duration of gestation, signs of difficult or prolonged parturition, apparent sex ratio (as ascertained by anogenital distances) of the offspring, feminization or masculinization of offspring, number of pups, stillbirths, gross pathology and histopathology of the vagina, uterus, ovaries, testis, epididymis, seminal vesicles, prostate, and any other identified target organs.*"

Test guidelines for the mammalian two-generation reproductive toxicity study as well as the new OECD test guideline for the extended one-generation reproductive toxicity study are proposed as EDSP Tier 2 tests. While the extended one-generation reproductive toxicity study was designed to provide the traditional spectrum of information from a reproductive study, it was enhanced to evaluate reproductive and developmental endpoints associated with the endocrine, nervous, and immune systems in male and female adult rodents and offspring at birth, weaning, and puberty, which may not necessarily be covered in other 40 CFR Part 158 test guideline studies.

For the estrogen hormonal pathway, 40 CFR Part 158 test guideline studies subpart F (or OECD equivalents) for human health effects may be sources of scientifically relevant information. There are estrogen-influenced endpoints in the two-generation and extended one-generation reproductive toxicity studies, developmental neurotoxicity

study as well as the subchronic, chronic, and cancer bioassays. Although test guidelines for these types of studies should be consulted for a full range of endpoints and specific details of measurement, some examples that are considered estrogen-influenced endpoints include:

- age at vaginal opening,
- estrous cyclicity,
- reproductive organ weights and corresponding histopathology, and
- fertility.

For the androgen hormonal pathway, essentially the same studies indicated in the preceding paragraph may be sources of scientifically relevant information. Although respective test guidelines for these types of studies should be consulted for a full range of endpoints and specific details of measurement, some examples that are considered androgen-influenced endpoints include:

- anogenital distance,
- age at preputial separation,
- hypospadias, epispadias, cleft phallus, and areola/nipple retention in male rodent pups,
- reproductive organ weights and corresponding histopathology, spermatogenesis, and
- fertility.

For the thyroid hormonal pathway, additional 40 CFR Part 158 test guideline studies subpart F (or OECD equivalents) may be sources of scientifically relevant information such as the 90-day rodent and dog studies, one-year chronic dog study, and chronic mouse and rat studies. Although test guidelines for these types of studies should be consulted for specific details of measurement, thyroid-specific endpoints include:

- thyroid organ weight and histopathology in the 90-day dog study,
- thyroid histopathology in the 90-day studies in mice and rats,
- thyroid histopathology in the chronic toxicity study in dogs, chronic, toxicity/carcinogenicity studies in rats and the carcinogenicity study in mice, and

- thyroid hormones (T3, T4, and TSH), thyroid weight and histopathology may be carried out if the test chemical is known or suspected of affecting the thyroid gland.

While thyroid hormone endpoints are not included in the traditional mammalian two-generation reproductive toxicity test guideline, they are recommended in the extended one-generation reproductive toxicity test guideline. Additionally, pituitary (TSH) and thyroid gland (T3, T4) hormone measures may be available in special studies.

The 40 CFR Part 158 test guideline studies subpart G for ecological effects are also potential sources of scientifically relevant information. Although these studies include endpoints that may be informative of a chemical to potentially interact with the endocrine system, they are not considered diagnostic. While the avian reproduction, fish full-life cycle, and fish early-life stage test guidelines should be consulted for a full range of endpoints and specific details of measurement, endpoints that may be informative of a potential interaction with the endocrine system include:

- fecundity,
- reproductive success,
- egg development, and
- embryo/larval survival and growth.

Notably, if gross morphologies of internal organs including the gonads are observed to be pathologic, histological analyses may be conducted.

Invertebrates may also provide information that a chemical has the potential to interact with the endocrine system. Apparently, the invertebrate endocrine system depends on a family of steroid-like hormones (*e.g.,* ecdysone) that regulate molting. Although invertebrates are not known to possess functional estrogen or androgen receptors similar to those in mammals, ecdysones bind to nuclear receptors that are part of a superfamily of conserved nuclear receptors that include the estrogen receptor. Currently, the relationship of ecdysones to estrogen binding and transactivation as

evaluated in respective EDSP Tier 1 screening assays is not thoroughly understood. Nonetheless, the chronic full-life cycle test guideline studies are typically conducted on the freshwater flea (*Daphnia magna*) and the estuarine or marine mysid shrimp (*Americamysis bahia*). While respective invertebrate test guidelines should be consulted for a full range of endpoints and specific details of measurement, endpoints that may be informative of a potential interaction with the endocrine system include:

- growth,
- reproduction, and
- survival.

In summary, although EPA regulations in 40 CFR Part 158, subparts F and G define the toxicological data requirements for health and ecological effects, respectively, they were not specifically designed to test for the potential of a chemical to interact with the E, A, or T hormonal pathways. Nonetheless, certain EPA or OECD equivalent test guideline studies may provide contributing scientific and technical information regarding effects related to the endocrine system. The potential effects of a chemical in a broader context of toxicity (*e.g.,* dose-response relationships and adversity) that may be directly related to non-endocrine systems may provide a better understanding of potential indirect effects on the endocrine system. Thus, EPA Part 158 test guideline studies and OECD equivalents may provide relevant scientific and technical information to be considered along with the results of Tier 1 screening in a WoE analysis to determine whether or not a chemical has the potential to interact directly or indirectly with the E, A, or T hormonal pathways.

3.2.2. Published or Publically Available Peer-reviewed Studies

Published or publically available peer-reviewed studies are used by EPA to inform the Agency's understanding of the potential for adverse health and ecological effects associated with chemicals in the context of risk assessment. Correspondingly, the availability of peer reviewed studies may provide additional information along with the results of Tier 1 screening in a WoE evaluation to determine whether or not a chemical has the potential to interact with the E, A, or T hormonal pathways.

In general, published or publically available peer-reviewed studies are conducted in accord with standard scientific methods that include hypothesis development and testing through observation, experimentation, and verification. However, unlike test guideline studies, published studies in the open literature (*i.e.,* non-guideline) do not typically adhere to GLP. Thus, for non-guideline, as well as guideline studies, to be considered as primary or secondary sources of information in a WoE evaluation with Tier 1 screening results, EPA would generally evaluate the quality and relevance of the information indicated in EPA Information Quality Guidelines (USEPA, 2002b).

4. QUALITY OF SCIENTIFIC AND TECHNICAL INFORMATION

Evaluation of the quality of scientific and technical information contained in test guideline, as well as non-guideline studies, is fundamental for consideration of that information in a WoE analysis that would support a regulatory decision. Developed in response to the Office of Management and Budget (OMB, 2002), the *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency* (USEPA, 2002b) contain EPA's policy and procedural guidance for ensuring and maximizing the quality of information. This and other guidance documents relevant to evaluation of the quality of scientific and technical information can be found at the EPA website (http://www.epa.gov/quality/informationguidelines/).

EPA also recognizes there are other reports that document a process for evaluating the quality of data which appear to be based on a seminal report of a systematic approach for evaluating the quality of toxicological and ecotoxicological data (Klimisch *et al.,* 1997).

- OECD Manual for Investigation of High Production Volume (HPV) Chemicals (OECD, 2005),
- Australian Ecotoxicity Database Quality Assessment Scheme (Hobbs *et al*., 2005), and

- Toxicological data Reliability assessment Tool (ToxRTool; Schneider *et al*., 2009).

All these approaches appear similar in principle, primarily relying on "*adequacy,*" "*reliability,"* and "*relevance,*" where reliability (*i.e*., validity or soundness and integrity) of a study is emphasized as the core evaluation criterion.

Within the EPA, Information Quality Guidelines (USEPA, 2002b) build upon the Agency's numerous existing systems, practices, and guidelines that address information quality. In this section, the basis for these guidelines is summarized with an overview of the General Assessment Factors (USEPA, 2003) and considerations for evaluating the quality of scientific and technical information.

4.1. General Assessment Factors (GAF)

In response to the Office of Management and Budget (OMB, 2002), EPA developed and made available EPA Information Quality Guidelines (USEPA, 2002b), which evolved, in part, from other agency-wide and program-specific policies. The information quality guidelines set forth the Agency's policy and procedural guidance for ensuring and maximizing the quality of information, regardless of the source of information. They enhance the transparency of EPA's general approach to evaluation of quality of scientific and technical information that is voluntarily submitted, or gathered or generated by the Agency. EPA's Science Policy Council (SPC) recommended the use of five General Assessment Factors (GAF; USEPA, 2003):

1. *soundness*,
2. *applicability and utility*,
3. *clarity and completeness*,
4. *uncertainty and variability*, and
5. *evaluation and review*.

The GAFs were drawn from existing information in EPA quality systems, practices, and guidelines that describe Agency's considerations for evaluating the quality and

relevance of scientific and technical information used in support of Agency actions USEPA, 2002b). The GAFs do not constitute new quality-related considerations or describe a new process for evaluating information. The GAFs may be applied to individual pieces of information as was done in the selection of studies for an EPA Exposure Handbook (USEPA, 2009b) or to a body of evidence that is collectively evaluated using a WoE approach. As was already defined (USEPA, 1999; EDSTAC, 1998), a WoE approach is an interpretive process that considers all scientifically relevant information in an integrative analysis. This process takes into account various kinds of available evidence, quality and quantity of that evidence, strengths and limitations associated with each type of evidence, and explains how the various types of evidence fit together to support a conclusion.

The following sections provide a general description of the five GAFs as outlined by the SPC (USEPA, 2003). In addition, illustrative considerations are given under each GAF as guidance for evaluating the quality of scientific and technical information that may be submitted as scientifically relevant information. The considerations are not all inclusive and may overlap from one GAF to another.

4.1.1. Soundness

Scientific and technical procedures, measures, methods or models employed to generate the information are reasonable for, and consistent with, the intended purpose.

Considerations: 1) adequacy of the test methods to detect the effect of interest; 2) conduct of studies according to the scientific method of hypothesis development and testing through observation, experimentation, and verification; 3) ability to distinguish between a specific versus a nonspecific outcome according to the intended purpose of the study; and 4) interpretation of results and conclusions that are statistically significant, biologically plausible, and consistent with the data.

4.1.2. Applicability and Utility

The information is relevant for the Agency's intended use.

Considerations: 1) appropriateness of test materials and methods, study design, and endpoints based on rationale, objectives, and hypotheses related to the intended purpose of the study; 2) evidence of competence in collection, analysis, presentation, and interpretation of data and conclusions; and 3) reliability of information from traditional as well as new methodologies.

4.1.3. Clarity and Completeness

The degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations and analyses employed to generate the information are documented.

Considerations: 1) transparency of authors, co-authors, contributors, and acknowledgement of respective institutions or organizations as well as sponsors; 2) background information or rationale, study objectives, hypotheses that are being tested, and experimental design, including controls and number of observations/groups related to the intended purpose of the study; 3) degree of standardization or scientifically valid methodology that supports repeatability with accuracy and precision; 4) availability of raw data; 5) statistical analysis approach; and 6) interpretation of statistical significance, plausibility of biological outcomes, and scientifically sound conclusions.

4.1.4. Uncertainty and Variability

The uncertainty and variability (quantitative and qualitative) in the information or the procedures, measures, methods or models are evaluated and characterized.

Considerations: 1) citation of references pertaining to the specificity and sensitivity of test methods or models, experimental designs, or endpoints; 2) evidence of reproducibility or repeatability of the test method; 3) performance criteria and quality control or assurance measures that may include historical or reference control information, coefficients of variation, GLP compliance, or independent peer review; and

4) number of animals or observations/groups and statistical analysis approach to sufficiently and adequately detect differences between or among groups.

4.1.5. Evaluation and Review

The information or the procedures, measures, methods or models are independently verified, validated, and peer reviewed.

Considerations: 1) explanation or reference of the process for verification or validation to evaluate relevance and reliability of test methods and endpoints as specific and sensitive units of measure; 2) general acceptance of the method in the peer reviewed literature; 3) availability of validation results; and 4) availability of performance or evaluation criteria.

4.2. Standard Evaluation Procedure (SEP) and Data Evaluation Record (DER)

In general, EPA uses SEPs that have been developed for each test guideline study as guidance for evaluating the conduct of each study and interpretation of results. Subsequent to evaluation, a DER is prepared for each test guideline study. In addition, EPA also develops reviews of non-guideline studies submitted as additional information that are used in the WoE analysis. A DER is an official Agency record of review that contains a summary of how well the study was conducted and conforms to the guideline and provides the interpretation and conclusions supported by the data. An SEP and corresponding DER template have been developed for each of the EDSP Tier 1 screening assays and will be used accordingly to evaluate, interpret, and summarize the results for use in a WoE analysis. EDSP Tier 1 screening assay SEPs and DER templates are available on the EDSP website (http://epa.gov/endo/pubs/toresources/index.htm).

5. WEIGHT-OF-EVIDENCE APPROACH

This section of the document will describe the principles, criteria, and approach used in the WoE determination on the potential of a substance to interact with endocrine-

mediated processes (*i.e.,* E, A, or T hormonal pathways) in support of Tier 2 testing decisions.

Generally, WoE is defined as the process for characterizing the extent to which the available data support a hypothesis that an agent causes a particular effect (USEPA 1999; 2002a; 2005). This process involves a number of steps starting with assembling the relevant data, evaluating that data for quality and relevance followed by an integration of the different lines of evidence to support conclusions concerning a property of the substance. WoE is not a simple tallying of the number of positive and negative studies (US EPA 2002a). Rather it relies on professional judgment. Thus, transparency is important to any WoE analysis. A WoE assessment explains the kinds of data available, how they were selected and evaluated, and how the different lines of evidence fit together in drawing conclusions. The significant issues, strengths, and limitations of the data and the uncertainties that deserve serious consideration are presented, and the major points of interpretation highlighted.

As explained in Section 2, the Tier 1 assays were specifically designed to evaluate a number of key biological events including potential effects on receptor binding (estrogen and androgen agonist and antagonist), steroidogenesis, and other effects on the hypothalamic-pituitary-gonadal (HPG) and -thyroidal (HPT) axes. Thus, the WoE approach in this case involves consideration of data from the EDSP Tier 1 assays in reaching and supporting a conclusion to determine whether or not a substance has the potential to interact with the E, A, or T hormonal pathways. As discussed in Section 3.2, other sources of information may be considered as appropriate.

As explained earlier in this document, the purpose of this WoE analysis is to support a determination of whether or not further evaluation of the chemical of interest with EDSP Tier 2 testing is warranted. Given the purpose of this WoE analysis, the following would typically be relevant considerations included in the WoE analysis:

- Do the existing data provide relevant, robust, and consistent evidence (*e.g.,* agreement among the outcomes within an individual assay and among the

different assays or studies) that the substance of interest has the potential to interact with the normal function of the E, A, or T hormonal pathways?

- If the data indicate a potential to interact with those specific endocrine pathways, which hormonal pathway(s) is impacted (E, A, or T) and what kind of Tier 2 testing is appropriate?

The WoE approach to be followed can be characterized as a hypothesis-based approach (USEPA 2005; Boobis *et al,* 2006 and 2008; Rhomberg, 2010). In particular, to bring structure, rigor and transparency to the evaluation of MoA data, a WoE framework was put forth in conjunction with work by the EPA (2005) International Programme for Chemical Safety (IPCS) (Boobis *et al.,* 2006; 2008). The criteria used in the EPA and IPCS MoA/WoE framework are applicable to the EDSP WoE evaluation including considerations of biological plausibility and coherence, strength, and consistency of the body of evidence. Multiple lines of evidence, reflecting the complex nature of endocrine-mediated processes, would be evaluated under this WoE framework to address the hypothesis or question of whether a compound interacts with the E, A, or T hormonal pathways. This question can be generally approached by considering effects at different levels of biological organization using the Tier 1 assays. An illustration of the application of the hypothesis-based approach supported by corroborating evidence at different levels of biological organization follows:

- The interaction of the chemical with a molecular target, such as estrogen receptor antagonism (as measured in the *in vitro ER* binding assay).
- This leads to an altered functional cellular response, which may be indicated by diminished vitellogenin production in females (as measured in the fish short-term reproductive assay).
- This is corroborated by an altered structural response at the organ or tissue level, such as decreased gonado-somatic index (GSI) or altered oocyte or ovarian follicle development *(*as evaluated in females in the fish short term reproductive assay).
- Ultimately, estrogen receptor antagonism may lead to an adverse outcome at the whole organism level, such as decreased fecundity.

In this example, the Tier 1 screening serves to identify the potential of the chemical to interact with endocrine-mediated processes as illustrated in the first three bullets. Tier 2 *in vivo* assays provide information on adverse effects and their dose response at the whole organism level as noted in the last bullet.

The individual assays that comprise the EDSP Tier 1 battery were designed to be complementary to one another as discussed in Section 2. As a consequence, a more thorough understanding of an E, A, or T endocrine interaction is obtained by the combined analysis of the Tier 1 assays. A fundamental point made throughout this document is that multiple lines of evidence are evaluated in an integrated manner during the WoE evaluation wherein no one study or endpoint is generally expected to be sufficiently robust to support a decision of whether or not Tier 2 testing is needed.

This WoE analysis is conducted on a case-by-case basis by first assembling and assessing the individual lines of evidence (*i.e.,* the specific assays, Section 5.1), and then performing an integrated analysis of those lines of evidence (*i.e.,* all assays, Section 5.2).

As described in the next section, all data considered in the WoE analysis need to be documented and scientifically acceptable.

5.1. Assembling and Evaluating the Individual Studies

A WoE analysis typically begins with a careful evaluation of each individual study. The process of evaluating the individual lines of evidence includes assembling the data, evaluating that data against current acceptance and quality criteria, and presenting the conclusions regarding the results for each study. The reviews of the available studies need to be transparent about what studies were considered or not, and how the quality of a study was judged. As discussed later in Section 5.2, the results of individual studies can be tabulated by study type and by endpoint to provide a structured and transparent approach to facilitate the WoE determination.

When assembling and evaluating the data, the information considered need to not only be scientifically sound but relevant to addressing whether there is a need for additional testing in Tier 2. As discussed in Section 2, the EDSP Tier 1 assays have been designed to determine whether or not a substance interacts with E, A, or T hormonal pathways, and are conducted using scientifically peer reviewed study protocols. If the Tier 1 studies are performed properly, the quality of the data would generally be expected to be sound and appropriate for determining whether or not a compound interacts with E, A, or T. Thus, an important aspect of the evaluation is consideration of the methodological strengths and limitations of each study to detect a potential interaction. For example, some of the strengths and potential limitations of the individual Tier 1 assays can be found in EPA Integrated Summary Reports or OECD Final Reports for individual assays at the EDSP website, the 2008 FIFRA SAP report (SAP, 2008), as well as in other reviews (*e.g.,* Eldridge and Laws, 2010; Bogert *et al.,* 2011). For each study, the Agency will review the test methods employed and the conditions under which the studies (both guideline and non-guideline studies) were conducted to assess the standard of scientific quality, and thus, the level of confidence in the study findings to contribute to the WoE determination. In addition to evaluation of the quality and relevance of scientific and technical information presented in a general context in Section 4, the evaluation of individual EDSP Tier 1 assays or collection of assays or studies in the context of this WoE determination is facilitated by using the questions below to guide the analysis. Not all questions are relevant in every case. In addition, there may be other questions an individual reviewer may find appropriate.

Considerations for all sources of data:

The Quality/Validity of the Method

- For the EDSP Tier 1 assays, how well was the test guideline followed for the specific assay under consideration? Were there any deviations, and were they clearly described? Do the deviations have an impact on the study outcome or its interpretability?

- For assays that are non-guideline studies, are the experimental procedures, methods and models scientifically sound, well documented, and appropriate for the evaluation of the E, A, or T endocrine activity of concern?
- For non-guideline studies, were the methods described in sufficient detail to permit an independent evaluation of the material used, equipment requirements, measurement procedures, controls, and test strengths and limitations? Were the results reported in sufficient detail to allow for an independent evaluation?
- For assays not following EDSP Tier 1 protocols, did the studies meet other quality criteria? (Note: EPA and equivalent OECD test guideline studies are typically conducted in accordance with GLP, 40 CFR Part 160 and Part 792).

The Reliability of the Results

- Although the EDSP Tier 1 assay guidelines have been validated, did the laboratory sufficiently demonstrate they can conduct the Tier 1 assays reliably based on assay performance or criteria as described in the test guidelines and SEPs?
- For assays not following EDSP Tier 1 protocols, is there confidence in the measurement of the endpoint(s) evaluated?
- Was the experimental design adequate (*e.g.,* purity and stability of test material, vehicle or solvent used adequate, dosing regimen, adequate number of animals tested, species and strain of animals)?
- Did the study include the appropriate positive and negative controls to evaluate the experimental design and performance (when applicable)?
- Did the number of animals or *in vitro* replicates follow the test guideline recommendations? Was the rationale for dose selection clearly presented and were the doses selected appropriate?
- Was there an adequate description of the statistical analysis? Was the proper statistical analysis selected and was it performed correctly?

The Nature of the Effect(s) Observed

- Was the effect of the test substance clearly described?
- Were the responses observed in the positive and negative controls (when applicable) appropriate?
- Under what experimental conditions were the responses reported? For example, what were the environmental and physiologic conditions of the test system? Could the route of exposure affect the response?
- What was the degree of the response? If several treatment doses were evaluated, what was the nature of the dose response?
- For *in vitro* studies, what was the shape of the concentration response? Was there evidence of cytotoxicity? What was the cytotoxicity assay employed? Was testing performed over an adequate range of concentrations? Was the effect observed only at cytotoxic concentrations?
- Were there issues with solubility of the test chemical (applies to both *in vitro* and *in vivo* tests)? Were the limits of solubility for the test material provided? How was this identified and handled in the study?
- For *in vivo* tests, what clinical signs, body weight changes, and other non-target changes in the animal's health were noted?
- Depending on the type of study and effect measured, was the effect severe or mild, persistent, reversible, or transient if evaluated?
- Was there substantial variability associated with the responses? Is the response within normal variation for the assay or species or strain?

Consistency and Interrelationship among Endpoints Reported in an Individual Assay

- For those studies (generally applies to *in vivo)* evaluating multiple endpoints, was a consistent pattern of effects found among the measured endpoints to support a potential interaction with E, A, or T hormonal pathways?

Relevance, Specificity, and Sensitivity of the Endpoint(s) Measured

- For non-Tier 1 guideline studies: Did the assay measure endpoints that provide useful information for evaluating the potential of a chemical to interact with E, A, or T hormonal pathways?
- Were the studies conducted using a sensitive model during a sensitive or susceptible period [*e.g.,* female peripubertal animals exposed to the test material during postnatal days (PND) 22 to 42 with examination for vaginal opening and body weight beginning on PND 22]?
- Could the reported effects arise from non-endocrine initiating events (*e.g.,* general systemic toxicity)?

As discussed earlier (Section 4.2), the documentation of study quality and conclusions regarding the results of a study are generally done in the DER. A DER will be prepared for each EDSP Tier 1 assay. Each DER will include a statement of whether the assay satisfied or did not satisfy the test order requirement. Reviews are also prepared for published or publicly available peer-reviewed studies that are considered in the WoE evaluation. DERs are also available for the standard toxicity test guideline studies used to meet 40 CFR Part 158 data requirements.

To aid in determining the level of confidence in any study, the strengths of the study and any attendant limitations and uncertainties are generally assessed, explained and reported. Where complex issues are being assessed in certain studies, such as potential effects on E, A, or T hormonal pathways, it is critical for the Agency to have

detailed information on the methods and data associated with the study. This detailed information will be used to determine the overall adequacy and reliability of the test method (*e.g.,* sufficient experimental group size, appropriate controls, adequate dosing). Studies of good quality are generally represented by those that conform to scientifically acceptable methodology and that sufficiently document both the methods and data. In general, greater confidence in the value of the information contained in a study will come from those submissions conforming to GLP and conducted using peer reviewed test guidelines or those studies meeting other quality assurance or standards. Consequently, inconsistencies or deviations with recommended methodologies would be relevant considerations in any WoE evaluation. Test guidelines or SEPs can provide helpful guidance for gauging the reliability of a study. Studies that use poorly documented or unacceptable methods or that have irreconcilable deficiencies in their design, conduct, or reporting of findings are generally considered of unacceptable quality, and are not considered to provide useful and reliable information.

A study measuring endpoints that are informative to E, A, or T hormonal pathways or an individual study that measures several endpoints showing consistent responses among interrelated endpoints under E, A, or T hormonal influence can provide key lines of contributing evidence in the overall WoE analysis (Section 5.2). Any study that shows an inconsistent pattern of findings among the interrelated endpoints measured within that study without a valid explanation, or a study confounded by variability or other complicating factors such as excessive cytotoxicity would generally not provide useful and reliable evidence that the substance interacts with an E, A, or T hormonal pathway. If deficiencies are found within a Tier 1 assay, on a case-by-case basis, after consideration of all relevant information including the potential contribution of the study or endpoint data to the WoE determination, a request may be made to repeat a Tier 1 assay or conduct a tailored study to address the identified deficiency(ies).

Consideration and characterization of whether the responses are marginal or clearly positive would also provide relevant information, as it may help to discriminate among compounds that are of high concern from those of lower concern for their potential to

interact with the E, A, or T hormonal pathways and, by inference, their potential to produce an adverse effect on human health and ecological populations that may be addressed in Tier 2 testing.

The determination of whether Tier 2 testing is warranted for an individual chemical is based on the totality of the evidence *(i.e.,* analysis of the available EDSP Tier 1 assays in combination with other scientifically relevant published or publically available peer-reviewed studies) as discussed below in Section 5.2.

5.2. Integrating the Different Lines of Evidence

An integrated analysis of the data means that the results from all scientifically relevant published or publically available peer-reviewed studies, which are of sufficient quality and reliability, are evaluated across studies and endpoints into an overall assessment. Determinations of whether or not a chemical has the potential to interact with E, A, or T hormonal pathways and is a candidate for EDSP Tier 2 testing is likely to be primarily based on an integrated analysis of the results of Tier 1 screening, but may also include results from other scientifically relevant information.

In general, the WoE analysis examines multiple lines of evidence and considers the following:

- nature of the effects within and across studies, including number, type, and severity/magnitude of effects,
- conditions under which effects occurred (*e.g.*, dose, route, duration),
- consistency, pattern, range, and interrelationships of effects observed within and among studies, species, strains, and sexes,
- strengths and limitations of the *in vitro* and *in vivo* information, and
- biological plausibility of the potential for an interaction with the E, A, or T hormonal pathways.

In this WoE determination, it is important to consider the biological plausibility of the findings from the different lines of evidence by examining the consistency, coherence, and interrelationships among the measured endpoints within and across studies (*i.e.,* do the majority of studies show findings in the same direction that would be expected for a specific interaction with the E, A, or T hormonal pathways; and is the pattern of effects expected based on the biological understanding of that endocrine MoA?). Because some endocrine MoAs may evoke a number of phenotypic consequences other than those evaluated in the Tier 1 assays, the available toxicity database on the substance may contribute to the WoE evaluation, such as the findings from standard toxicology studies on reproductive effects or tumor responses, which can be associated with hormonal influences. A question to be addressed, therefore, in the WoE analysis, is whether or not the toxicity database is internally consistent with the purported or hypothesized endocrine interaction.

Because the Tier 1 battery was designed to provide complementary endpoint data, a tabular representation of the array of data can be helpful as shown in Table 2. Such a tabular array may aid in the discrimination of consistent effects versus isolated and discordant responses among studies and facilitate the WoE determination.

Table 2. A Suggested Format for Organizing the Individual Lines of Evidence: This example is illustrative rather than comprehensive. The assays included, as well as the organization thereof, depend on the amount of data and on the quality of the studies. This table presents one possible way to organize different lines of evidence indicative of a potential interaction of a compound with the estrogen hormonal pathway. [Key: Positive (P), Negative (N) or Equivocal (E) observation; Arrows (↑↓) indicate the direction of the response; Dashes (--) indicate that parameter was not evaluated.]

Lines of Evidence Indicating Potential Interaction with the Estrogen Pathway (Anti-Estrogenicity)												
Study/ Citation	ER Binding	ER Transactivation	Sex Steroids	Uterine Weight	Ovarian Weight	Ovarian Histopathology	Pituitary Weight	Pituitary Histopathology	Estrous Cyclicity (Age, Length and % of Animals Cycling)	Fertility	Age and Weight at Vaginal Opening	Vitellogenin
Study 1	--	--	↓	↓	↓	--	N	--	--	--	--	--
Study 2	P	N	--	--	--	--	--	--	--	--	--	--
Study 3	--	--	--	--	--	--	--	--	--	--	--	--
Study 4	--	--	E	--	--	N	--	--	--	--	--	↓
Study 5	--	--	--	--	↓	P	N	N	↓	--	↓	--
Study 6	--	--	--	--	--	--	--	--	↓	--	--	--
Study 7	--	--	--	--	↓	--	N	N	E	↓	↓	--
Study 8	--	--	--	--	↓	--	N	N	--	--	--	--
Study 9	--	--	--	--	E	N	N	N	--	--	↓	--

Within the context of the current EDSP Tier 1 battery, results of *in vitro* assays alone would not generally be expected to provide a sufficient basis to support the need for Tier 2 testing. When weighing the different lines of evidence and examining the balance of positive and negative results, EPA expects that *in vivo* evidence would typically be given greater overall influence in the WoE evaluation than *in vitro* findings because of the inherent limitations of such assays. Although *in vitro* assays can provide insight into the MoA, there are some limitations of *in vitro* assays including the inability to account for normal metabolic activation and clearance of the compound, as well as normal intact physiological conditions (*e.g.,* the ability of an animal to compensate for endocrine alterations). The relative sensitivity and specificity of the measured endpoints from the different lines of evidence would also be relevant considerations. It is important to consider and rule out other explanations for the observed results (*e.g.,* secondary consequences of non-endocrine MoA or general toxicity) to the extent possible given the available data. Thus, an important and relevant consideration in the WoE evaluation is to consider the potential for the alternative hypothesis and whether the alternative has evidence in its favor (*i.e.,* Do the data truly reflect an interaction of the chemical with the E, A, or T hormonal pathways?).

Available results from Tier 1 screening together with other sources of scientifically relevant information will span multiple levels of biological organization from molecular, cellular, and tissue or organ effects derived from *in vitro* or *in vivo* systems. The relationship among endpoints at different levels of biological organization and their impact on normal E, A, or T hormonally-mediated processes are important factors in the determination of whether a chemical interacts with these hormonal pathways. Concordant and consistent effects observed among multiple interrelated endpoints reflective of the same interaction with an E, A, or T hormonal pathway can indicate a high potential that the chemical will interfere with endocrine function. In contrast, isolated or discordant effects lower confidence that the chemical would interfere with E, A, or T. For example, a chemical that interacts with the androgen pathway through receptor antagonism could be supported by evidence showing:

- binding to the androgen receptor *in vitro* along with corroborating findings in the Hershberger assay with,
- collaborative measures from androgen sensitive tissues in the *in vivo* mammalian studies (*e.g.,* delayed puberty, testicular histopathology, decrease in epididymis weight, altered testosterone), and
- in fish *in vivo* tests, for example, diminished male secondary sex characteristics, testicular degeneration, and male gonad weight and gonadosomatic index.

Differences in age and body weight, compared to controls, at the time of female sexual maturation and vaginal opening can result from the complex interplay of many factors and can be affected by both endocrine and non-endocrine MoAs. However, a consistent pattern of findings in the female pubertal assay provides supporting evidence of estrogenicity including acceleration of vaginal opening, increases in uterine weight, persistent vaginal expression of estrus and anovulation, and increased blood levels of estradiol along with interaction of the chemical with the estrogen receptor (as measured in the *in vitro* ER binding and transcriptional activation assays). The case for estrogenicity would be further strengthened if evidence was found across taxa, such as increased vitellogenin production and altered secondary sex characteristics in male fish (as measured in the fish short-term reproductive assay).

Data from both the *in vitro* and *in vivo* assays may provide the necessary information to determine whether or not the compound affects steroidogenesis. If hormone production is affected only in the *in vivo* assays along with expected tissue responses and effects on reproductive and developmental processes, with no *in vitro* verification, one possible consideration is the likelihood that the substance impairs hypothalamic-pituitary function, and subsequently alters gonadotropin synthesis or secretion. Other explanations for the reported findings should be considered and ruled out (*e.g.,* secondary consequences of non-endocrine MoA or general toxicity). The examples above illustrate a WoE analysis for a substance producing a consistent pattern of responses within and across different study types and endpoints indicative of endocrine

interactions that could support a recommendation that a chemical be considered as a candidate for Tier 2 testing.

In addition to the EDSP Tier 1 results, consideration of other sources of scientifically relevant information (*e.g.,* 40 CFR Part 158 test guideline studies and published or publicly available peer-reviewed studies) may be appropriate in determining whether or not further testing is needed in Tier 2. Other scientific information relevant to the evaluation may be useful in confirming the results of Tier 1 screening, and may be of special importance if marginal or weak or inconsistent relationships exist within or among the assay results. These data may be useful when considering other explanations for the reported findings (*e.g.,* secondary consequences of non-endocrine MoA or general toxicity). Other supportive evidence that may provide additional insight could include data on the presence or absence of effects that would be anticipated from an impact on normal endocrine function (*e.g.,* certain reproductive or tumor findings). Published or publically available peer-reviewed studies may include data of a similar nature to the Tier 1 assays which may be helpful in interpreting Tier 1 results. Information on related compounds, predictions from computational models (*e.g.,* QSAR) may also help interpret Tier 1 results. Data on metabolism, toxicokinetics, or molecular conservation of the targets being perturbed (*e.g.,* ER, AR, steroidogenic enzymes) may be helpful in evaluating whether species not specifically tested (*e.g.,* wildlife or humans) will respond in a similar manner, given physiological differences and anticipated environmental exposure conditions (*e.g.,* route of exposure).

5.3. Weight-of-Evidence Narrative/Characterization

A summary WoE narrative or characterization generally accompanies the detailed analysis of the individual studies and the integrative analysis of the multiple lines of evidence. Inclusion of a WoE narrative is common in WoE assessments and judgments (USEPA, 2005). The narrative/characterization is intended to be transparent and allow the reader to clearly understand the reasoning behind the conclusions as to whether or not Tier 2 testing is needed. The judgment of whether or not a substance interacts with E, A, or T hormonal pathways should be supported by available data that take into

account the analysis as a whole. The narrative will generally explain the selection of the studies or effects used as the main lines of evidence and relevant basis for conclusions for determining whether or not the test chemical interacts with the E, A, or T hormonal pathways. Essentially, this characterization is intended to capture the most important and relevant information and key conclusions. In general, a summary statement for the WoE analysis would typically be expected to address, among others, the following considerations:

- main lines of evidence for the effect of the test chemical on the E, A, and T hormonal pathways (as well as for each sex and species of concern, if known) including the coverage and selectivity or specificity for interacting with the identified endocrine-mediated process,
- uncertainties and the extent to which these uncertainties impact the conclusions,
- discussion of the studies considered key to the conclusion and why they are considered key,
- a description of inconsistent or conflicting data and whether there is an explanation for these discordant results should be identified (*e.g.,* species, metabolic, dose or route differences),
- conclusions on the need for Tier 2 testing should describe whether available data provided any other possible explanations for the observed results (*e.g.,* secondary consequences of non-endocrine MoAs or general toxicity), and
- what, if any, additional data (*e.g.,* Tier 1, Tier 2 or specialized studies) are needed and why.

The overall strength of the evidence supporting a conclusion (*i.e.,* indicating whether a test chemical has the potential to interact with the E, A, or T hormonal pathways) from the WoE evaluation needs to be described.

5.4. EDSP Tier 2 Testing Recommendations

Given the extensive amount of time, effort, and resources, including the increased animal use required for EDSP Tier 2 testing, the decision to move into Tier 2 needs to

be based on a scientifically sound and robust WoE evaluation for determining whether or not the substance interacts with the E, A, or T hormonal pathways.

- EDSTAC indicated "… *the screening and testing strategy should require the minimal number of screens and tests necessary to make sound decisions, thereby reducing the time needed to make these decisions.*" (EDSTAC, 1998).

EPA generally intends that the need for further evaluating apical endpoints and establishing dose responses would be based on a hypothesis-driven approach (*i.e.,* the understanding of both the chemistry and biology) that takes into account and integrates all existing knowledge about that chemical (Tier 1 results in combination with other available information, including exposure and hazard information) to determine if additional *in vivo* testing to characterize the most likely hazards and risks of concern is necessary. The following are some general considerations relating to determining the need and type of further testing that may be warranted beyond the Tier 1 screening battery.

Further testing would not be supported for compounds with an overall WoE that indicates no potential interaction with the E, A, or T hormonal pathways across multiple lines of evidence. On the other hand, compounds showing positive findings with a consistent pattern of results across multiple lines of evidence would be further considered for Tier 2 testing. This consideration would address what additional data, if any, EPA would need to assess risks arising from the endocrine activity of the chemical. Like the determination of whether a chemical interacts with E, A, or T, EPA would consider multiple data sets and would make a WoE determination regarding additional data requirements. Some suggested questions to consider in deciding whether to require additional testing beyond the Tier 1 screen are:

- For the specific endocrine interaction found in the Tier 1 screen, what apical effects would be anticipated across taxa in standard toxicity guideline studies (including reproductive tests)? Were these apical endpoints adequately measured? Are differences across taxa found or would they be anticipated? Does the existing apical endpoint data in combination with the available Tier 1 *in*

vivo apical data provide a sufficient basis to characterize the apical effects for hazard and risk assessment purposes?

- How can the specific endocrine interaction of interest and endpoints of concern be most efficiently characterized? As an example, if there is evidence in the pubertal assays for a chemical to impact only the mammalian thyroid hormone pathway, the Tier 2 two-generation mammalian reproduction study may not be the most scientifically effective or efficient method to address this concern. A tailored short-term study that includes perinatal exposure and a dose-response evaluation of thyroid axis disruption, for example, could be a more efficient and effective approach.
- If a risk assessment is already available for the compound of interest, what toxicities were used for selecting a point of departure or other features for the risk assessment? Are the results from the Tier 1 screen along with other available information consistent with the chemical's current risk assessment? If so, how likely is it that additional testing would impact the regulatory endpoints?

After considering all relevant hazard and exposure information, the Agency may decide additional testing is not necessary. In other cases, the Agency may determine that additional Tier 2 testing would contribute to the risk assessment of the substance. The totality of information on the substance may indicate that standard or guideline Tier 2 testing is not the most effective method to characterize the relevant apical endpoints or establish dose-response. Therefore, the Agency may recommend a targeted or tailored study that addresses the specific regulatory need.

If Tier 2 testing is indicated, EPA generally expects that the summary narrative and characterization would not only explain the basis of the conclusions regarding potential interactions with the E, A, or T hormonal pathways, but also provide a discussion on why additional data are needed for the chemical's risk assessment. Final recommendations for Tier 2 testing would generally be expected to clearly describe the kind of testing that is appropriate and why specific studies are required. Any suggested

modifications of specific validated Tier 2 test protocols that are tailored to the chemical and specific endpoint(s) of interest should be explained, so they can be evaluated.

6. SUMMARY

In summary, interpretation of the EDSP Tier 1 battery of screening assays and the decision of whether or not to move to Tier 2 testing will be based on the totality of the scientific evidence. The results of the Tier 1 screening assays are likely to be the primary source of data to be considered, along with other scientifically relevant information on the chemical, as appropriate (*e.g.,* 40 CFR Part 158 test guidelines and published or publicly available peer-reviewed studies). WoE analyses rely on professional judgment. The basis of conclusions regarding the potential of a substance to interact with the E, A, or T hormonal pathways should be clearly stated and presented in a transparent manner for those who rely on the analysis. Uncertainties and inconsistencies should be flagged. For any data situation, there are likely to be some uncertainties and inconsistencies. Thus, the extent that these uncertainties or inconsistencies impact the conclusions would be relevant considerations to be explained, as well as whether the evidence is still sufficient to support the endocrine interaction or lack thereof, and overcome the limitations in the database. In some cases, the limitations of the submitted data may be too great, and thus repeating Tier 1 studies or endpoints, or conducting some modification of a Tier 1 assay, or a more targeted study design may be warranted. If the WoE analysis leads to a recommendation that Tier 2 testing is sufficiently supported by all the available evidence, EPA intends to present the rationale for which Tier 2 tests are indicated and why these are critical data for the risk assessment. If a different approach to Tier 2 testing is appropriate, EPA also intends to provide the scientific basis for the modified or tailored approach.

7. REFERENCES

Bogert, C.J., Mihaich, E.M., Quill, T.F., Marty, M.S., Levine, S.L., Becker, R.A. (2011b) Evaluation of EPA's Tier 1 endocrine screening battery and recommendations for improving the interpretation of screening results. Regulatory Toxicology and Pharmacology 59:397-411.

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W. (2006) IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical Reviews in Toxicology. 36(10):781-92.

Boobis, A.R., Doe, J.E., Heinrich-Hirsch, B., Meek, M.E., Munn, S., Ruchirawat, M,. Schlatter, J., Seed, J., Vickers, C. (2008) IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Critical Reviews in Toxicology. 38(2):87-96.

EDSTAC (1998) Endocrine Disruptor Screening and Testing Advisory Committee, Final Report, Volume I-II. Accessed online at http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/finalrpt.htm

Eldridge, J.C., Laws, S.C. (2010) The U.S. EPA's Tier I screening for endocrine disruptor compounds. In: Endocrine Toxicology III Edition. J.C. Eldridge and J.T. Stevens eds. Informa Health Care, N.Y. 1-26.

Hobbs, D.A., Warne, M.St.J., Markich, S.J. (2005) Evaluation of criteria used to assess the quality of aquatic toxicity data. Integrated Environmental Assessment and Management 1:174-180.

Klimisch, H.J., Andreae, M., Tillmann, U. (1997) A systematic approach for evaluating the quality of experimental and toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25:1-5.

National Research Council (NRC). (2007) Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academy of Sciences, Washington, D.C. Accessed on line at http://www.nap.edu/catalog.php?record_id=11970#toc

OECD (2005) Data Evaluation In: Manual for Investigation of HPV Chemicals, Chapter 3. Accessed online at http://www.oecd.org/dataoecd/13/15/36045203.pdf

OMB (2002) Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. (67 FR 8452). Accessed online at http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf

Rhomberg L. (2010) Hypothesis-based weight of evidence: a tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in

proposing a carcinogenic mode of action--naphthalene as an example. Critical Reviews in Toxicology 40:671-696.

SAB/SAP (1999) Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. Accessed online at http://yosemite.epa.gov/sab/sabproduct.nsf/C8ABD410E357DBCF85257193004C42C4/$File/ec13.pdf

SAP (2008) Meeting Minutes of the FIFRA Scientific Advisory Panel to review and consider the Endocrine Disrupter Screening Program (EDSP) Proposed Tier 1 Screening Battery. SAP Minutes No. 2008-03. Accessed online at http://www.epa.gov/scipoly/sap/meetings/2008/march/minutes2008-03-25.pdf

Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung, T., Hoffmann, S. (2009) ToxRTool, a new tool to assess the reliability of toxicological data. Toxicology Letters 189:138-144.

USEPA (1991) Guidelines for Developmental Toxicity Risk Assessment. Federal Register 56(234):63798-63826. Accessed online at http://www.epa.gov/raf/publications/pdfs/DEVTOX.PDF

USEPA (1996) Guidelines for Reproductive Toxicity Risk Assessment. Federal Register 61(212):56274-56322. Accessed online at http://www.epa.gov/raf/publications/guidelines-reproductive-tox-risk-assessment.htm

USEPA (1998) Guidelines for Ecological Risk Assessment, PA/630/R095/002F, Risk Assessment Forum. Accessed online at http://www.epa.gov/raf/publications/pdfs/ECOTXTBX.PDF

USEPA (1999) Guidelines for Carcinogen Risk Assessment, Review Draft, CEA-F-0644, Office of Research and Development. Accessed online at http://cfpub.epa.gov/ncea/raf/cancer.cfm

USEPA (2002a) A Review of the Reference Dose and Reference Concentration Processes. Risk Assessment Forum, Washington DC EPA/630/P-02/002F December 2002 Final Report. Accessed online at http://www.epa.gov/raf/publications/pdfs/rfd-final.pdf

USEPA (2002b) Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency, EPA 260R-02-008, Office of Environmental Information. Accessed online at http://www.epa.gov/quality/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf

USEPA (2003) A summary of general assessment factors for evaluating the quality of scientific and technical information. Science Policy Council, Washington DC. EPA/100/B-03/001. Accessed online at http://www.epa.gov/osa/spc/pdfs/assess2.pdf

USEPA (2005) Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum, Washington, DC. EPA/630/P-03/001F. Accessed online at http://www.epa.gov/raf/publications/pdfs/CANCER_GUIDELINES_FINAL_3-25-05.PDF

USEPA (2007) Validation of Screening and Testing Assays Proposed for the EDSP. Version 5.4, July, 2007. Accessed online at http://www.epa.gov/scipoly/oscpendo/pubs/edsp_validation_paper_v%205.4.pdf

USEPA (2009a) EPA's Approach for Considering Other Scientifically Relevant Information (OSRI) Under the Endocrine Disruptor Screening Program. Accessed at Federal Docket Management System, EPA-HQ-OPPT-2007-1080-0032

USEPA (2009b) Exposure Factors Handbook: 2009 Update. External Review Draft. EPA/600/R-09/052A. Access online at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=209866#Download

USEPA (2011a) Corrections and Clarifications on Technical Aspects of the Test Guidelines for the Endocrine Disruptor Screening Program Tier 1 Assays (OCSPP Test Guideline Series 890). Accessed online at http://www.epa.gov/scipoly/oscpendo/pubs/assayvalidation/clarificationdoc.pdf