Endocrine disruption: Where are we now: Tier 2 testing

Mary Ann Ottinger [Corresp., 1], **Kevin Flynn** [2], **Sig Degitz** [3], **Tim Verslycke** [4], **Hank Krueger** [5], **Leslie Touart** [6], **Allen Olmstead** [7], **Anne Goumelon** [8]

[1] Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA
[2] USEPA NHEERL MED, Duluth, Minnesota, USA
[3] USEPA, Grantsburg, Wisconsin, USA
[4] Gradient, Cambridge, Massachusetts, USA
[5] Wildlife International Ltd., Easton, Maryland, USA
[6] Equiparent Consulting, Woodbridge, Virginia, USA
[7] Bayer CropScience, Research Triangle Park, North Carolina, USA
[8] Organization for Economic Cooperation and Development, Paris, France

Corresponding Author: Mary Ann Ottinger
Email address: maotting@uh.edu

The focus of this session was on the proposed USEPA Endocrine Disruption Screening Program (EDSP) Tier 2 testing protocols. Tier 2 tests have been developed to evaluate the potential impacts of endocrine disrupting chemicals (EDCs) over the life cycle across organisms representing vertebrate and invertebrate classes. Key aspects of these Tier 2 testing protocols rely on selecting appropriate measurement end points to reveal differential sensitivity and adverse impacts across an organism's life stages. To this end, certain Tier 2 tests utilize a multigenerational protocol, which detect both short- and long-term effects. However, multigenerational testing protocols can be time consuming and costly. As such, other testing protocols have also been considered, including partial life-cycle and extended one-generation tests. Regardless of the specifics of the multigenerational protocol, it is critical to identify key measurement end points that are responsive, reliable, and repeatable indicators of exposure to endocrine disrupting chemicals; these measures should also provide information to enable initial assessments of risk translated from individual to potential population level effects across a variety of living organisms. Presentations in Session three of the Society of Environmental Toxicology and Chemistry (SETAC) North America Focused Topic Meeting: Endocrine Disruption (February 4 – 6, 2014) focused on the current state of the science for EPA EDSP Tier 2 testing. Presentations in this session considered the strengths and weaknesses of the Tier 2 assays across several classes of organisms, and provided an industry perspective on Tier 2 testing. The interactive panel discussion provided an interesting perspective that balanced regulatory needs for reliable testing protocols that are highly repeatable and utilize consistent indices of exposure and adverse effect.

Title: Endocrine Disruption: Where Are We Now: Tier 2 Testing

Running head: Extended and multi-generation testing protocols for evaluating EDCs

Mary Ann Ottinger[†*], Kevin Flynn[‡], Sig Degitz[§], Tim Verslycke[ǁ], Hank Krueger[#], Leslie Touart[††], Allen Olmstead[‡‡], and Anne Goumelon[§§1]

[†]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204 USA, maotting@central.uh.edu, ; [‡] USEPA NHEERL MED, Duluth, MN 55804-2595; [§] USEPA, Grantsburg, Wisconsin; [ǁ]Gradient, 20 University Road, Cambridge, MA 02138; [#]Wildlife International Ltd., 8598 Commerce Drive, Easton, MD 21601, hkrueger@eag.com; [††] Equiparent Consulting, Woodbridge, VA 22191, email: les.touart@equiparentconsulting.com, [‡‡] Bayer CropScience, Research Triangle Park, NC 27709, allen.olmstead@bayer.com [§§] Organization for Economic Cooperation and Development (OECD) and rue André-Pascal, 75775 Paris Cédex 16-France, Email : anne.gourmelon@oecd.org

[1]This text reflects the views of the author and does not necessarily represent the official position of the OECD and its member countries.

* Corresponding author:

Mary Ann Ottinger

University of Houston, Houston, TX 77204 USA

Email address: maotting@central.uh.edu.

Corresponding Author

ABSTRACT

The focus of this session was on the proposed USEPA Endocrine Disruption Screening Program (EDSP) Tier 2 testing protocols. Tier 2 tests have been developed to evaluate the potential impacts of endocrine disrupting chemicals (EDCs) over the life cycle across organisms representing vertebrate and invertebrate classes. Key aspects of these Tier 2 testing protocols rely on selecting appropriate measurement end points to reveal differential sensitivity and adverse impacts across an organism's life stages. To this end, certain Tier 2 tests utilize a multigenerational protocol, which detect both short- and long-term effects. However, multigenerational testing protocols can be time consuming and costly. As such, other testing protocols have also been considered, including partial life-cycle and extended one-generation tests. Regardless of the specifics of the multigenerational protocol, it is critical to identify key measurement end points that are responsive, reliable, and repeatable indicators of exposure to endocrine disrupting chemicals; these measures should also provide information to enable initial assessments of risk translated from individual to potential population level effects across a variety of living organisms. Presentations in Session three of the Society of Environmental Toxicology and Chemistry (SETAC) North America Focused Topic Meeting: Endocrine Disruption (February 4 – 6, 2014) focused on the current state of the science for EPA EDSP Tier 2 testing. Presentations in this session considered the strengths and weaknesses of the Tier 2 assays across several classes of organisms, and provided an industry perspective on Tier 2 testing. The interactive panel discussion provided an interesting perspective that balanced regulatory needs for reliable testing protocols that are highly repeatable and utilize consistent indices of exposure and adverse effect.

Key words: Endocrine disrupting chemicals, multi-generation tests, Endocrine Disrupter Screening Program, Tier 2.

INTRODUCTION

There are a number of components that constitute the USEPA Endocrine Disruption Screening Program (EDSP) that, subsequent to priority setting candidate chemicals, include screening and testing programs to be implemented through Tier 1 and Tier 2 testing protocols. The goal of Session Three of the Society of Environmental Toxicology and Chemistry (SETAC) North America Focused Topic Meeting: Endocrine Disruption (February 4 – 6, 2014) was to provide an overview of the strengths and weaknesses of the Tier 2 test protocols across a range of species and classes of organisms; and importantly to integrate the industry perspective into the conduct and efficacy of these testing protocols to assess endocrine disrupting compounds (EDCs). The session was chaired by Gary Ankley and Mary Ann Ottinger and included eight presentations and a panel and audience discussion. Leslie Touart (subsequently retired from USEPA) overviewed the status of the Tier 2 tests; Kevin Flynn provided insight into the status of the Tier 2 Medaka Extended One-Generation Reproduction Test (MEOGRT). Sig Degitz discussed the development for the Larval Amphibian Growth and Development Assay (LAGDA) protocol; and Tim Verslycke provided an overview of the validation of the Mysid Two-Generation Toxicity Test for EDCs. Lastly, Mary Ann Ottinger overviewed the Tier 2 Japanese Quail (*Coturnix japonica*) Toxicity Test. These presentations provided a basis for a series of presentations in which Anne Goumelon addressed the OECD perspective; Allen Olmstead discussed the Tier 2 EDSP Assays Viewed Through the Lens of Ecological Risk Assessment; Hank Krueger presented the Contract Laboratory Perspective on Higher-Tier Endocrine Testing; and finally a Panel and Audience Discussion was held with the speakers and Ed Perkins.

SESSION PRESENTATION SUMMARIES

USEPA's Endocrine Disruptor Screening Program Tier 2 Ecotoxicology Test Methods

by Leslie Touart[1]

USEPA established the Endocrine Disruptor Screening Program (EDSP) in response to a US Congressional mandate "to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or such other effects as USEPA may designate" (21 U.S.C. 346a(p)) (USEPA 2011). As part of the EDSP, USEPA is validating assays to identify and characterize the endocrine activity of pesticides, commercial chemicals, and environmental contaminants, specifically in relation to estrogen, androgen, and thyroid hormones. This talk presented a brief historical summary of the development and validation of the candidate test methods including a mammalian two-generation test, a Japanese quail two-generation test, the Larval Amphibian Growth and Development Assay (LAGDA), a medaka multi-generation test, and an invertebrate test. Although a medaka multi-generation test was the principal fish method considered, an abbreviated medaka reproduction (extended one-generation) test was also proposed. Additionally, a mysid two generation toxicity test is recommended as the preferred invertebrate *in vivo* Tier 2 EDSP test, but a harpacticoid copepod reproduction and development test was also considered as a potential alternative or option. The reasoning and judgments leading to the various studies that were conducted as part of the development, demonstration, and validation of the various test methods was discussed. In addition, the outcome and recommendations of a FIFRA SAP review (USEPA 2013) (www.epa.gov/scipoly/sap/meetings/2013/june/062513minutes.pdf) of the proposed methods and public comments of the revised methods were summarized and discussed. The current status of the final test guidelines, at the time, was presented.

1 DISCLAIMER: The opinions presented are those of the author and may not reflect EPA policies.

The Proposed Tier 2 Medaka Extended One-Generation Reproduction Test (MEOGRT)

By: Kevin Flynn

The Medaka Extended One Generation Reproduction Test (MEOGRT) has been proposed as part of the Tier 2 testing within the USEPA EDSP. As part of definitive Tier 2 testing, the MEOGRT should determine whether a substance adversely affects a test organism through endocrine-mediated pathways, and to quantitatively evaluate those effects incorporating exposure during the most sensitive life stages and provide the opportunity for identification of dose-response effects. The MEOGRT characterizes the nature, likelihood, and dose-response relationship of apical adverse outcomes from potential endocrine disruption via estrogenic, androgenic, and possibly thyroid pathways. In general, to meet the goals of an EDSP Tier 2 test protocol, the MEOGRT encompasses all the life stages of at least one full generation (F1) including effects on fertility and mating, embryonic development, sensitive neonatal growth and development, and transformation from the juvenile life stage to sexual maturity. In addition, a substantial exposure time is called for in the F0 generation that starts the exposure phase of the test as adults to allow for loading of the gametes with the chemical of interest to account for possible maternal transfer of chemical. Lastly, a continued exposure into an additional generation (F2) is allowed if adequate information is present to suggest the possibility of different sensitivity or the manifestation of different effects between successive equivalent generations.

Summary timeline, replication and sampling information was presented for the proposed MEOGRT design. Briefly, the typical test done with the MEOGRT protocol without the additional F2 generation would last 19 weeks (4 weeks of F0 exposure; 15 weeks of F1 exposure) with samples taken at 9 weeks post-fertilization (sub-adult lifestage) and 15 weeks post-fertilization (adult lifestage). A reproductive assessment is done during post-fertilization weeks 12 – 14. The MEOGRT protocol as a proposed part of the USEPA EDSP Tier 2 testing strategy is

anticipated to enter a public commenting period[2]. It is possible that in response to comments received both from internal and external stakeholders, the USEPA may change aspects of the protocol that would not have been presented at the SETAC North America Focused Topic Meeting (FTM) on Endocrine Disruption: Chemical Testing and Risk Assessment Approaches and Implications. However, it is not anticipated that fundamental changes to the structure of the test, for instance, starting exposure with adult F0 and continuing through a complete F2 generation, would occur after the commenting period.

Information that was used to arrive at the proposed replicate structure was summarized. The MEOGRT has a 2:1 replication design: twice as many control replicates as each exposure replicate. For most of the test, there are 12 control replicates and 6 replicates in each of five exposure levels; however, during the reproductive assessment, the replication doubles so there are 24 control replicates and 12 replicates in each of the five exposure levels. A power analysis based upon Monte Carlo simulation of fecundity data was done that provided the necessary information to make recommendations regarding replicate structure within the MEOGRT (Figure 1). Note that at 12 control replicates/6 exposure replicates per treatment, there is a small but noticeable probability of not detecting a reduction of 50%, about a 75% probability at detecting a reduction of 40%, a less than 50% chance at detecting a reduction of 30%, and a very little chance at detecting a reduction of 20%. At 24 control replicates/12 exposure replicates per treatment, the probability of not detecting a reduction of 40% or greater is near zero, and there is probability of greater than 80% of detecting a reduction as low 30%. During discussions on replicate structure of the MEOGRT, consideration was given not only to the power analysis, but also to the possibility of mortality, especially in the control replicates. It has been our experience that a very small percentage of the adults, irrespective of treatment, may die, and in addition,

2 Note from the Guest Editor: Since the Focused Topic Meeting was held, the USEPA EDSP Tier 2 MEOGRT Guideline has been finalised (USEPA 2014a).

even with skilled technicians, there is a possibility of handling-induced mortalities as well. To be conservative, 12 breeding pairs (replicates) in treatments and 24 control breeding pairs (replicates) was chosen to mitigate the consequences to statistical power.

The MEOGRT provides data about the primary apical endpoint of reproduction, as well as the toxicity endpoints of growth, hatch, survival, and liver pathology, and finally data providing insight into adverse outcome pathways (secondary sexual characteristics, vitellogenin gene expression and gonad pathology). This data is either a ratio, ordinal, or continuous in nature. Typical control values, the expected minimum and maximum values, and the proposed acceptance criteria are presented for each of the endpoints specified in the MEOGRT below (Table 1). Data from future MEOGRTs that fail to meet the acceptance criteria put the validity of the individual test at risk by potentially reducing power to unacceptable levels or loss of entire exposure levels.

Based upon the molecular initiating event of an adverse outcome pathway, a certain pattern of responses in the above endpoints might be expected (Table 2). These data expectations provide a potential means to identify the adverse outcome pathway(s) that an unknown EDC activates to produce a negative biological impact. While there are substantial data gaps for various adverse outcome pathways, the expected outcomes based upon the molecular initiating event are presented in Table 2.

In conclusion, the presentation of the MEOGRT protocol at the SETAC FTM on Endocrine Disruption was intended to provide a summary of the protocol, rationale for the proposed replication structure, typical output data from the protocol, and the impacts on the measured endpoints based upon molecular initiating event. We also assert that the MEOGRT protocol fulfills the EDSTAC-defined purpose of a Tier 2 test in that it 1) includes endpoints to assess whether a test substance adversely affects a test organism through endocrine-mediated

pathways, 2) includes exposure during life stages that may potentially be more sensitive than those tested in Tier 1, 3) Includes potential effects of parental transfer of chemical and other endogenous factors with exposure during gametogenesis, and 4) characterizes the dose-response.

Development of the Larval Amphibian Growth and Development (LAGDA), by: Sigmund Degitz

The Food Quality Protection Act of 1996 requires EPA to develop and implement a program using valid tests for determining the potential endocrine effects from pesticides. The EPA established advisory group, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC, USEPA 1997), recommended EPA develop a two-tiered approach: Tier 1 would identify the potential of a substance to interact with the endocrine system whereas Tier 2 would further identify and characterize chemical-induced interactions with estrogen, androgen and thyroid hormones for risk assessment to inform regulatory decisions. One of the Tier 2 tests recommended by EDSTAC is an amphibian full life cycle test to evaluate the adverse consequences of putative endocrine disrupting chemicals, especially those active within the hypothalamic-pituitary-thyroid (HPT) and hypothalamic-pituitary-gonadal (HPG) systems, on the development, growth and reproduction of amphibians (specifically the frog). The LAGDA is intended to serve as a higher tier test with an amphibian for collecting definitive concentration-response information on adverse effects suitable for use in ecological risk assessment. Specifically, the design enables the collection of amphibian hormone-regulated endpoint data (*e.g.*, metamorphosis, gonadal development) and information concerning various aspects of the reproductive biology and life-stage viability.

The LAGDA protocol describes a chronic toxicity test with an amphibian species that considers growth and development from fertilization through the early juvenile period[3]. It also enables measurement of a suite of other endpoints that allows for diagnostic evaluation of endocrine disrupting chemicals or other types of developmental and reproductive toxicants. The LAGDA is a relatively long-term assay (normally 130 days or longer) that assesses early development, growth, and partial reproductive maturation. The test is designed to detect both endocrine and non-endocrine mechanisms by including diagnostic endpoints specific to key endocrine mechanisms. It should be noted that prior to development of the LAGDA, no validated assay existed which could serves this function for amphibians.

The general experimental design entails exposing Nieuwkoop Faber (NF) stage 8 *Xenopus laevis* embryos to four different concentrations of a test chemical and a control until 10 weeks after the median time to completion of metamorphosis (NF stage 62) in the control with one interim sub-sample at NF stage 62 (See Nieuwkoop and Faber 1994 for staging details). There are four replicates in each test concentration with eight replicates for the control. Endpoints evaluated during the course of the exposure include those indicative of generalized toxicity, *i.e.*, mortality, abnormal behavior, and growth determinations (length and weight), as well as endpoints designed to characterize specific endocrine toxicity modes of action targeting estrogen (E)-, androgen (A)-, or thyroid (T)-mediated pathways.

During standardization and optimization, studies were performed geared toward refining, optimizing, and standardizing the protocol, and initially assessing protocol transferability and performance. Individual and inter-laboratory evaluations of the LADGA were conducted to evaluate the practical transferability of the assay protocol and quantitative reproducibility of the results. The inter-laboratory validation evaluated the ability of four labs to conduct and evaluate

3 Note from the Guest Editor: Since the Focused Topic Meeting was held, the final USEPA EDSP Tier 2 LAGDA Guideline has been finalised (USEPA 2014b).

196 the LAGDA assay (USEPA, 2013). The following chemicals were evaluated across individual or
197 multiple laboratories: prochloraz (aromatase inhibitor, AR agonist), 4-*tert*-octylphenol (ER
198 agonist), 17-β trenbolone (AR agonist), and benzophenone-2 (ER agonist, TPO inhibitor).
199 Prochloraz was tested in four labs, and 4-*tert*-octylphenol was tested in three labs. Trenbolone
200 and benzophenone-2 were tested in single laboratories and these studies serve to demonstrate the
201 responsiveness of the LAGDA to additional modes of action.

202 The LAGDA proved to be an effective test model. All four chemicals produced
203 endocrine-related effects. Of the two chemicals available for inter-laboratory comparison,
204 prochloraz resulted in thyroid gland pathologies consistent with a hypothyroid condition in 3 of
205 the 4 labs, and vitellogenin (VTG) induction and gonad/reproductive duct pathologies were noted
206 in all 4 laboratories. The second chemical, 4-*tert*-octylphenol, produced thyroid gland
207 pathologies consistent with a hypothyroid condition and delayed development in only 1 of the 3
208 laboratories. However, VTG production and mild gonad/reproductive duct pathologies were
209 observed in all laboratory studies. 17-β Trenbolone and benzopehone-2, although only tested in
210 single laboratories, produced endocrine-related effects involving the thyroid gland, delayed
211 metamorphosis, VTG production and reproductive tract pathologies.

212 ***Validation of the Mysid Two-Generation Toxicity Test for the Regulatory Testing of Endocrine***
213 ***Active Compounds, by Tim Verslycke***

214 This presentation provided a summary of the validation results for the mysid two-
215 generation toxicity test (MTTT) which is being proposed as a Tier 2 invertebrate assay in
216 USEPA's EDSP. Full validation results for the MTTT as well as the harpacticoid copepod
217 development and reproduction test (HCDRT), which was evaluated as a potential alternative to

the MTTT, are presented in the Integrated Summary Report (EPA-HQ-OPP-2013-0182-0007, USEPA 2013).

Invertebrates comprise 95% of the world's animal species (Wilson 1988), and certainly a larger percentage of the Earth's total animal abundance. Many invertebrate toxicity test protocols are routinely used in regulatory testing; however, few have been designed with endocrine-specific endpoints in mind. Although many aspects of invertebrate physiology and life cycle are known to be under endocrine control, the hormones produced and used by invertebrates are not directly analogous to those of vertebrates. For example, crustaceans and other ecdysozoans account for more than 75% of all known animal species, yet they rely largely on invertebrate-specific ecdysteroid and juvenile hormones to regulate their physiology (Chang 1993; deFur et al. 1999; Subramoniam 2000). On the other hand, crustaceans have true endocrine glands derived from epithelial tissue and functioning similar to vertebrate glands (deFur 2004) and their endocrine systems are relatively well understood compared to those of other invertebrates (Oehlmann and Schulte-Oehlmann 2003; LeBlanc 2007). Given that endocrine disruption has been reported in crustaceans (OECD 2006), an invertebrate test method that uses crustaceans for evaluating potential effects of endocrine disrupting chemicals (EDCs) is relevant.

Mysid crustaceans have been used in regulatory (and other) toxicity testing for more than 30 years and standard testing protocols have been developed for several species. Beyond certain insect growth regulators (IGRs), there have been few direct links between potential EDCs and endocrine disruption in mysids. Still, mysids have the ecological relevance and sensitivity to stressors required of a taxon that would be suitable for evaluation of endocrine disruption in marine and estuarine invertebrates and could serve as a surrogate for other crustacean species. Further, the proposed test species, *Americamysis bahia*, has widespread availability, is relatively easy to culture, has a short life cycle (17-20 days), and has been widely used in toxicity testing.

Finally, our knowledge of hormone regulation in mysids continues to grow and several EDC-related endpoints in mysids have been proposed over the last decade (Verslycke *et al*. 2004, 2007; Ghekiere *et al*. 2005; 2006a; 2006b; 2007; Yokota *et al.* 2011).

McKenney (2005) first demonstrated transgenerational effects in *A. bahia* using a two-generation exposure protocol. These studies led to the development of the proposed Tier 2 invertebrate assay. The proposed MTTT is a relatively long-term assay (normally 60 days or longer) that assesses early development, growth, and reproduction in two generations. It is an extension of existing standard practice for conducting a mysid life-cycle test (ASTM 2004; McKenney 1986, 1998; and Nimmo et al. 1977, 1978) and is intended to serve as a higher tier test with an aquatic arthropod for collecting definitive concentration-response information on adverse effects suitable for use in ecological risk assessment. The MTTT guideline includes 25 different endpoints (8 growth, 11 reproduction, and 9 survival endpoints), some are recorded per mysid or composite of mysids, some are recorded per breeding pair, and some are recorded per replicate tank.

The MTTT guideline was used in demonstration and optimization studies using a number of endocrine-active chemicals (fenoxycarb, 3,5-dichlorophenol, fipronil, prochloraz, flutamide, ketaconazole, 4-*tert*-octylphenol, lindane, atrazine, perfluorodecanoic acid) in two different laboratories. Subsequently, the MTTT guideline was used in an inter-laboratory validation study with three participating laboratories and using three endocrine-active chemicals (lindane, vinclozolin, 4-*tert*-octylphenol). Two out of the three laboratories were able to successfully execute the draft method. Large inter-laboratory and intra-laboratory variability was observed in the control endpoint responses. Significant differences were also observed in lab proficiency as estimated by the variability in the endpoint responses in the control groups, indicating difficulties in the transferability of the MTTT between laboratories. Further, where the same chemical was

tested in two laboratories, dose-response relationships were generally not consistent (based on comparisons of significantly affected endpoints at each treatment level) among the laboratories.

A number of strengths of the MTTT were highlighted during the validation studies. Laboratories have established experience with *A. bahia* and were generally able to successfully perform the MTTT within the recommended acceptability criteria. Further, the MTTT can be conducted in continuous or intermittent flow ensuring consistent water quality and chemical exposure concentrations. Also, control variability for several endpoints indicates that these should be able to detect significant adverse effects with adequate statistical power. Finally, common population modeling approaches can be employed based on the data obtained in the MTTT to estimate population-level effects (Raimondo and McKenney 2005) and several mechanistic endpoints (*e.g.*, vitellin and hormone levels, hormone receptor expression) could be added to the MTTT to allow for the collection of mechanistic data.

Similarly, a number of limitations of the MTTT were highlighted during the validation studies. Considerable variability was observed in endpoint responses between different laboratories, resulting in reduced power to detect significant differences. Some of the endpoints were consistently non-responsive (*e.g.*, sex ratio, time to maturation), and their value may need to be evaluated further. There was a lack of treatment-related responses and responses were inconsistent between laboratories. Several aspects of the MTTT are time-consuming and resource demanding. Specifically, the addition of a second generation significantly adds to the time and resources required to perform the MTTT and the value of the additional information obtained from the second generation was not obvious. Finally, the appropriateness and adequacy of the current Tier 1 screen for identifying chemicals that may interfere with invertebrate hormone axes or the endpoints measured in the MTTT remains unclear since Tier 1 screening is

focused on identifying chemicals that may interfere with the vertebrate estrogen, androgen, or thyroid hormone axes.

A number of further refinements of the MTTT are suggested. For example, the selected chemicals in the validation studies were representative endocrine active chemicals in vertebrates, but may not exhibit endocrine toxicity in arthropods. Further validation with known arthropod endocrine active chemicals (*e.g.*, insect growth regulators, ecdysone agonists, etc.) may provide better data for evaluating strengths and limitations of the proposed MTTT. Such testing could lead to the identification of a positive control chemical to be used for the MTTT. Given the amount of endpoints included in the MTTT and the inter-relatedness of the endpoints, endpoints could be reduced and refined which will augment replication and statistical power. Additional guidance on statistical evaluation of endpoint data and conducting range-finding experiments for dose selection for the MTTT would also be beneficial.

While this presentation was focused on the MTTT, comparative strengths and limitations of the HCDRT were also discussed. Based on this comparison, the MTTT was recommended as the preferred Tier 2 test for several reasons, including the extensive experience of contract laboratories with mysids, the availability of an extensive mysid ecotoxicology database, a greater understanding of mysid biology, easier achievement of chemical exposure concentrations, greater body size of mysids, and well-established growth endpoints (which appeared to be among the most sensitive endpoints in the MTTT).

The Tier 2 Japanese Quail (Coturnix japonica) Avian Toxicity Test, by: Mary Ann Ottinger

The purpose of the Tier 2 Japanese Quail (*Coturnix japonica*) Avian Toxicity Test is to detect both short and long term impacts from exposure to Endocrine Disrupting Chemicals (EDCs).

There are compelling reasons for having either an extended one-generation or a multigenerational avian protocol in the EDC testing schedule. Birds have a number of unique characteristics that may predispose them to adverse impacts from EDCs as listed below. In addition, EDCs are readily transferable into the egg where they concentrate into the yolk if they are lipophilic or the albumin for water soluble compounds.

- Metabolic systems
  - High body temperature (105ºF) with high metabolic rate
  - Associated accelerated toxicokinetics
  - Migratory associated energy drain and mobilization of lipid reserves
  - Thyroid system function in precocial and altricial birds
- Reproductive system
  - Female has one functional ovary
  - Altered gonadal differentiation results in ovotestes
- Hormones and behavior
  - Males adversely impacted by xenoestrogens and androgenic compounds
  - Female behavior (receptivity) less sensitive to xenoestrogens
- Sexual differentiation
  - Males are the homogametic sex having ZZ; females are ZW
  - HPG axis relies on relative exposure to estradiol and testosterone
    - Males—primary exposure to testosterone
    - Females—primary exposure to estradiol
  - HPG axis and song system differ in precocial and altricial birds
- Growth and migration
  - High metabolism requires sufficient nutrient utilization
  - Rapid growth rate especially for migratory species
  - Thyroid system critical for pre-migratory fattening
- Lifespan
  - Long-lived birds produce few offspring annually over many years.

In addition, there is an argument for retaining all the proposed generations. Core endpoints are survival, growth, and reproduction. All have multilayers of associated endpoints that reflect toxicity, direct effects of chemicals on organ systems, and impacts on reproductive, metabolic/thyroid systems, and adrenal/stress axes. The P0 (parent generation) provides maturation and adult responses to exposure. Although these phases of the life cycle are potentially less vulnerable, it is important to assess effects on adults, especially relative to adverse

effects on reproductive and metabolic endocrine function. Moreover, because the Japanese quail has the same neuroendocrine circuitry regulating reproduction as other avian species, any impact would be translatable to field birds. Behavioral impacts observed in Japanese quail would be indicative of potential greater impacts on songbirds because the song control system is steroid dependent as are the neural systems that modulate singing behavior. Because the F1 (first generation) birds are exposed both via maternal deposition and from the diet (same treatments as their parents), they would be impacted by endocrine disruption during embryonic development and sexual differentiation as well as experiencing any impacts due to endocrine disruption during activation of reproduction during maturation and in adults. Finally, the F2 (second generation) is exposed to maternally deposited EDCs, thereby exhibiting effects of endocrine disruption during embryonic development. As such, the importance of the F2 generation is to reveal potential transgenerational effects and isolate embryonic effects of EDCs.

The Japanese quail is a precocial bird that has advantages for a multigenerational testing protocol because this species is relatively domesticated, rapidly maturing, easily maintained in the laboratory, and is a well-characterized avian model. Studies have been conducted to inform the design of an avian two-generation testing protocol and to ascertain key measurement endpoints that provide reliable indicators of EDC exposure. These studies have included egg injection and several types of dietary studies that have considered a range of compounds. Egg injection studies take advantage of avian embryonic development in the egg, independent of parental input. As such, egg injection studies mimic maternal deposition of chemicals, providing an opportunity to dose the embryo with known concentrations of compound and track effects throughout ontogeny. These studies have shown impacts of EDCs on reproductive and metabolic endocrine systems, behavior, and heart function, especially with exposure during embryonic development (Ottinger et al 2005; 2009; Ottinger and Dean, 2011). Findings from comparison of

existing studies reveal that many EDCs do impact avian species in support of observations of wild populations (Rattner et al, 2004). These studies also emphasize the unique characteristics of avian species, which must be considered by a testing protocol, including high body temperature, migration associated energy demands, precocial and altricial birds, high metabolic rate, and mechanisms and role of steroid hormones in sexual differentiation. In addition, potential sources of variability occur due to strain differences and between species relative to sensitivity to EDCs. Analyzing core endpoints of survival, growth, and reproduction across generations will reveal potential impacts on reproductive, metabolic/thyroid systems, and adrenal/stress axes as well as general toxicity. It is important to assess measurement end points reflective of neural mechanisms regulating reproductive endocrine function and behavioral response, metabolic and stress axis function, and functional measures indicative of adverse physiological outcomes. Future applications will use these data to assess risk across the wide range of breeding strategies and diversity of life histories with consideration of sensitivity and period(s) of vulnerability in order to protect avian populations. [4]

Status of OECD work on the Development of Harmonized Test Methods for Endocrine Disrupters, by: Anne Gourmelon

The protection of human health and the environment from endocrine disrupters is currently a high priority for regulatory authorities in most OECD countries/regions, and it has been proposed by UNEP as a SAICM[5] policy emerging issue. Indeed, the OECD Test Guidelines Program has spent approximately half of its resources since 1996 to develop Test Guidelines and other tools to support countries' needs related to testing and assessment of chemicals for

4 Note from the Guest Editor: Since the Focused Topic Meeting was held, the Avian Two-Generation Toxicity Test in the Japanese Quail has been finalised (USEPA 2014c).

5 SAICM is the Strategic Approach to International Chemicals Management, see www.saicm.org

endocrine disruption.These works have been made possible because some OECD countries like the United States had a dedicated program, a clear goal and resources to undertake the validation studies. OECD Test Guidelines are standardized, validated and harmonized test methods used across countries adhering to the Mutual Acceptance of Data, a government-to-government agreement aiming at reducing non-tariff barriers to trade and avoiding duplicative testing. After more than 15 years working on the validation and development of methods for screening and testing chemicals for endocrine disruption, a Conceptual Framework functioning as a toolbox has been developed and refined; more than 10 OECD Test Guidelines specific to ED have been validated and adopted; a large number of guidance and reviews documents have been published, and validation reports and workshop reports have been agreed and published in support of the Test Guidelines. A few long-term test methods for wildlife species are still under discussion and the OECD is keen on developing harmonized methods for these important and resource-intensive assays. In looking towards the future, OECD countries are conscious of the need to assess more chemicals more efficiently. OECD is providing a forum to discuss and harmonize ways to integrate new technologies and novel approaches in the testing and assessment of ED, based on knowledge of the modes of action leading to adverse outcomes.

The work on endocrine disrupters testing and assessment is overseen by the Working Group of National Coordinators of the Test Guidelines Programme (WNT) and managed by four main expert groups:

- An advisory group on endocrine disrupters testing and assessment (EDTA AG)
- A validation management group on ecotoxicity testing
- A validation management group on non-animal testing

- A validation management group for mammalian testing

The EDTA AG is an advisory group to the WNT and to the VMGs. National experts nominated by the National Coordinators and the European Commission, and representatives from the Business and Industry Advisory Committee, Environmental NGOs, and International Council on Animal Protection in OECD Programmes participate in the work.

After more than 10 years working on the validation and development of methods for screening and testing chemicals for endocrine disruption, the *Workshop on OECD Countries' Activities Regarding Testing, Assessment and Management of Endocrine Disrupters* (OECD, 2010), held in September 2009 in Copenhagen, recommended further work for OECD, and in particular (i) the development of a guidance document for the assessment of endocrine disrupters, (ii) the revision of the 2002 Conceptual Framework for Testing and Assessment of Endocrine Disrupters, and (iii) the development of a detailed review paper on endpoints that are not included in existing Test Guidelines. In parallel with the continuous development of Test Guidelines for the screening and testing of endocrine disrupters, other documents recommended by the Copenhagen workshop have been developed (see below).

A Conceptual Framework (CF) for the Testing and Assessment of Endocrine Disrupters was adopted in 2002. The CF is not a testing strategy; it is not prescriptive and simply reflects the type of information the tests provide at the different levels, such as informing endocrine toxicity outcome pathways, moving from *in silico* to *in vitro* and *in vivo*. It should be noted that information on mechanisms/pathways is particularly important for assessing chemicals for endocrine disruption. An updated CF was approved by the WNT in April 2012. It includes all published Test Guidelines listed in Table 3 of this document; test methods for which inclusion in the Test Guidelines work plan has been approved by the WNT (Table 4); some existing Test Guidelines not specifically developed for screening/testing of chemicals for endocrine disruption

(Table 5), and a few non OECD test methods. The updated CF is attached as an annex to the *Guidance Document on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disrupters* (OECD, 2012a). The revised description of the five levels of the draft CF is as follows:

- Level 1. Existing data and non test information
- Level 2. In vitro assays providing data about selected endocrine mechanism(s)/pathway(s)
- Level 3. In vivo assays providing data about selected endocrine mechanism(s)/pathway(s)
- Level 4. In vivo assays providing data on adverse effects on endocrine relevant endpoints
- Level 5. In vivo assays providing more comprehensive data on adverse effects on endocrine relevant endpoints over extensive parts of the life cycle of the organisms.

Information/tools from lower levels can be used to determine what specific higher level tests are needed for a specific chemical to increase evidence that it is/it is not an endocrine disrupter. This approach is illustrated in the *Guidance Document on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption (OECD, 2012a); guidance document on standardized test guidelines for evaluating chemicals for endocrine disruption*

The Guidance Document No. 150 in the OECD Series on Testing and Assessment (OECD, 2012a) was developed to support regulatory authorities' decisions related to the hazard of specific chemicals and toxicologically-relevant metabolites when they receive test results from a Test Guideline or draft Test Guideline for the screening/testing of chemicals for endocrine disruption. The guidance is worded to permit flexible interpretation in the context of different domestic legislation, policies and practice. It also provides guidance on how to interpret the outcome of individual tests, taking into account existing information, and how to increase

462 evidence on whether or not a substance may be an endocrine disrupter. It recommends test
463 methods that may be performed if regulatory authorities need more evidence. The test methods
464 are defined precisely so that countries' possible testing requirements can be harmonized and
465 hence ensure the Mutual Acceptance of Data.

466 The project to develop the Detailed Review Paper on the State of Science on Novel *in*
467 *vitro* and *in vivo* Screening and Testing Methods and Endpoints for Evaluating Endocrine
468 Disrupters (OECD, 2012b) was led by the United States, in cooperation with the European
469 Commission. To date, OECD work related to endocrine disrupters focused on oestrogen/androgen
470 and thyroid pathways. However, other endocrine and neuro-endocrine pathways may also have
471 adverse outcomes, such as symptoms of metabolic syndrome, reproductive dysfunction, altered
472 fetal development.

473 A number of Test Guidelines have been published in 2007-2012[6] and are available free of
474 charge from the OECD ilibrary (http://www.oecd-ilibrary.org/content/package/chem_guide_pkg-
475 en). The work plan of the Test Guideline Programme includes projects for other Test Guidelines
476 for screening/testing chemicals for endocrine disruption.

477 ***Tier 2 EDSP Assays Viewed Through the Lens of Ecological Risk Assessment, by: Allen***
478 ***Olmstead***

479 Ecological risk assessment is the process through which the likelihood that adverse effects
480 in the environment occur due to a stressor. Generally for chemical substances, assessments are
481 made at the level of the individual on processes of survival, growth, and reproduction with the

9 6 Note from the Guest Editor: Since the Focused Topic Meeting was held, OECD guideline 240 of the Medaka
10 Extended One Generation Medaka Reproduction Test (MEOGRT) and OECD guideline 241 of the Larval Amphibian
11 Growth and Development Test (LAGDA) have been finalized (OECD 2015a, 2015b)

assumption that by safeguarding these, populations would be protected from adverse effects. Endocrine toxicity represents one of many means through which these processes may be affected. While currently a large battery of ecotoxicology tests is employed to assess the hazard, these are not tailored specifically to endocrine toxicity. The EDSP should evaluate hazard that is not covered by current testing. The Tier 2 EDSP assays should be evaluated based on what additional hazard information that has a meaningful impact on ecological risk assessment is generated beyond that from current test guidelines. Further, the endpoints measured in these test should be optimized with respect to their utility in evaluating ecological risk.

The Contract Lab Perspective on Higher Tier Endocrine Tests Part 2, by: Hank Krueger

Higher tiered endocrine testing will be conducted in Contract Laboratories that will be challenged by the size and complexity of these studies. I would like to thank many contributors that expressed opinions and provided comments in preparing for this presentation. Their contribution represents many years of experience in the contract laboratory environment. Translating the concepts of Tier 2 testing into reality provide many practical challenges that have not been thoroughly discussed or incorporated into guidelines. Among the challenges is the selection of test concentrations, physical constraints on our ability to achieve test concentration in test systems, finding ways to fill data gaps to have the necessary information for the design of Tier 2 tests, ways to improve Tier 2 tests, and managing projects with higher degrees of complexity.

When it comes to selecting test concentrations several issues need to be addressed. The first is to determine the range and spacing of concentrations. Knowing how high to test becomes critical, because it is desirable to be testing at levels that are free from the effects of general toxicity. Testing at concentrations that require separating classical effects of toxicity from

endocrine effects should be avoided. While it is desirable to be testing at a maximum tolerated dose, defining that dose and achieving it experimentally can be difficult. Tier 1 testing used the criteria for setting the highest test concentration as 100 mg/L, the water solubility limit, or 1/3 the LC50 as an estimate of the maximum tolerated dose. While these high concentrations represent extreme levels, they may be very different from relevant environmental concentrations. This leads to a more general question of setting concentrations that are environmentally relevant for risk assessment or setting them to determine hazard. There is also the concern of low dose effects, which means more guidance will need to be provided to labs on how to set and space test concentrations.

There are physical limitations as to what can be done in laboratories. The chemicals that were chosen to develop the tests and then used in the test validation process were well studied and in most cases had desirable physical properties and modes of action. The chemicals selected for the validation of the Medaka 2-Gen study all were very soluble and for the most part were easy to deliver since the concentrations were well over 100 times their solubility limit. However, to test at concentrations near the solubility limit, the volumes of stock solutions that need to be prepared for testing become limiting. Other physical constraints on spacing of test concentrations occur for materials with very low solubility. In some cases, the distance between the limit of solubility and the analytical limit of quantitation (LOQ) may be too small to accommodate the desired range of concentrations.

Test systems used in aquatic toxicity tests have been designed to deliver concentrated stocks to mixing chambers where a clean source of dilution water (well water) is mixed with the stock ideally at a ratio of 1: ≤100 of stock solution to dilution water. Testing at water solubility limit means there is no dilution and that the highest test concentration receives nothing but the water stock prepared at the solubility limit. The worst case scenario for endocrine testing is the

mysid 2-generation test which would consume 1600 L of stock per day. If the highest test concentration was 100 times lower than the solubility limit, then one would could prepare a water stock at the solubility limit and then dilute it 100 fold which would result in the consumption of only 16 L of stock per day, a much more manageable volume.

While a stock is being used on a test it must be stable. If a material degrades as a result of hydrolysis or photolysis, or is lost to the system through volatility or adsorption while the test is being conducted, then the use of a water stock is limited. This is one of the key reasons solvents have been used in aquatic toxicology. One can prepare a concentrated stock in solvent that is both stable and concentrated. There are guideline limits on how much solvent can be used in a test, with limits of 0.1 ml/L for general testing and 0.02 ml/L for endocrine tests resulting in 10,000 and 50,000 fold dilutions of stock, respectively. While solvents should be avoided they still can have a role when testing at concentrations near the solubility limit or when testing materials that are not stable in water stocks.

Contract labs are also concerned about the limited amounts of data that may be available when asked to conduct a Tier 2 test. Data gaps will exist that will need to be filled to design a larger scale test. Information from Tier 1 testing is limited. Tier 1 tests do not look at liver and kidney histology which are new endpoints for the Tier 2 test. How does one address these endpoints when selecting test concentrations? Designing higher tier endocrine tests involves taxonomic leaps of faith extrapolating data from rodent assays to fish, frogs, and birds; from fathead minnow to medaka, from mallard and bobwhite quail to Japanese quail, and from receptor to whole organism.

The Tier 1 Fish Short Term Reproduction Assay (FSTRA) is a 21-day assay using adult fathead minnows. While the study provides data on adults, data on sexual development from earlier stages of development are missing. Knacker et al (2010) demonstrates that for most

553 modes of action for endocrine disruption in zebrafish, sexual development is the most sensitive
554 stage to look for effects. Such data gaps demonstrate the need for pilot studies that are smaller
555 and more focused experiments that provide additional information for designing a higher tiered
556 test. Pilot studies may also incorporate newer techniques that provide better data and may
557 provide enough information so that the Tier 2 test is not needed.

558 Lastly, endocrine testing raises new concerns as to how we manage studies in contract
559 laboratories. Ron Biever of Smithers Viscient points out that in the past all that was needed in a
560 project team was a chemist to evaluate exposure by measuring concentrations of test substances
561 in water and a biologist to serve as a study director to oversee a test from start to finish, interpret
562 results, and author a report. Endocrine studies have added a level of complexity that requires a
563 more complex project management structure and have redefined the role of the study director. In
564 addition to a biologist and chemist, other members of the project team now include individuals
565 that specialize in the measurement and evaluation of biomarkers that include vitellogenin,
566 steroids, determination of genetic sex, gene expression and histopathology. A statistician is also
567 needed on the project team with all these additional endpoints and rigorous statistical analyses
568 being required in the guidelines. The study director's new role is to integrate all these disciplines
569 into one report and that will require very knowledgeable and experienced individuals.

570 **CONCLUSIONS**

571 Multigenerational tests provide critical information about the potential for impact by
572 endocrine active compounds or EDCs over the life cycle across vertebrates and invertebrates.
573 These Tier 2 testing protocols rely on selecting appropriate measurement end points to reveal
574 differential sensitivity and adverse impacts across an organism's life stages. Further, it is

important to understand life stages that are most sensitive or vulnerable to the effects these environmental contaminants. It has become clear that traditional methods of assessing potential risk and impact to an individual or population may not reveal EDC associated adverse effects. As more is known about the timing and sensitivity of organisms to suspected EDCs, a suite of targeted measurement end points as part of an extended one generation or multigenerational test will augment estimated toxicity from measures such as toxic equivalency quotient (TEQ) or toxic equivalency factor (TEF). Finally, Tier 2 tests will detect both short- and long-term effects as well as other potentially long-term effects from epigenetic change. However, multigenerational testing protocols can be time consuming and costly. As such, other testing protocols have also been considered, including extending the one-generation test. Regardless of the specifics of the multigenerational protocol, it is critical to identify key measurement end points that are responsive, reliable, and repeatable indicators of exposure to endocrine disrupting chemicals; these measures should also provide information to enable initial assessments of risk translated from individual to potential population level effects across a variety of living organisms.

ACKNOWLEDGMENTS

The authors wish to sincerely thank our collaborators; you all have been critical to the research and scientific conclusions expressed in this manuscript. Our colleagues include the following: Dean Hammermeister, Doug Lothenbach, Frank Whiteman, Rodney Johnson, J.T. Haselman, David B. Mayfield, Vince Brown, David Houchens, Patience Browne, Michael Quinn, Emma Lavoie, Karen Dean, Meredith Bohannon, Tim Springer.

DISCLAIMER

This text reflects the views of the authors and does not necessarily represent the official position of the OECD and its member countries. In addition, this text reflects the views of the authors and does not necessarily represent the official position of the EPA.

REFERENCES

[ASTM] American Society for Testing and Materials. 2004. Standard Guide for Conducting Renewal Microplate-Based Life-Cycle Toxicity Tests with a Marine Meiobenthic Copepod. ASTM Standard No. E2317-04. ASTM, Philadelphia, pp. 1-16.

Chang ES. 1993. Comparative Endocrinology of Molting and Reproduction: Insects and Crustaceans. *Annual Reviews of Entomology*. 38:161-180.

deFur PL. 2004. Use and role of invertebrate models in endocrine disruptor research and testing. ILAR journal/National Research Council, *Institute of Laboratory Animal Resources*. 45(4), 484-493.

deFur PL, Crane M, Ingershold C, Tattersfield L. 1999. Endocrine Disruption in Invertebrates: Endocrinology, Testing and Assessment. Society of Environmental Toxicology and Chemistry, Pensacola, FL, USA.

Ghekiere A, Fenske M, Verslycke T, Tyler C, Janssen CR. 2005. Development of a quantitative enzyme-linked immunosorbent assay to study vitellogenesis in the mysid *Neomysis integer* (Crustacea: Mysidacea*). Comp. Biochem. Physiol*. A. 142(1):43–49.

Ghekiere A, Verslycke T, and Janssen C. 2006a. Effects of methoprene, nonylphenol, and estrone on the vitellogenesis of the mysid *Neomysis integer*. *General and Comparative Endocrinology*. 147(2):190-195.

Ghekiere A, Verslycke T, Fockedey N, Janssen CR. 2006b. Non-target effects of the insecticide methoprene on molting in the estuarine crustacean Neomysis integer (Crustacea: Mysidacea). *Journal of Experimental Marine Biology and Ecology*. 332(2):226-234.

620 Ghekiere A, Fockedey N, Verslycke T, Vincx M, Janssen, CR. 2007. Marsupial development in
621 the mysid *Neomysis integer* (Crustacea: Mysidacea) to evaluate the effects of endocrine-
622 disrupting chemicals. *Ecotoxicology and Environmental Safety*. 66(1):9-15.

623 McKenney, CL Jr. 1986. Influence of the organophosphate insecticide fenthion on *Mysidopsis*
624 *bahia* exposed during a complete life cycle I. Survival, reproduction, and age-specific
625 growth. *Dis Aquat Organ*. 1:131–139.

626 McKenney CL Jr. 1998. Physiological dysfunction in estuarine mysids and larval decapods with
627 chronic pesticide exposure. In: Wells PG, Lee K, Blaise C (eds) Microscale testing in
628 aquatic toxicology: advances, techniques, and practice. CRC Press, Boca Raton, FL,
629 USA, pp 465–476.

630 McKenney CL Jr. 2005. The influence of insect juvenile hormone agonists on metamorphosis and
631 reproduction in estuarine crustaceans. *Integr. Comp. Biol*. 45:97–105.

632 Knacker T, Boettcher M, Frische T, Rufli H, Stolzenberg HC, Teigeler M, Zok S, Braunbeck
633 T, Schäfers C. 2010. Environmental effect assessment for sexual endocrine-disrupting
634 chemicals: Fish testing strategy. *Integr Environ Assess Manag*. 6(4):653-62.

635 Nieuwkoop PD, Faber J. 1994. Normal Table of *Xenopus laevis* (Daudin). Garland Publishing
636 Inc., New York, N.Y.

637 Nimmo DR, Bahner LH, Rigby RA, Sheppard JM, Wilson AJ Jr. 1977. *Mysidopsis bahia*: An
638 Estuarine Species Suitable for Life-Cycle Toxicity Tests to Determine the Effects of a
639 Pollutant. In: Aquatic Toxicology and Hazard Evaluation, Mayer F.L. and Hamelink, J.L.
640 eds. Pp. 109-116. American Society for Testing and Materials. Philadelphia, PA.

Nimmo DR, Hamaker TL, Sommers CA. 1978. Entire Life-Cycle Toxicity Test using Mysids (*Mysidopsis bahia*) in Flowing Water. EPA-600/9-78-010. U.S. Environmental Protection Agency, Environmental Research Laboratory, Gulf Breeze, FL. (ERL, GB X107).

[OECD] Organization for Economic Cooperation and Development. 2006. Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects. Organisation for Economic Co-Operation and Development Series on Testing and Assessment Number 55. Environment Directorate, Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology. ENV/JM/MONO(2006)22. JT03212400. 125 pages. Available at: http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2006)22&doclanguage=en.

[OECD] Organization for Economic Cooperation and Development. 2010. Workshop Report on OECD Countries Activities Regarding Testing, Assessment and Management of Endocrine Disrupters, Series on Testing and Assessment No. 118, OECD, Paris.

[OECD] Organization for Economic Cooperation and Development. 2012a. Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, Series on Testing and Assessment No. 150, OECD, Paris.

[OECD]. Organization for Economic Cooperation and Development. 2012b. Detailed Review Paper on the State of the Science on Novel *In Vitro* and *In Vivo* Screening and Testing

663 Methods and Endpoints for Evaluating Endocrine Disruptors, Series on Testing and
664 Assessment No. 178, OECD, Paris.

665 [OECD]Organization for Economic Cooperation and Development) i-library: (http://www.oecd-
666 ilibrary.org/content/package/chem_guide_pkg-en)

667 [OECD]. Organization for Economic Cooperation and Development). 2014a. New Guidance
668 Document on Harpacticoid Copepod Development and Reproduction Test with
669 Amphiascus. Series on Testing and Assessment No. 201, OECD, Paris.

670 [OECD] Organization for Economic Cooperation and Development. 2014b. New Scoping
671 Document on *in vitro* and *ex vivo* Assays for the Identification of Modulators of Thyroid
672 Hormone Signalling. Series on Testing and Assessment No. 207, OECD, Paris.

673 [OECD] Organization for Economic Cooperation and Development. 2015a. Test No. 240:
674 Medaka Extended One Generation Reproduction Test (MEOGRT), OECD Publishing,
675 Paris. [Cited 2016-09-08] Available from **http://www.oecd-**
676 **ilibrary.org/content/package/chem_guide_pkg-en**

677 [OECD] Organization for Economic Cooperation and Development. 2015b. Test No. 241: Larval
678 Amphibian Growth and Development Test (LAGDA), OECD Publishing, Paris. [Cited
679 2016-09-08] Available from **http://www.oecd-**
680 **ilibrary.org/content/package/chem_guide_pkg-en**

681 Oehlmann J, Schulte-Oehlmann U. 2003. Endocrine Disruption in Invertebrates. *Pure and*
682 *Applied Chemistry* 75(11-12):2207-2218.

683 Ottinger MA, Dean KM. 2011. Neuroendocrine impacts of endocrine-disrupting chemicals in
684 birds: life stage and species sensitivities. *Journal of Toxicology and Environmental*
685 *Health Part B, Critical reviews* 14:413-422.

686 Ottinger MA, Lavoie ET, Abdelnabi M, Quinn MJ Jr., Marcell A, Dean K. 2009a. An overview of
687 dioxin-like compounds, PCB, and pesticide exposures associated with sexual
688 differentiation of neuroendocrine systems, fluctuating asymmetry, and behavioral effects
689 in birds. *Journal of Environmental Science and Health Part C, Environmental*
690 *Carcinogenesis & Ecotoxicology Reviews* 27:286-300.

691 Ottinger MA, Quinn Jr MJ, Lavoie E, Abdelnabi MA, Thompson N, Hazelton JL, et al. 2005.
692 Consequences of endocrine disrupting chemicals on reproductive endocrine function in
693 birds: Establishing reliable end points of exposure. *Domestic Animal Endocrinology*
694 29:411-419.

695 Raimondo S, McKenney CL. 2005. Projecting population-level responses of mysids exposed to
696 an endocrine disrupting chemical. *Integrative and Comparative Biology*. 45(1):151-157.

697 Rattner BA, McGowan PC, Golden NH, Hatfield JS, Toschik PC, Lukei RF, Jr., et al. 2004.
698 Contaminant exposure and reproductive success of ospreys (*Pandion haliaetus*) nesting
699 in Chesapeake Bay regions of concern. *Archives of Environmental Contamination and*
700 *Toxicology* 47:126-140.

701 Subramoniam T. 2000. Crustacean ecdysteroids in reproduction and embryogenesis.
702 *Comparative Biochemistry and Physiology C*. 125:135-156.

703 [USEPA] U.S.- Environmental Protection Agency: Endocrine Disruptor Screening Program
704 Advisory Committee (EDSTAC) (1997). EDSTAC final report volume 1 and 2.

705 [USEPA] US Environmental Protection Agency. 2011. Endocrine Screening Program. Available
706 from: http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/development.htm

707 [USEPA] US Environmental Protection Agency .2013. Validation of the Larval Amphibian
708 Growth and Development Assay - Integrated Summary Report, EPA-HQ-OPP-2013-
709 0182-0006.

710 [USEPA] US Environmental Protection Agency .2013. Validation of the Mysid Two-Generation
711 Toxicity Test and Harpacticoid Copepod Reproduction and Development Test -
712 Integrated Summary Report, EPA-HQ-OPP-2013-0182-0007

713 [USEPA] US Environmental Protection Agency. 2013. Science Advisory Panels
714 www.epa.gov/scipoly/sap/ meetings/2013/june/062513minutes.pdf

715 [USEPA] US Environmental Protection Agency. 2014a. Endocrine Disruptor Screening Program
716 Test Guideline OCSPP 890.2200, Medaka Extended One-Generation Test (MEOGRT).
717 EPA-HQ-OPPT-2014-0766-0021.

718 [USEPA] US Environmental Protection Agency. 2014b. Endocrine Disruptor Screening Program
719 Test Guideline OCSPP 890.2300, Larval Amphibian Growth and Development Test
720 (LAGDA). EPA-HQ-OPPT-2014-0766-0020

721 [USEPA] US Environmental Protection Agency. 2014c. Endocrine Disruptor Screening Program
722 Test Guideline OCSPP 890.2100, Avian Two-Generation Toxicity Test in the Japanese
723 Quail. EPA-HQ-OPPT-2014-0766-0019

724 Verslycke T, Fockedey N, McKenney CL, Roast SD, Jones MB, Mees J, Janssen CR. 2004.
725 Mysid crustaceans as potential test organisms for the evalution of environmental
726 endocrine disruption: a review. *Environ Toxicol Chem*. 23(5):1219–1234.

727 Verslycke T, Ghekiere A, Raimondo S, Janssen C. 2007. Mysid crustaceans as standard models
728 for the screening and testing of endocrine-disrupting chemicals. *Ecotoxicology.* 16: 205-
729 219.

730 Wilson, EO (ed.). 1988. Biodiversity. National Academy Press. Washington DC.

731 Yokota H, Eguchi S, Nakai M. 2011. Development of an in vitro binding assay for ecdysone
732 receptor of mysid shrimp (*Americamysis bahia*). *Aquatic Toxicology*. 105(3-4):708-716.

733 Figure 1. The effect of the number of replicates on power at several levels of reduction of
734 fecundity. Simulated power is on the y-axis and number of pairs in a single exposure level is on
735 the x-axis.

736 Table 1: Medaka Extended One Generation Reproduction Test Endpoints and Proposed

737 Acceptance Criteria

Endpoint	Mean	Range	Acceptance Criteria	Comments
Fecundity	23 eggs/pair-day	0 to 50	>20	SD = 9
Fertility	82%	0 to 100%	>75%	Occasional low fertility
Anal fin papillae (subadult)	Male: 70	0 to 120	>20	Occasional 0
	Female: 0	0	Near 0	Occasional XX male
Vitellogenin	Male: $2x10^4$	$1x10^2$-$2x10^4$	Female mean at least 200 x male	
	Female: $5x10^6$	$3x10^4$-$1x10^8$		Occasional XX male
Hatch	80%	0 to 100%	>70%	Occasional 0%
Weight	Male: 470 mg	330 to 690	> 170 mg	SD = 97
	Female: 570 mg	420 to 750	> 195 mg	SD = 70
Survival (ELS)	80%	0 to 100%	>65%	

738 Table 2: Potential Means to Identify Adverse Outcome Pathways for Endocrine Disrupting
739 Effects

MIE*	Male				Female				Reproduction	
	Vtg	SSC	Gonad	Growth	Vtg	SSC	Gonad	Growth	Fecund.	Fertility
ER agonist	↑	↓	sex reversal intersex	–	(↑)	–	–	–	↓	–
AR agonist	–	(↑)	–	–	↓	↑	sex reversal perifollicular hyper/hypo	–	↓	–
AR antagonist	–	(↓)	?	–	–	–	–	–	↓	–
Steroidogenesis inhibitor	–	–	?	–	↓	–	Perifollicular hyper/hypo	–	↓	–
Toxicity	↓		hypoplastic	↓			Hypoplastic	↓	↓	?

740 MIE = Molecular Initiating Event; ER = Estrogen Receptor; AR = Androgen Receptor; vtg =
741 vitellogenin, SSC = Secondary Sex Characteristics

742 Table 3: Published Test Guidelines Specifically Developed or Updated for the Screening or
743 Testing of Chemicals for Endocrine Disruption

TG	Title	Adoption year
440	Uterotrophic Bioassay in rodents: A short-term Screening Assay for Oestrogenic Properties	2007
407 (updated)	Repeated Dose 28-day Oral Toxicity Study in Rodents	2008
211 (updated)	Daphnia Magna Reproduction Test	2011
441	Hershberger Bioassay in rats: A Short-Term Screening Assay for (Anti)Androgenic Properties	2009
229	Fish Short Term Reproduction Assay	2009
230	21-Day Fish Assay: A Short-Term Screening for Oestrogenic and Androgenic Activity, and Aromatase Inhibition	2009
231	Amphibian Metamorphosis Assay	2009
455	Stably Transfected Human Oestrogen Receptor-α Transcriptional Activation Assay for the Detection of Oestrogenic Agonist Activity of Chemicals	2009
234	Fish Sexual Development Test	2011
443	Extended One –Generation Reproductive Toxicity Study	2011
456	H295R Steroidogenesis Assay	2011
457	BG1Luc Estrogen Receptor Transactivation in vitro Assay to Detect Estrogen Receptor Agonists and Antagonists	2012

744 Table 4: Projects for the Screening or Testing of Chemicals for Endocrine Disruption, Currently
745 on the Work Plan

Project	Lead Country
Fish Life-Cycle Test/Medaka Multi-Generation Test	USA/JPN
Larval Amphibian Growth and Development Assay	USA/JPN
Xenopus Embryonnic Thyroid Assay	FRA
Zebrafish Embryo Assay for the detection of endocrine active substances acting through the estrogen receptor	FRA
Copepod Reproduction and Development Test, published (OECD, 2014a)	SWE
Mollusc Reproductive Toxicity Tests – Development and Validation of Test Guidelines	DEU/GBR/FRA/DNK
Avian 2-Generation Reproductive Toxicity Assay	USA
Human Recombinant Oestrogen Receptor Alpha Binding Assay	USA/EC/DEU/JPN
STTA Assay for the Detection of Androgenic and Anti-Androgenic Activity	JPN
STTA Assay for the detection of Anti-Oestrogenic activity of chemicals	JPN
Performance-Based Test Guideline for the Androgen Receptor Transactivation Assay	EC
Transcriptional Assay for the Detection of Estrogenic and Anti-Estrogenic Compounds using MELN Cells	EC
Thyroid Scoping Document, published (OECD, 2014b)	OECD Secr.
Update of TG 421 and TG 422 with ED-relevant endpoints	DK

746 A number of existing Test Guidelines may also provide useful information for the
747 assessment of endocrine disrupters. They are available free of charge from the ilibrary
748 (http://www.oecd-ilibrary.org/content/package/chem_guide_pkg-en).

749 Table 5: Adopted Test Guidelines That May Provide Useful Information, Although Not
750 Specifically Developed for Screening/Testing Chemicals for Endocrine Disruption

Name	TG Number	Year of Adoption
One-Generation Reproduction Toxicity Study	TG 415	1983
Two-Generation Reproduction Toxicity	TG 416	2001
Reproduction/Developmental Toxicity Screening Test	TG 421	1995
Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test	TG 422	1996
Carcinogenicity and Reproductive Toxicity Studies	TG 451-453	2009
Prenatal Development Toxicity Study	TG 414	2001
Repeated Dose 90-Day Oral Toxicity Study in Rodents	TG 408	1998
Development Neurotoxicity Study	TG 426	2007
Avian Reproduction	TG 206	1984
Chironomid Toxicity Test	TG 218-219	2004
Sediment-Water Chironomid Life-Cycle Toxicity Test Using Spiked Water or Spiked Sediment	TG 233	2010