Validation of the Medaka Multigeneration Test: Integrated Summary Report

U.S. Environmental Protection Agency
Endocrine Disruptor Screening Program
Washington, D.C.

28 May 2013

Contents

EXECUTIVE SUMMARY

This Integrated Summary Report (ISR) provides the validation record of the EDSP Tier 2 Medaka Multigeneration Test (MMT) aimed at characterizing potential disruption of the endocrine system by putative Endocrine Disrupting Chemicals (EDCs). Pertinent literature on fish chronic toxicity tests has previously been summarized in an in-depth review (Appendix B). Relevant studies completed as part of the validation process for the MMT are summarized that helped inform the ultimate design of the recommended protocol. Studies include tests with estrogen receptor agonists (17β-estradiol, 4-*t*-octylphenol, *o*,*p*'-DDT,4-chloro-3-methylphenol, and tamoxifen), an androgen receptor agonist and antagonist (trenbolone and vinclozolin, respectively), and a steroid synthesis modulator (prochloraz).

The research MMT protocol describes a 2½ generation toxicity test with a small freshwater fish that evaluates reproductive fitness as an integrated measure of toxicant effects across 2½ consecutive generations using pairs of adult medaka. It also enables measurement of a suite of other endpoints for diagnostic and quantitative evaluation of EDCs or other types of reproductive toxicants. The MMT is a relatively long-term assay (29 weeks of in-life exposure) that assesses hatch, growth, survival, gonadal development, and reproduction in 2½generations (F_0, F_1, and part of F_2). Vinclozolin was chosen for an interlaboratory demonstration of the protocol.

The research objectives were to: a) develop a multigeneration test protocol using the Japanese medaka that could be used to evaluate the effects of endocrine disruption chemicals (EDCs) on reproduction and reproductive development; b) to include diagnostic or intermediate effects endpoints that could be evaluated for associations with the apical outcomes to help elucidate various EDC-induced adverse outcome pathways (AOPs); c) address the question of the relative importance of an additional F_2 generation; and d) develop a scientifically-defensible Tier 2 fish test protocol for characterizing concentration-response for ED endpoints.

The MMT method was successfully performed in characterizing adverse effects and concentration-response for several known endocrine active substances. In the case of 4-*t*-octylphenol, tests at different laboratories and at different times in one laboratory comported very well. However, in the interlaboratory trial with vinclozolin, the comparative analysis indicated that for many endpoints, the control response was not consistent between laboratories. The median responses from F_1 reproductive measurements, F_1, F_2 length and weight measurements were all significantly different ($p < 0.01$) between laboratories. Closer inspection of the results suggest some of the significant inter-lab variability can be attributed to the relatively small size at termination of F_1 and F_2 fish in a single laboratory. The small size of this laboratory F_1 females appears to have contributed to the high inter-lab variability in fecundity. Normalizing F_1 fecundity for the body mass of the female eliminated the significant difference. Overall the interlaboratory can be considered comparable insomuch as all three labs did not see decreased reproduction at the concentrations of vinclozolin tested. This is also unfortunate for the very same reason. Other endpoints were also not significantly affected by vinclozolin with the exception of SSC which was negatively responsive to vinclozolin exposure in all labs that recorded this endpoint. SSC was also decreased in the vinclozolin MMT performed by the US EPA/ORD/MED. So while few significant effects were observed in the three interlaboratory MMTs, there was consistency in the observation of effects on the SSC.

Unfortunately, the previously discussed issues with the interlaboratory data cannot simply be ignored, and they highlight that this is a relatively complicated procedure that even with experience is not a trivial undertaking. Many of the directives within the protocol are derived from extensive experience at the US EPA/ORD/MED lab, and without *a priori* knowledge, should be followed without deviation.

The outcomes of the various MMT trials have provided enough information to recommend a medaka reproduction test (MRT) for use as the fish test in Tier 2 of the EDSP. Two major changes from the MMT are proposed, *i.e.*, an increase in the number of replicates per treatment for evaluating effects on reproduction, and terminating the test after the embryos hatch in F_2. Other proposed changes include minimizing the collection of endpoint data from F_0, and evaluating pathology in only the F_1 adults sampled after the assessment of reproduction.

Overall, the authors conclude that both the MMT and the MRT are transferable methods and are capable of adequately characterizing potential disruption of the endocrine system by putative endocrine disrupting chemicals. However, the MRT is recommended as the preferred EDSP Tier 2 test method for fish because it is less resource intensive with improved statistical power, appears to be as sensitive, and is better able to ensure consistent findings when performed routinely by testing laboratories.

OBJECTIVE

The objective of this integrated summary report (ISR) is to provide a detailed account of the validation process for the Medaka Multigeneration Test (MMT) so that Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) members may address the validation charge questions provided.

1 INTRODUCTION

Section 408(p) of the Federal Food Drug and Cosmetic Act (FFDCA) requires the U.S. Environmental Protection Agency (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)].*

Subsequent to passage of the Food Quality Protection Act in 1996, which amended FFDCA and FIFRA, the EPA formed the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), a committee of scientists and stakeholders that was charged by the EPA to provide recommendations on how to implement its Endocrine Disruptor Screening Program (EDSP). The EDSP is described in detail at the following website: http://www.epa.gov/scipoly/oscpendo/. Upon recommendations from the EDSTAC (EDSTAC 1998), the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormonal systems as well as the endocrine systems of wildlife. Following broader international concerns and the creation of similar programs in other countries, the Organization for Economic Co-operation and Development (OECD) established the Endocrine Disrupters Testing and Assessment (EDTA) Task Force in 1998 within its Test Guidelines Program. EDTA is charged with developing an internationally harmonized testing strategy for the screening and testing of endocrine disrupting chemicals, taking into account the consequences of such a testing strategy on the development and validation of Test Guidelines, and on existing regulatory systems for new and existing substances.

The MMT is one of the assays proposed as part of the EDSP Tier 2 and is the subject of this report. This report, along with the detailed information provided in the attached supporting materials introduces the purpose of the MMT, the scientific rationale for the assay, the development and optimization of the assay protocol, as well as how it fits into the larger EDSP. It also synthesizes the information gained during the validation process and addresses the advantages and limitations of this fish test method based on its, practicality, reproducibility, reliability, protocol transferability and fit for purpose.

1.1 Endocrine Disruptor Screening Program

To comply with its mandate, the EPA chartered a Federal advisory committee (EDSTAC) to provide advice and guidance on the development of a screening and testing program. In 1998, it recommended to the EPA a conceptual two-tiered approach that involved screening and testing chemical compounds for effects on the estrogen (E), androgen (A), and thyroid (T) hormone axes collectively referred to as EAT. The ultimate goal of these assays is to provide input into hazard identification to assess risk of adverse consequences to humans and wildlife (EDSTAC, 1998).

The EPA submitted a proposal of the EDSP for public review and comment in a Federal Register Notice (FRN, 1998) as well as peer review by a joint subcommittee of the EPA Science Advisory Board (SAB) and FIFRA Scientific Advisory Panel (SAB-SAP, 1999). A complete description of the program proposal can be found in the FRN. Briefly, the EDSP proposed by EDSTAC allows for: 1) initial sorting and prioritization of chemical compounds, 2) identification of chemicals for further testing using a Tier 1 screening battery that includes *in vitro* and *in vivo* mammalian, amphibian and fish assays, and 3) characterization of adverse consequences resulting from possible endocrine disruption and establishment of dose-response relationships for hazard identification using Tier 2 testing.

In comparison to the more refined, detailed, and definitive tests in Tier 2, the EDSTAC indicated that the *in vitro* and *in vivo* screening assays in the Tier 1 battery should:

- be relatively fast and efficient;
- be standardized and validated;
- be more sensitive than specific to minimize false negatives without an unreasonable rate of false positives;
- be comprised of multiple endpoints that reflect as many modes of endocrine action as possible;
- have a sufficient range of taxonomic groups among test organisms represented; and
- yield data that can be interpreted as either negative or positive for determining the necessity and manner in which to conduct Tier 2 tests.

Together, the suite of Tier 1 assays form a battery in which some endocrine axis redundancy is incorporated (*e.g.*. two different assays may cover some similar aspects of the estrogenic response). This redundancy allows for a weight-of-evidence (WOE) approach, as recommended by EDSTAC, to determine whether a chemical undergoes further, more definitive, testing.

The assays in Table 1-1, recommended by EDSTAC, have undergone validation and have been adopted as the EDSP Tier 1 screening battery (Table 1-1). These assays are meant to detect chemicals that may affect the E, AE, A and TT hormone axes.

Table 1-1. Assays recommended for consideration for the Tier 1 screening battery.

Assay	Reason for Inclusion
Estrogen receptor binding or transcriptional activation assay	An *in vitro* test to detect chemicals that may affect the endocrine system by binding to the estrogen receptor.
Estrogen receptor transcriptional activation assay	An *in vitro* test to detect chemicals that may affect the endocrine system by binding to the estrogen receptor.
Androgen receptor binding assay	An *in vitro* test to detect chemicals that may affect the endocrine system by binding to the androgen receptor.
In vitro steroidogenesis assay	An *in vitro* test to detect chemicals that interfere with the synthesis of the sex steroid hormones
Placental Aromatase Assay	An assay to detect interference with aromatase.
Uterotrophic Assay	An *in vivo* assay to detect estrogenic chemicals.
Hershberger Assay	An *in vivo* assay to detect androgenic and antiandrogenic chemicals.
Pubertal Male	An *in vivo* assay to detect chemicals that act on androgen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system. This assay could in part substitute for the female pubertal assay.
Pubertal female assay	An *in vivo* assay to detect chemicals that act on estrogen or through the HPG axis that controls the estrogen and androgen hormone systems. It is also enhanced to detect chemicals that interfere with the thyroid system.
Amphibian metamorphosis assay	An *in vivo* assay for detection of chemicals that interfere with the thyroid hormone system.
Fish screening assay	An *in vivo* assay for detection chemicals that interfere with the HPG axes.

A weight of the evidence (WOE) review of the Tier 1 screening battery and other scientifically relevant information (OSRI), will be made to ascertain if a chemical is deemed sufficiently endocrine active to prompt requirements for Tier 2 testing. The following tests, as generally recommended by EDSTAC, are being considered as part of the EDSP Tier 2:

- Two-generation Rat Reproduction Test
- Medaka Multigeneration Test
- Japanese Quail Two-generation Test
- Mysid Two-generation Toxicity Test
- Larval Amphibian Growth and Development Test

The Two-generation Rat Reproduction Test (OCSPP 870.3800) has already been adopted. The remaining four Tier 2 tests are the subject of this FIFRA Scientific Advisory Panel review.

1.2 Validation

Validation has been defined as "the process by which the reliability and relevance of a test method are evaluated for a particular use" (NIEHS, 1997; OECD, 1996).

Reliability is defined as the reproducibility of results from an assay within and between laboratories.

Relevance describes whether a test is meaningful and useful for a particular purpose (OECD, 1996). For Tier I EDSP assays, relevance can be defined as the ability of an assay to detect the effects of chemicals with the potential to interact with the endocrine system.

Validation is generally recognized as necessary for the regulatory acceptance of new and revised test methods, and is now an integral component of the international development and acceptance of these methods (OECD, 2005). The criteria used to guide the validation process for the MMT were based on the principles of validation developed by the U.S. Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) (NIEHS, 1997) and the OECD (OECD, 2005). These criteria as stated by ICCVAM (NIEHS, 1997) are as follows:

1. The scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.

2. The relationship of the endpoints determined by the test method to the *in vivo* biologic effect and toxicity of interest must be addressed.

3. A formal detailed protocol must be provided and must be available in the public domain. It should be sufficiently detailed to enable the user to adhere to it and should include data analysis and decision criteria.

4. Within-test, intra-laboratory and interlaboratory variability and how these parameters vary with time should have been evaluated.

5. The test method's performance must have been demonstrated using a series of reference chemicals preferably coded to exclude bias.

6. Sufficient data should be provided to permit a comparison of the performance of a proposed substitute test to that of the test it is designed to replace.

7. The limitations of the test method must be described (*e.g.*, metabolic capability).

8. The data should be obtained in accordance with Good Laboratory Practices (GLPs).

9. All data supporting the assessment of the validity of the test methods including the full data set collected during the validation studies must be publicly available and, preferably, published in an independent, peer-reviewed publication.

The EPA has adopted these various validation criteria for the EDSP as described in Appendix A: Criteria for Validation (EDSP, 2007) which were considered met in the review of the Tier 1 assays. Validation of these assays proved a challenge to comply with Criterion No. 6 because these are not replacement assays (Validation Criterion No. 6). Many of them are novel assays; consequently, large data bases do not exist as a reference to establish their predictive capacity (*e.g.*, determination of false positive and false negative rates).

In general, the EPA uses a five-part or stage validation process outlined by ICCVAM (NIEHS, 1997). The EPA also believes it is important to recognize that this process was specifically developed for alternative assays intended to replace or refine existing assays. A rudimentary problem confronting the EPA is how to adapt and work with this process for rodent and ecological *in vivo* assays in Tiers 1 and 2 that are being considered as the standards and not as

alternatives to some other assay. Nonetheless, the stages of the process outlined by the ICCVAM are discussed below.

The first stage of the process as employed for the MMT was "test development", an applied research function which culminated in an initial protocol. As part of this phase, EPA drafted a Detailed Review Paper (DRP) to explain the purpose of the test method, the context in which it will be used, and the scientific bases upon which the assay's protocol, endpoints, and relevance rest (Appendix B: Detailed Review Paper). The DRP reviewed the scientific literature for candidate protocols and evaluated them with respect to a number of considerations, such as whether candidate protocols met the assay's intended purpose, costs, and other practical considerations. The DRP also identified the developmental status and questions related to each protocol; the information needed to answer the questions; and, when possible, recommended an initial protocol for the initiation of the second stage of validation, *i.e*., "standardization and optimization". During standardization and optimization, studies were performed geared toward refining, optimizing, and standardizing the protocol, and initially assessing protocol transferability and performance. In the third stage, *i.e*., "interlaboratory validation" studies were conducted in several independent laboratories with the refined protocol. The results of these studies were used to determine interlaboratory variability and to set or cross-check performance criteria. The report on the interlaboratory trials for the MMT is provided in Appendix C: Vinclozolin Interlaboratory Report. Interlaboratory validation is followed by the "peer review" phase, involving an independent scientific review by qualified experts, and by then by the final phase, "regulatory acceptance", *i.e.*, adoption for regulatory use by an agency. EPA has developed extensive guidance on the conduct of peer reviews because the Agency believes that peer review is an important step in ensuring the quality of science that underlies its regulatory decisions (USEPA, 2006).

The purpose of this ISR is to provide a summary of the development and validation of an MMT protocol (Appendix D: MMT Protocol) and a final proposed Medaka Reproduction Test (MRT; Appendix E: MRT Protocol) to be used as an *in vivo* assay for EDSP Tier-2 testing. The reasoning and judgments leading to the various studies, and conclusions concerning the strengths and limitations of the MMT research assay in its initial form, and advantages of the final proposed MRT are presented.

2 HISTORICAL OVERVIEW OF FISH TESTS

Fish are an important vertebrate class in both freshwater and marine systems. Although many fish toxicity test protocols are routinely used in regulatory testing, few have been designed with endocrine-specific endpoints in mind. Further, it is known that growth, reproduction, development, and other aspects of fish physiology and life cycle are under endocrine control. Thus, vertebrate test method that uses fish for evaluating potential effects of endocrine disrupting chemicals (EDCs) is relevant.

2.1 Purpose of the Medaka Multigeneration Test in EDSP

The purpose of definitive Tier 2 testing is to determine whether a test substance adversely affects a test organism through endocrine-mediated pathways, and to quantitatively evaluate those effects with respect to the estrogen, androgen, and thyroid systems. In addition, the tests must include exposure during the most sensitive life stages, provide the opportunity for identification of dose-response effects, and encompass a variety of taxa.

The recommended fish (Tier 2) test will characterize the nature, likelihood, and dose-response relationship of any adverse outcomes from potential endocrine disruption via E, AE, A, and/or TT pathways. In combination with other ecotoxicity tests in the EDSP Tier 2 and Part 158 studies, a comprehensive ecological risk assessment could be performed. The Tier 2 tests should complement Tier 1; however, results from well run Tier 2 tests would supersede Tier 1 results. Tier 2 is the final phase of the screening and testing program and, thus, provides more detailed information regarding the endocrine disruption activity of a tested chemical or mixture. To fulfill this purpose, these tests are much longer-term studies designed to encompass critical life states and processes, a broad range of doses, and administration by a relevant route of exposure. In addition, the effects associated with an EDC may be latent and not manifested until later in life or may not be apparent until reproductive processes occur in an organism's life history. Thus, tests for endocrine disruption will often encompass additional generations, including effects on fertility and mating, embryonic development, sensitive neonatal growth and development, and transformation from the juvenile life state to sexual maturity. The results from the Tier 2 testing should be conclusive, and a discernible cause-effect relationship should be manifest during the test if one exists from the chemical exposure. In summary, the following elements should be part of the test:

- establish exposure/concentrations/timing and effects relationships
- be sensitive and specific
- assess relevant endpoints
- include all life stages across generations
- include a dose range for full characterization of effects
- capable of being routinely conducted in accordance with Good Laboratory Practices (GLP).

2.2 Medaka multigeneration toxicity tests for the detection of possible endocrine disruption

2.2.1 Advantages of the medaka test model

The medaka (*Oryzias latipes*) is a freshwater fish belonging to the family of Asian rice fishes (Adrianichthyidae) indigenous to areas of Japan, Taiwan, and southeastern Asia, where ambient temperatures range from 5ºC to 35ºC (Kirchen and West, 1976). The medaka has a long history as an experimental animal and a complete presentation of their biology can be found in Yamamoto (Yamamoto, 1975), which lists over 1000 references dating from the early 1900s. The culture and handling of the medaka has been exhaustively studied and detailed guidelines are available in EPA/600/3-91/064 (Denny et al., 1991). Medaka are small (25 to 50 mm in total length) and are easy to rear and maintain in laboratory aquaria. Adults can be maintained in an aquarium with little space and can tolerate low dissolved oxygen and a wide range of temperatures and salinity (Yamamoto, 1975). Medaka are sexually dimorphic and although the sex of fully grown fish can be determined by observing the outline of the anal and dorsal fins, the observation of an anesthetized fish under a low-power microscope is recommended for confirmation and for sexing an immature fish (Yamamoto, 1975).

Medaka have a generation interval of 2 to 3 months, but can be induced to spawn throughout the year by controlling the temperature and photoperiod. Spawning is highly predictable in time, usually within 1 hr of first daylight (Hirshfield, 1980; Koger et al., 1999). Measures of medaka reproductive output, including gonadal morphology, fecundity, and fertility, demonstrate that through laboratory control of temperature and photoperiod, reproductive capacity of breeding groups can be calibrated before exposure, thereby establishing a baseline from which to assess effects (Koger et al., 1999). Individual females can lay an average of 25 eggs/day for up to 4 months under proper conditions of photoperiod, temperature, and food supply (Hirshfield, 1980). A cluster typically contains 10 to 30 eggs that are attached to the female's vent by filaments for a number of hours until they are brushed off onto a spawning substrate simulating aquatic plants (Yamamoto, 1975). Even when spawning substrates such as a spawning sponge are provided, some females retain their eggs requiring manual stripping (Denny et al., 1991). The eggs should be collected as soon as possible after spawning to prevent their predation by adults. Fertilization can be easily assessed with low magnification because of the transparency of the egg chorion. The egg incubation period is approximately 1 week when kept at 28ºC (Yamamoto, 1975), and the embryos will tolerate a temperature range of 7ºC to 38ºC (Kirchen and West, 1976).

2.2.2 Conceptual basis for a multigeneration test paradigm

Endocrine active substances are known to disrupt endocrine regulated organizational as well as activational processes in an organism. It is also known that in oviparous animals there is maternal transfer of xenobiotic as well as endogenous factors to the embryo, *in ovo* (Adkins-Regan et al., 2013; Almasi et al., 2012; Foran et al., 2000; Hayward and Wingfield, 2004; Henriksen et al., 2011; Lin et al., 2004; Metcalfe et al., 2000; Schwaiger et al., 2002; Schweitzer

et al., 2013). Further, organizational effects during development and maturation may not fully manifest as adverse until a later life stage or in future progeny (reproduction or sex ratio) (Adkins-Regan et al., 1995; Edmunds et al., 2000; Quinn et al., 2007; Quinn et al., 2006; Yamamoto, 1958; Yamamoto, 1953). Therefore, a multigeneration test is intended to evaluate these possible trans-generation effects. A Medaka Multigeneration Test design was conceived to most efficiently maximize the adverse impact-potential of an endocrine active substance. The initial F_0 generation is exposed during spawning to allow test compound and maternal factors to be transferred into the eggs of the F_1 generation. This transfer establishes the initial chemical environment of the embryo. Conceptually, beginning exposure at the egg stage alone would limit chemical exposure due to the attenuation of chemical absorption by the egg chorion and it will not present an altered state of maternal factors, *e.g.*, base estrogen levels to the embryo. Any potential organizational perturbation to the embryo *in ovo* would therefore be missed. If an F_0 generation is exposed after onset of reproduction, then some transfer of factors that present potential organizational effects may be considered. However, the maternal transfer at this stage may be partially affected or otherwise modified since the parent would have normally developed and matured prior to exposure. If that is the case, the F_1 animals would be subject to a modified receipt of maternal factors as well as increased test compound accumulation from continuous exposure post-hatch and is expected to be fully affected in their development and maturation. Their progeny (F_2) would therefore be expected to receive the full spectrum of modified factors which may further express and manifest in developmental or maturational disturbances. The distinction between activational effects to sexually differentiated fishes exposed to EDCs compared to the organizational effects that happen where exposure *in ovo* of endogenous factors and/or a bioaccumulated xenobiotic EDC that occurs during sexual differentiation has been generally reviewed by Denslow and Sepulveda (Denslow and Sepúlveda, 2007). It is the sexual differentiation and reproductive competence of the F_2 generation that would conceptually be most severely affected if there are organizational changes that result from a chemical's endocrine action to the developing embryo *in ovo*. If that is indeed the case, then a multigeneration test paradigm would be needed to fully characterize the potential endocrine disrupting impact.

Multigeneration fish tests in some form have been performed on several fish species including sheepshead minnow (*Cyprinodon variegatus*) (Cripe et al., 2009; Cripe et al., 2010; Raimondo et al., 2009), fathead minnow (*Pimephales promelas*) (Staples et al., 2011), zebrafish (*Danio rerio*) (Nakari and Erkomaa, 2003), Chinese rare minnow (*Gobiocypris rarus*) (Zha et al., 2008), Australian crimson-spotted rainbowfish (*Melanotaenia fluviatilis*) (Holdway et al., 2008), convict cichlid (*Amatitlania nigrofasciata*) (Newsome, 1980), as well as in medaka (Patyna and Cooper, 2000; Patyna et al., 1999). From the few studies and few chemicals that have been investigated, the importance of inclusion of additional generations is somewhat equivocal beyond the need to inform population models in probabilistic assessments or investigate putative epigenetic effects. Comparing sensitivities between activational and organizational effects is important in assessing the additional value of multiple generation tests. A key objective of the MMT method development is to address the question of the relative importance of an additional full F_2 generation.

3 MEDAKA MULTIGENERATION TEST PROTOCOL DEVELOPMENT

As discussed in **Section 2**, the focus of the ISR is on multigeneration studies conducted using medaka, a small fish test model. This section summarizes the initial full version of the MMT protocol. This protocol incorporated redundant endpoint measurements at comparable developmental life-stages in all generations of the test, F_0, F_1, and F_2. As such, it is considered a research protocol to be used for evaluating the value added of various endpoints and additional generations to a possible Tier 2 fish reproduction test protocol. A complete presentation of the MMT research protocol is included as Appendix D: MMT Protocol.

3.1 Overview of the Medaka Multigeneration Test (MMT) Research Protocol

The research MMT protocol describes a 2½ generation toxicity test with a small freshwater fish that evaluates reproductive fitness as an integrated measure of toxicant effects across 2½ consecutive generations using pairs of adult medaka. It also enables measurement of a suite of other endpoints for diagnostic and quantitative evaluation of EDCs or other types of reproductive toxicants. The MMT is a relatively long-term assay (29 weeks of in-life exposure) that assesses hatch, growth, survival, gonadal development, and reproduction in 2½ generations (F_0, F_1, and part of F_2). The MMT is an extension of existing standard Fish Full Life-Cycle test protocols developed by Japanese scientists (Seki et al., 2003b; Yokota et al., 2001).

The research objectives were to: a) develop a multigeneration test protocol using the Japanese medaka that could be used to evaluate the effects of endocrine disruption chemicals (EDCs) on reproduction and reproductive development; b) to include diagnostic or intermediate effects endpoints that could be evaluated for associations with the apical outcomes to help elucidate various EDC-induced adverse outcome pathways (AOPs); and, c) develop a scientifically-defensible Tier 2 fish test protocol for characterizing concentration-response for ED endpoints. The project involved developing methods for evaluating population-relevant endpoints such as fecundity, fertility, reproductive behavior, gonadal development, and the genotypic sex of individual fish compared to phenotypic sex within each generation tested. Also, a number of molecular and histological measures that are diagnostic of initiating and subsequent key events along AOPs were developed to allow endpoint comparisons across test designs and chemical modes of action.

The MMT provides data that can be used to simultaneously evaluate two general types of adverse outcome pathways (AOPs) ending in reproductive impairment: a) those primarily involving disruption of the hypothalamus-pituitary-gonadal (HPG) endocrine axis (i.e., endocrine-mediated EDCs); and, b) those that cause reductions in apical effects on survival, growth, hatch, etc through other endocrine axes (*e.g.*, HPT) and non-endocrine mediated toxicity pathways. Some of the EDC endpoints, such as the presence of anal fin papillae in medaka males, are biomarkers only minimally linked to adverse reproductive outcomes; whereas other EDC endpoints such as fecundity and fertility can be directly linked to adverse population outcomes through population models. The non-endocrine mediated endpoints typically measured in full life-cycle and early life-stage (ELS) chronic toxicity tests are also included in the MMT. These are used to evaluate the concentration-response of a chemical working through any endocrine or non-endocrine mediated AOP. The question of whether a chemical has endocrine-mediated effects at lower concentrations than non-endocrine-mediated toxicities cannot be unambiguously addressed by the apical endpoints alone and will rely on other diagnostic

biochemical (*e.g.*, vtg induction), histopathological (*e.g.*, Leydig cell hyperplasia), or changes in SSC.

The test conditions for the research MMT are outlined in Table 3-1. Briefly, the MMT protocol consists of continuously exposing 2½ generations of fish to the test chemical starting with maturing adults in the first generation (F0) and exposing for several weeks to allow chemical accumulation. There are 6 treatments (if a solvent is not used); five chemical treatments and an unexposed control treatment. There are six replicates aquaria in each treatment. The measured endpoints include embryo hatch, growth, survival, fecundity and fertility, liver vitellogenin mRNA levels in both sexes, the number of anal fin papillae (a sexual phenotype marker), histological evaluation of the gonad phenotype, and histopathology evaluation of kidney, liver and gonad. All of these endpoints are evaluated in the context of the genetic sex of each fish. Chemical concentrations are held constant in each treatment throughout the bioassay to allow endpoint sensitivity comparisons across and within generations. Thus, although not discussed here in detail, precise dependable flow-through exposure and dilution systems, as well as high-quality analytical chemistry, are an integral part of the protocol. The in-life duration of the MMT is nominally 29 weeks. Note that each generation overlaps with the subsequent generation by three weeks. Details of the endpoints measurements are presented in Section 3.2, and a detailed protocol and sampling timeline is in Section 3.3 indicating when each MMT measurement endpoint is evaluated.

Table 3-1. Test Conditions for the Medaka Multigeneration Test (MMT) Protocol

Recommended species	Medaka (Oryzias latipes)
Test type	Continuous flow-through
Water temperature	26 ± 1°C
Illumination quality	Fluorescent bulbs (wide spectrum and ~1100 lumens)
Photoperiod	16 h light, 8 h dark
Loading rate	F_0- 2 fish/replicate; F_1 and F_2- initiated with maximum of 40 larvae/replicate eventually culled to 2 fish/replicate
Test replicate chamber size	18x9x15 cm
Test replicate solution volume	~1.8 liters
Volume exchanges of test solutions	12 volume replacements/day (15 ml/min flow).
Age of test organisms at initiation	F_0-11 weeks post-fertilization, F_1 and F_2 continuously from fertilization
No. of organisms per replicate	F_0- 2 adults; F_1 and F_2- initiated with maximum of 40 larval fish eventually culled to 2 adults
No. of treatments	6 (5 toxicant treatment plus 1 control)
No. replicates per treatment	6
No. of organisms per test concentration	F_{0-} 12 for each test concentration and for control; F_1 and F_2- start with a maximum of 240 and eventually cull to 12 for fecundity assessment.
Feeding regime	Fish were fed 24-hour old brine shrimp (*Artemia* spp.) nauplii at the following rates:

Time (weeks post-hatch)	**Brine Shrimp (mg dry weight/fish/day)**
1	0.77
2	3.06
3	5.48
4	7.75
5	9.12
6-7	11.4
8-15	13.7

Aeration	None unless DO reaches <60% saturation

Dilution water	50 micron sand-filtered Lake Superior Water, UV sterilized in the lab and further filtered to 5 microns and degassed to 8.0 mg O_2/L.
Exposure period	Trans-generational test: F_0 from 11-16 weeks post-fertilization (wpf),), F_1 and F_2 from 0-16 wpf
Primary endpoints	Survival: F_0 from 11-16 wpf, F_1 and F_2 from 1-3 wpf and from 3-16 wpf Growth : length, weight, (F_0, F_1, F_2 at 16 wpf termination; F_1 and F_2 7 wpf) Hatch: $F_1$and F_2 only Anal Fin Papillae: F_0, F_1, F_2 assessed at 16 wpf termination; F_1 and F_2 assessed at7 wpf Fecundity: assessed for each breeding pair T,W,Th,F of weeks 11-15 for F_0, F_1 and F_2 Fertility: assessed for each breeding pair T,W,Th,F of weeks 11-15 for F_0, F_1 and F_2 Vitellogenin: determined at 7 and 16 wpf for F_0, F_1 and F_2
Optional endpoints	Time to first spawn for F_1 and F_2 generations.
Test acceptability	Dissolved oxygen $\geq$60% of saturation; mean temperature of 26 ± 1°C; ≥70% survival in control treatments of each generation; successful reproduction in at least 65% of control females with mean daily fecundity in the control treatment of at least 15 eggs.

3.2 MMT Endpoints

An overview of the endpoints incorporated into research MMT are listed in Table 3-2. Some of the endpoints are related to EDC effects while others are traditional ecotoxicology endpoints generally considered to be a result of systemic toxicity, and not necessarily endocrine effects.

Table 3-2. Endpoint overview of the medaka multigeneration test (MMT)

Life-stage	Endpoint	Endocrine-mediated	Non-Endocrine-Mediated	Population Relevance
ELS[1] (1 wpf[2])	Hatch		Y	Y
ELS (3 wpf)	Survival		Y	Y
Subadult (8 wpf)	Survival		Y	Y
	Growth (weight)		Y	Y
	Vitellogenin (Vtg)	Y		
	Anal fin papillae	Y		
	Intersex	Y		Maybe
	Sex Reversal	Y	Y	Y
Adult (12-16 wpf)	Fecundity	Y	Y	Y
	Fertility	Y	Y	Y
Adult (16 wpf)	Survival		Y	Y
	Growth (weight)		Y	
	Vitellogenin (Vtg)	Y		
	Anal fin papillae	Y		
	Intersex	Y		Maybe
	Sex Reversal	Y	Y	Y
	Histopathology	Y	Y	Maybe

[1] ELS: early life-stage
[2] wpf: weeks post-fertilization

The primary emphasis in the MMT is on endpoints that can be associated with potential adverse effects on population-relevant parameters such as survival, gross development, growth and the reproductive measures of fecundity and fertility. To help differentiate endocrine-mediated effects from systemic toxicity, additional endocrine-mediated effects such as liver vitellogenin mRNA

levels, the number of papillae on the anal fin (an external sexual phenotype marker), and the histological evaluation of gonadal sex and intersex, and finally, histopathology evaluation of kidney, liver, gonad, and other tissues are also measured (see endpoint list in Table 3-2). All endpoints are evaluated in the context of the genetic sex of each fish. As mentioned previously, traditional ecotoxicology endpoints that may not be endocrine-mediated such as embryo hatch, growth, and survival are also measured.

3.2.1 Genotypic Sex (F_O, F_1, F_2)

The genetic sex of each fish, for each generation, is determined by non-destructively obtaining a small tissue (fin) sample and, using polymerase chain reaction (PCR) methods, identifying the presence or absence of the sex-determining gene (DMY) in each fish. Under normal conditions XX individuals express the female phenotype and XY individuals the male phenotype.

3.2.2 Survival (F_O, F_1, F_2)

Survival is assessed daily for F_0, F_1, and F_2 individuals throughout the MMT to quantify chemically-induced mortality. The survival rates of larval fish at 3 weeks post-fertilization (wpf) are evaluated in F_1 and F_2 based on the difference in the number of hatched embryos placed in each replicate and the number of survivors collected. Subadult survival (F_1, F_2; 8 wpf), and adult survival (F_0, F_1, F_2; 16 wpf) are evaluated based on the number of fish present in each replicate at the sampling time for each life-stage (see Section 3.3 for additional information on sampling timeline).

3.2.3 Growth (F_O, F_1, F_2)

A common toxic effect in response to chemical exposure is reduced growth, thus growth is incorporated into the MMT to allow the evaluation of possible non-endocrine mediated overt toxicity as a possible explanation for treatment induced reductions in reproductive outcomes. While it is true that endocrine processes can be linked with growth outcomes, it is likely that an integrated evaluation of all the test data will allow discrimination of endocrine versus non-endocrine mediated AOPs. In the MMT, growth is determined on days when fish are being removed from aquaria to reduce the number of fish per replicate when starting a new phase of the test protocol. Specifically, wet weight and length are measured in all F_1 and F_2 fish shortly after metamorphosis to juveniles as the number of fish per replicate aquarium are reduced to 10. Subadult growth is measured in the F_1 and F_2 fish remaining after reproductive pairs have been selected. Finally, adult growth is measured in F_0, F_1, and F_2 after reproduction has been assessed and the fish in that generation are terminated. Growth is measured by measuring weight to the nearest 1 mg and length to the nearest 0.1 mm.

3.2.4 Reproduction (F_O, F_1, F_2)

The fecundity and fertility of reproductive pairs is assessed four days per week for 5 weeks in each generation (F_0, F_1, and F_2). On the morning of each day (Tuesday through Friday of each

week) egg clumps are removed from females and from the bottom of each aquarium and counted for fecundity and evaluated under a dissecting microscope for fertility.

3.2.5 Liver Vitellogenin mRNA (F_O, F_1, F_2)

Vitellogenin (Vtg), a common biomarker of EDC exposure, is measured by determining the Vtg mRNA copy number per nanogram (ng) of total mRNA in the liver by quantitative PCR (QPCR). QPCR is highly sensitive and has a large dynamic range (~7 orders of magnitude), making it an ideal tool for efficiently assessing concentrations of a target gene. Specifically, Vtg is measured in adults at the end of each generation (F_0, F_1, and F_2) and in subadults in F_1 and F_2. The gene expression data in each treatment are categorized by gender genotype (XX or XY) and then compared to the gender-specific control data from each particular generation and life-stage.

3.2.6 Secondary Sex Characters (F_O, F_1, F_2)

Male medaka normally have anal fin papillae that develop as a secondary sexual characteristic (SSC) during the transition from subadult to adult. These bony structures are easily counted and are responsive, negatively or positively, to various EDC exposures. In males, it is common for estrogen receptor (ER) agonists (*e.g.*, estradiol and 4-tert octylphenol) to lower the number of papillae or prevent their development altogether; conversely, androgen receptor (AR) agonists (*e.g.*, 17β-trenbolone) can induce the development of papillae in females. In all three generations, during necropsy of adult and subadult samples, the portion of the fish posterior to the vent is removed and fixed in Davidson's solution. After fixation, the anal fin papillae are counted with the aid of a dissecting microscope, and recorded as the SSC measurement.

3.2.7 Hatch Success (F_O, F_1, F_2)

Eggs are placed in an incubator that is suspended in the same aquarium used to expose the spawning pair. The developing eggs are continually agitated within the incubator by air bubbles. Incubators are checked daily for mortalities. Dead eggs are recorded and removed from the incubators. To promote hatching, the aeration of the incubators is stopped on the 8th day and the eggs settle to the bottom of the incubator. The majority of control eggs hatch over the next two days. For each treatment, hatchlings are counted, and systematically distributed to each of the replicate aquaria. Embryos that have not hatched by the 14th day post fertilization are considered non-viable and discarded.

3.2.8 Pathology and Gonad Phenotype (F_O, F_1, F_2)

Particularly in the subadult samples, where the sample number is greater, but also in the adult samples, the phenotype of the gonad is determined. At its simplest, the determination is either ovary or testis. However, an EDC exposure will often induce some level of intersex condition in which a single gonad will contain both ovarian and testicular tissue of varying proportions, presumably correlated to the potency of the EDC exposure. In addition a small percentage of control males will have testes with a few minimally developed oocytes. The prevalence of this

appears to be somewhat strain-dependent (Grim et al., 2007), but in the EPA cultures, it seems to be less than 2%. The assessment of gonad phenotype in subadults involves transverse oblique sectioning of the gonad (sections of the kidney are often collected as well) to optimize the sectioning plane. These sections are routinely hematoxylin and eosin (H&E) stained and viewed by a trained pathologist that categorizes them as one of the following: 1 = testis only; 2 = mostly testicular tissue, some ovarian tissue; 3 = approximately equal prevalence of testicular and ovarian tissues; 4 = mostly ovarian tissue, some testicular tissue; 5 = ovary only. The same categories are used in the assessment of adults, but a sectioning protocol optimized to produce sections of multiple organs and tissues is used.

While observations of histopathology are made on the subadult sectioned obliquely, the primary histopathological analysis is done on the adult samples which can then be compared with the reproductive output of the specific pair. Therefore, the adults are sectioned parallel to the sagittal plane including one midline level and four symmetrical parasagittal levels to produce sections, primarily, of gonad, kidney, and liver while other secondary tissues can often be observed as well. Diagnoses are scored for severity by the pathologist and these severity scores are entered into StatCHARRMS for statistical analysis. Details of the histotechnical procedures and pathology guidance documents have been provided (Appendix F: Pathology Guidance). In addition, an internet-based resource is being developed that would not only include histotechnical and pathology guidance documents from various sources, but also would be an interactive atlas to help harmonize diagnoses between independent pathologists. A description of this resource is available in (Appendix F: Pathology Guidance).

3.3 Research MMT Timeline

The timeline for the MMT protocol is illustrated in Figure 3-1. The test begins with exposure of adult F_0 fish, and continues the exposure through the full medaka life-cycle for two additional generations, F_1 and F_2. The test is complete after assessing reproductive endpoints in the F_2 generation.

Figure 3-1. The time-line and endpoints in the research MMT. The MMT is 29 weeks long and spans 2 ½ generations; F_0, F_1, and F_2. The rows to the right of each endpoint listed on the left indicate the study-week when the endpoint is measured in each generation. The life-stages of the test animals are color-coded.

MMT Exposure and endpoint timeline

Gen																														
	F0	1	2	3	4	5																								
	F1			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15												
	F2															1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Study Week		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Lifestage Key		Embryo							Larvae					Juvenile					Subadult							Adult				
Endpoints																														
Fecundity		F_0												F_1												F_2				
Fertility		F_0												F_1												F_2				
Hatch					F_1												F_2													
Survival						F_0	F_1			F_1								F_1	F_2			F_2								F_2
Growth						F_0				F_1								F_1				F_2								F_2
Vitellogenin						F_0				F_1								F1				F2								F2
Secondary sex						F_0				F_1								F1				F2								F2
Histopathology						F_0				F_1								F1				F2								F2
Study Week		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

3.3.1 F_0; Study weeks 1-5

Young adult breeding pairs (11 wpf; females >250 mg; males >200 mg) of the first generation (F_0) are moved into the exposure system at the start of Study week 1 and continually exposed for the next five weeks. During this exposure period developing oocytes and sperm within the breeding adults accumulate the test chemical and/or its metabolites. Depending on toxicokinetic factors, the concentration of the test chemical in the gametes may a steady state concentration relative to each treatment levels. On Study Day 1, a minimum of six female-male breeding pairs are placed in an exposure system that has six treatment levels including controls with six replicate aquaria per treatment concentration. Each replicate aquarium houses a single breeding fish pair for a test total of 36 breeding pairs. The fecundity of each breeding pair is assessed for the next 5 weeks by collecting all spawned eggs from each breeding pair on weekdays T, W, T, F. The F_0 breeding pairs are sampled at the start of Study week 6 for the endpoints listed in Table 3-2 and Figure 3-3.

3.3.2 F_1 and F_2; Weeks 1-3

The first week for generations F_1 and F_2 (study weeks 3 and 15, respectively), eggs are collected from each breeding pair of the previous generation and a maximum of 20 eggs are loaded into a glass incubator which is then placed in a separate incubator aquarium and sufficient aeration is supplied to the incubator to keep the eggs gently suspended in the water column of the incubator. This process is repeated the next day, resulting in 2 incubators, each containing a maximum of 20 eggs, from each breeding pair. Aeration is continued until 7 days post-fertilization (dpf), at

which time the aeration is turned off and the eggs begin to hatch. Each day, the larval fish that have hatched are collected, counted and placed in the incubator aquarium (*i.e.*, same aquarium and separate for each parental pair, but outside egg incubator enclosure).

3.3.3 F_1 and F_2; Week 4

At the start of week 4, all of the surviving larval fish from each treatment are counted, pooled and redistributed to the same aquaria that contained the breeding pairs with a maximum of 10 fish per aquarium. The survival to 3 wpf is determined at this point.

3.3.4 F_1 and F_2; Weeks 4-6

The exposure continues as the larval fish develop into sub-adults.

3.3.5 F_1 and F_2; Week 7

At the start of week 7, a small sample is taken from the caudal fin of each fish to determine the genotypic sex of the individual. This information is used to randomly establish female-male breeding pairs. The fish not selected as breeding pairs are sampled for the various endpoints outlined in Figure 3-3.

3.3.6 F_1 and F_2; Weeks 7-10

The exposure continues as the sub-adult breeding pairs develop into adults.

3.3.7 F_1 and F_2; Weeks 11-15

The fecundity of each breeding pair is assessed for the next 5 weeks by collecting all spawned eggs from each breeding pair on T, W, Th, and FF.

3.3.8 F_1 and F_2; Week 16

The breeding pairs are sampled at the start of week 16 for the various endpoints outlined in Figure 3-3.

3.4 Statistical Analyses

The biological data generated by the MMT are one of three types: 1) a continuous response that is measured once per subject (fish), *e.g.*, vitellogenin and weight; 2) a continuous response that is measured multiple times per subject (breeding pair), *e.g.*, fecundity which is measured on 21 consecutive days; or, 3) ordinal data, such as histopathology findings that are expressed as a

severity score. These types of data are not unique to the MMT and, except for pathology data, many other appropriate statistical methodologies have been developed to properly analyze them depending on the characteristics of the data including normality, variance homogeneity, whether the study design lends itself to hypothesis testing or regression analysis, etc. In order to analyze the complete datasets from all the MMTs presented in this ISR, a somewhat statistically conservative, systematic approach was used as described below. For the prevalence/severity score pathology data, a new statistical approach, described below and attached as an appendix, was developed.

All statistics were done with a program termed StatCHARRMS (Statistical analysis of Chemistry, Histopathology, And Reproduction endpoints including Repeated measures and Multigeneration Studies), a Statistical Analysis System (SAS®;SAS® Institute, Cary, NC)-based program developed for this use by John Green (DuPont Applied Statistics Group, Newark, DE). For continuous data measured once per subject, normality (via Shapiro-Wilks test) and variance homogeneity (via Levene's test) were assessed and appropriate transforms applied by the program to determine the appropriate statistic for hypothesis testing. In addition, regardless of the results of those two tests, a step-down Jonckheere-Terpstra trend (JT) test was also applied to the data as long as a monotonic response was observed. To be consistent across MMTs, the statistical test that indicated the most sensitive lowest observed effect concentration (LOEC) was reported. In addition, StatCHARRMS has the ability to analyze continuous data across multiple generations, but to simplify interpretation of the data within this ISR, only single generation analyses are presented. Table 3-3 summarizes the statistical tests used by StatCHARRMS, while a thorough description of the StatCHARRMS package including statistical rationale can be found in Appendix G: Statistics.

Table 3-3. Decision-tree determining the statistical test used by StatCHARRMS.

Shapiro-Wilks	Levene's	Statistical test
Pass	Pass	Dunnett; JT
Fail	Pass	Modified Dunn's; JT
Pass	Fail	Tamhane-Dunnett; JT
Fail	Fail	JT

For fecundity and fertility, which are continuous responses that are measured multiple times, a repeated measures analysis of variance (ANOVA) with two main effects, *i.e.*, treatment and week (or day), and the treatment-by-week interaction are used for analysis. During the MMT research program, transformations of the fecundity and fertility data never produced normally distributions so the nonparametric analysis of a repeated measures ANOVA on the ranks of the data is used. When comparing levels of the treatment effect, *i.e.*, treatments averaged across weeks, against the controls, a Dunnett's test is used to adjust for the family-wise error rate. For the week effect, *i.e.*, comparisons of weeks averaged across treatments, if either the week main effect or the week-by-treatment interaction is significant at the 0.05 level, then the comparisons of weeks averaged across treatments or within a treatment are done at the 0.05 level with no adjustment. Similarly, if the treatment-by-week interaction is significant at the 0.05 level, then treatments within the same week are compared at the 0.05 level unadjusted p-values. Finally, if

the treatment-by-week interaction is not significant, then comparisons of weeks within a treatment or treatments within a week are compared using Bonferroni-Holm adjusted p-values.

For histopathology data which are in the form of severity scores, StatCHARRMS uses a new test termed RSCABS (Rao-Scott Cochran-Armitage by Slices). RSCABS uses a step-down Rao-Scott adjusted Cochran-Armitage trend test on each level of severity in a histopathology response. The Rao-Scott adjustment incorporates the replicate vessel experimental design into the test. The by-slices procedure incorporates the biological expectation that severity of effect tends to increase with increasing doses or concentrations, while retaining the individual subject scores and revealing the severity of any effect found. The RSCABS procedure not only determines which treatments are statistically different, *i.e.*, have more severe pathology than controls, but it also determines at which severity score the difference occurs providing much needed context to the analysis. A detailed description of this test can be found in Appendix G: Statistics.

3.5 Data Expectations

The types of data generated within the MMT are relevant to population dynamics (*e.g.*, reproduction), to the description of the affected AOP(s) (*e.g.*, Vtg mRNA induction and SSC), and to non-specific, or at least non-EDC related, toxicity (*e.g.*, growth parameters). It is expected that the complete suite of data from the MMT would not only define a LOEC for reproduction but also provide valuable information regarding the affected AOP, perhaps to the extent of providing a "fingerprint" to identify the test agent not only as an EDC but also as, for instance, an ER agonist. Since the MMT measures apical endpoints and does not measure direct consequences of the molecular initiating event (MIE) within an AOP, it does have limitations to its ability to resolve the various known or postulated AOPs. However, information from the Tier 1 studies will provide relevant information regarding the AOP of the test chemical.

For ER agonists (or perhaps more appropriately ER activation), the pathway, after agonism of the estrogen receptor, is expected to include, in male fish, production of vitellogenin (therefore up-regulated vitellogenin genes), feminized phenotypes of various somatic tissues normally influenced by the HPG, developmental abnormalities and feminization of the testis including, at the extreme, complete sex reversal, and finally impaired reproductive capacity (Ankley et al., 2010b). While ER agonists, in particular strong ER agonists, presumably induce adverse effects in females, the affects appear to be less specific. They may impair ovary and/or brain development which could reduce reproductive performance as well. The MMT does not assess reproductive capacity of exposed males and females independently (for instance by forming a breeding pair from an exposed male and an unexposed female) therefore, the MMT is unable to completely resolve impairment of one gender versus the other.

Several ER agonists were tested during MMT development, ranging from strong (17β-estradiol) to very weak *in vitro* agonists (4-chloro-3-methylphenol). Within these MMTs, the data closely reflected effects to the measured parameters that would be expected. Typically, the end result of exposure was reduced fecundity. Changes to biomarker endpoints were also observed including decreased SSC (in males), up-regulation of vitellogenin, and feminization of the gonad, all

without adversely impacting growth. This suite of outcomes is consistent with an ER activation AOP.

For AR agonists, the pathway after binding (the MIE) seems to include hormonal feedback that results in the reduced synthesis of testosterone and estrogen which leads to reduced vitellogenin production, impaired oocyte development, and finally impaired reproduction (Ankley et al., 2010a). In addition, AR agonists could be expected to produce masculinized somatic tissues normally influenced by the HPG and developmental abnormalities and changes in the phenotype of the ovary, including sex reversal. During the development of the MMT, a single strong AR agonist (17β-trenbolone) was used as a test agent. Near the concentrations that reduced fecundity, the MMT also provided data from biomarkers indicating the involvement of the AR agonism AOP including the presence of male-specific SSC in female fish, the reduction of vitellogenin, and the development of testes in XX fish. Again, the suite of data seemed to provide a hallmark pattern scientifically ascribable to the AR activation AOP.

Another MIE that leads to an AOP, albeit not as well defined, is the inhibition of aromatase which catalyzes the conversion of testosterone to 17*β*-estrogen. The literature suggests that aromatase inhibition reduces 17*β*-estrogen plasma concentrations, reduces vitellogenin uptake into oocytes, and reduces fecundity (Ankley et al., 2002; Sun et al., 2007b). Indeed the MMT with prochloraz, an aromatase inhibitor had the expected results: reduced fecundity and decreased vitellogenin with no effect on the development of SSC in males. While far from conclusive and needing further research, the MMT may serve as a means to differentiate between AR agonists and aromatase inhibitor/steriodogenesis inhibitor based upon the responses of the SSC and vitellogenin.

The responses of fish in various bioassay systems to AR antagonists have been somewhat ambiguous (Bayley et al., 2003; Kiparissis et al., 2003; Makynen et al., 2000). Enough so that testing potential of AR antagonists to reduce or eliminate the effects of a model androgen like 17*β*-trenbolone have been suggested as a means to identify AR antagonists (Ankley et al., 2010a). Therefore, expectations for any specific response to the AR antagonist, vinclozolin, are uncertain.

Table 3-4. Contingency table for effects of EDCs with different MIEs on vitellogenin, secondary sexual characteristics (SSC) and the histologically determined gonad phenotype. Arrows indicate direction of change and parentheses indicate a change that can occur, albeit not consistently. Based on (Ankley et al., 2010b).

MIE	Male			Female		
	Vitellogenin	SSC	Gonad Phenotype	Vitellogenin	SSC	Gonad Phenotype
ER agonist	↑	↓	feminized	(↑)	–	–
AR agonist	–	(↑)	–	↓	↑	masculinized
AR antagonist	–	(↓)	?	–	–	?
Steroidogenesis inhibitor	–	–	?	↓	–	?

4 MMT DATA SUMMARIES

This section provides a brief summary of the MMT tests performed in the U.S. and Japan beginning with the MMT protocol presented in **Section 3.** Each successive iteration of the MMT,

usually with a new/different test agent, was performed using a protocol slightly different than that used for the previous MMT. The differences were generally responses to one or more of the following considerations; a) experience and data from previous tests, b) discussions with Japanese scientists who were also performing MMT protocols, and c) results from ancillary experiments designed to provide information to improve the MMT protocol. Examples of modifications include: changes in the temperature of the test, duration and measurement times for reproduction endpoints, feeding protocols, egg incubation methods, histopathology methods, *etc*. Table 4-1 lists the MMTs that were performed in U.S. EPA or Japanese laboratories. Generally, EDCs with different AOPs were tested in the successive MMTs. Modifications of the MMT protocol were evaluated based on whether the change improved the efficiency of performing the test, or the statistical power for detecting endpoint effects by reducing variance or increasing the absolute value of a critical endpoint.

Table 4-1. Chemicals run in the MMT protocol

Chemical	**Laboratory**	**Protocol**	**AOP's MIE***
17β-Estradiol	CERI/Japan	Full MMT	ER agonist (strong)
4-*t*-Octylphenol	US EPA/MED	Full MMT	ER agonist (weak)
4-*t*-Octylphenol	NIES/Japan	Full MMT	ER agonist (weak)
o,p′-DDT	CERI/Japan	Full MMT	ER agonist (medium)
4-chloro-3-methylphenol	US EPA/MED	Full MMT	ER agonist in vitro (weak)
Tamoxifen	NIES/Japan	Abbreviated MMT[1]	ER binder/modulator; tissue-dependent antagonist/partial agonist
17*β*-Trenbolone	US EPA/MED	Full MMT	AR agonist
Prochloraz	US EPA/MED	Full MMT	Aromatase inhibitor; AR antagonist; steroid metabolism modulator
Vinclozolin	US EPA/MED	Abbreviated MMT[1]	AR antagonist
Vinclozolin	Interlab A	Abbreviated MMT[1]	AR antagonist
	Interlab B	Abbreviated MMT[1]	AR antagonist
	Interlab C	Abbreviated MMT[1]	AR antagonist
4-*t*-Octylphenol	US EPA/MED	Full MMT[2]	ER agonist

[1] Abbreviated MMT: same as Full MMT but without F2 reproduction
[2] Full MMT with additional replication for the assessment of reproduction in F1 and F2
*Molecular Initiating Event of an AOP

Taken *en masse*, the data indicate that the MMT is able to determine effect concentrations for several chemicals with MIE for various AOPs that are relevant to reproduction. In addition, the MMTs provided information regarding endpoints that might provide insight into intermediate markers (key events) that involve disruption of the HPG axis (endocrine-mediated chemical effects) and/or that cause reductions in survival, growth, hatch, *etc*. through non-endocrine mediated pathways. Table 4-2 summarizes the LOECs determined in each MMT for both

reproduction and the lowest LOEC for the non-reproduction endpoints, and representative LOECs for similar measurements from the literature involving the same test agent.

Table 4-2. Lowest observable effect concentration (LOEC) in each MMT for reproduction and lowest LOEC for any non-reproductive endpoint (biomarker) and generation for which it was detected, compared to values reported in the literature for the same test agent.

		MMT LOEC					Published Results	
Chemical	**Laboratory**	**Fecundity**		**Biomarker**			**Conc.**	**Effect**
17β-Estradiol	CERI/Japan	28 ng/L	F_1	2.8 ng/L	SSC	F_1	27.9 ng/L[a]	Fecundity
4-*t*-Octylphenol	EPA/MED	51 µg/L	F_2	13 µg/L	SSC	F_2	11.4 µg/L[b]	Vitellogenin
				13 µg/L	Gonad	F_1		
4-*t*-Octylphenol	NIES/Japan	100 µg/L	F_1	6.25 µg/L	SSC	F_1	11.4 µg/L[b]	Vitellogenin
				6.25 µg/L	Vtg	F_1/ F_2		
o,p′-DDT	CERI/Japan	1.9 µg/L	F_2	0.07 µg/L	Vtg	F_1/ F_2	0.1 µg/L[c]	Vitellogenin
CMP	EPA/MED	>345 µg/L	NA	88 µg/L	Vtg	F_0	No data	No data
Tamoxifen	NIES/Japan	20 µg/L	F_1	1.3 µg/L	Vtg	F_1/ F_2	625 µg/L[d]	Fecundity
				1.3 µg/L	SSC	F_2		
17β-Trenbolone	EPA/MED	32 ng/L	F_1	32 ng/L	SSC	F_1	39.7 ng/L[e]	Vitellogenin
				32 ng/L	Vtg	F_1		
				32 ng/L	Gonad	F_1		
Prochloraz	EPA/MED	25 µg/L	F_2	9 µg/L	SSC	F_1	30 µg/L[f]	Fecundity
				9 µg/L	Vtg	F_1/ F_2		
Vinclozolin	EPA/MED	>136 µg/L	NA	17 µg/L	SSC	F_1	2500 µg/L[g]	Spematogenesis
4-*t*-Octylphenol	EPA/MED	23 µg/L	F_1/ F_2	In process			11.4 µg/L[b]	Vitellogenin

a: Seki et al., 2005
b: Seki et al., 2003a
c: Uchida et al., 2010
d: Sun et al., 2007a
e: Seki et al., 2006
f: Zhang et al., 2008
g: Kiparissis et al., 2003

4.1 Results with ER-agonists

4.1.1 17β-Estradiol (CERI/Japan)

The primary reproductive endpoint (fecundity) was affected by exposure to 17β-estradiol, in both F1 and F2 generations with a LOEC of 28 ng/L. In F1, fecundity was further reduced by exposure to 84 ng/L (no breeding pairs were available for the 84 ng/L treatment in F2). On the other hand, reproduction was not affected by 17β-estradiol exposure in F0, presumably highlighting a greater sensitivity of organizational impacts over activational impacts on reproductive impairment.

The secondary endpoints that provide information about the potential AOP (in this case obviously identified as an ER agonist) responded in a manner and magnitude consistent with 17*β*-estradiol exposure reported in the literature (Cripe et al., 2009; Hirai et al., 2006; Raimondo et al., 2009; Seki et al., 2006; Seki et al., 2005). The LOEC, considering all endpoints equally, was 2.8 ng/L, based on SSC of F_1, XY subadult fish. This is the same LOEC reported from a fish full life-cycle test with medaka for F_1 (Seki et al., 2005). In further support of the ER agonist AOP, vitellogenin production in male livers was induced by 17*β*-estradiol exposure starting at 28 ng/L, and significant numbers of testicular oocytes started to appear at 28 ng/L as well. Growth parameters generally were not altered by 17*β*-estradiol exposure suggesting that systemic toxicity was not observed in the MMT. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 1-15.

4.1.2 4-t-Octylphenol (US EPA/MED)

4-t-Octylphenol (4tOP) is a relatively strong estrogen that has been shown to induce vitellogenin production and cause abnormal sex differentiation in medaka after a 60 day exposure (starting with embryos) at a concentration of 11.4 µg/L (Seki et al., 2003a). The literature clearly indicates that 4tOP is an ER agonist (Gray and Metcalfe, 1999; Gronen et al., 1999; Korner et al., 2001; Laws et al., 2000). Therefore, it is not surprising that it decreased reproduction in F_1 at 102 µg/L and in F_2 at 51 µg/L. At 102 µg/L, reproduction was essential shutdown while at 51 µg/L, 4tOP reduced the reproductive output by approximately 50%. On the other hand, reproduction was not affected by 4tOP exposure in F_0, presumably highlighting a greater sensitivity of developmental impacts over activational impacts on reproductive impairment.

Those endpoints that would be expected to be affected by an estrogen were responsive to 4tOP exposure. The overall LOEC for the MMT was 13 µg/L, based on SSC in F_2 XY fish, which matches the published LOECs from other protocols (Gray and Metcalfe, 1999; Knorr and Braunbeck, 2002; Segner et al., 2003; Seki et al., 2003a). In support of the ER agonist AOP, the male SSC LOEC was lowered in XY fish in both F_1 and F_2 at both subadult and adult samplings. Further, vitellogenin gene expression was induced in XY fish at both lifestages and in all generations with LOECs of either 25 µg/L or 51 µg/L, and the prevalence of testicular oocytes in XY fish increased at all samplings except F_0 with LOECs ranging from 13 µg/L to 51 µg/L.

Histological observations were also made to provide further information about both potential EDC and non-EDC effects on multiple organs including gonad, liver, and kidney. Generally, the observed pathologies in F0 were of low severity, occurred only after exposure to 102 µg/L, and were more prevalent in the kidney and liver than in the gonad. Of the pathologies reported, the majority were observed in XY individuals and could be linked to an ER agonist AOP: increased number and severity of testicular oocytes (LOEC = 13 µg/L), developmental perturbation within the gonad related to both cell types (immature spermatocysts; LOEC = 13 µg/L) and structural malformations (efferent duct; LOEC = 13 µg/L), testicular hypoplasia (*i.e.*, smaller than normal testes), and increased basophilia in the liver, presumably related to the production of vitellogenin (LOEC = 51 µg/L). There were also pathologies observed in the kidneys of XY fish that are most likely a secondary pathology induced by high concentrations of vitellogenin in the bloodstream of XY fish that normally lack a clearance mechanism for vitellogenin. There were very few, if any, pathologies that would be linked to non-EDC toxicity.

All of these findings demonstrate the MMTs potential to detect effects that are important both at the population level and at lower levels of biological organization. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 16-38.

4.1.3 4-t-Octylphenol (NIES/Japan)

In conjunction with the 4tOP MMT completed by MED and as part of a bilateral agreement, NIES (National Institute for Environmental Studies) in Japan conducted a similar MMT, again with 4tOP as the test agent. The timing of the subadult samplings were a little different (7 wpf at MED versus 8 wpf at NIES), no data were collected in F_0, and the suite of endpoints that were measured in the MED version of the MMT were reduced in the NIES MMT to just survival and SSC. Note, because the data provided by NIES were summary information and statistics, they could not be analyzed using StatCHARRMS.

Fecundity was reduced by about 80% in F_1 after exposure to the highest concentration tested, 100 µg/L, which resulted in no fish in that treatment in F_2. There were no changes in fecundity observed in F_2 (the highest concentration was 50 µg/L). These results are very similar to the MED MMT where the LOEC for fecundity was 102 µg/L in F_1 and 51 µg/L in F_2.

The secondary endpoints that provide AOP information responded as would be anticipated. SSCs in XY fish were reduced at a LOEC of 6.25 µg/L in F_1 subadults and 25 µg/L in F_2 subadults, while vitellogenin production in XY fish was induced at a LOEC of 6.25 µg/L in F_2 and only 25 µg/L in F_1. The mean hepatic vitellogenin concentration in the same treatment across F_1 and F_2 generations was not very great, 4.8 ng/mg liver in F_1 compared to 8 ng/mg liver in F_2, but F_2 was high enough to be statistically different than controls. The endpoint that seemed to respond differently between the MED and the NIES MMTs was gonad phenotype. While the MED MMT had many XY fish with 4tOP induced testicular oocytes, very few were reported in the NIES MMT so that this was a relatively insensitive endpoint. A comprehensive

report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 39-50.

4.1.4 Comparison of 4-t-Octylphenol MMTs in Support of Validation

Comparison of control parameters between MMTs can indicate whether the protocol was successfully transferred between laboratories. The 4tOP MMTs completed by EPA/MED and NIES Japan provide just such an opportunity. In Table 4-3, the only two parameters in adult control fish that were measured the same in both MMTs are compared, showing similar data across the two labs. Other data were collected in both, but are not directly comparable. For instance, vitellogenin was measured by QPCR to determine the copy number of vitellogenin mRNA in the MMT done at MED while NIES measured vitellogenin protein concentrations in the liver. Also, the subadult sampling done by MED was at 44 dpf while NIES sampled subadults at 56 dpf (or later as the data only indicates 8 wpf). During this time, medaka are undergoing rapid development and growth, thus growth, SSC, and vitellogenin production all increase dramatically within this fairly small window. So while the mean SSC in NIES subadults was 69, the mean for MED subadults was only 27; however, as seen in the table by the adult life stage, they are very similar regardless of lab.

Table 4-3. Comparison of adult control parameters in the two 4tOP MMTs.

	US EPA MED		NIES Japan	
	F1	F2	F1	F2
Fecundity	17 ± 2	20 ± 5	29 ± 5	27 ± 3
SSC	83 ± 9	130 ± 5	116 ± 12	120 ± 19

In general, the two MMTs (MED and NIES) were successful demonstrations of the protocol's reproducibility. Despite being run at different times, in different countries, and the language barrier between the two performing labs, they measured similar LOECs for reproduction and provided mutually corroborating information on the documented AOP of 4tOP (Table 4-4). These data indicate that the MMT is a reliable protocol which is reproducible between labs, a key component of the validation process.

Table 4-4. LOECs from selected endpoints from the 4tOP MMTs of ORD and NIES.

Endpoint	**US EPA/MED LOEC**	**NIES/Japan LOEC**
Fecundity	51 µg/L	100 µg/L
SSC	13 µg/L	6.25 µg/L
Vitellogenin	25 µg/L	6.25 µg/L
Gonad Phenotype	25 µg/L	100 µg/L (limited data)

4.1.5 o,p'-DDT (CERI/ Japan)

Essentially, the full research MMT was implemented by CERI (Chemicals Evaluation and Research Institute) as a part of the US-Japan bilateral agreement to test *o,p'*-DDT with only the adult growth measurements being omitted. This test agent was chosen because it is a known ER agonist that would be expected to bioaccumulate (Metcalfe et al., 2000; Uchida et al., 2010).

Reproduction was not effected by exposure in F_0, even at the highest treatment 1.9 µg/L. However, in F_1, 1.9 µg/L *o,p'*-DDT did reduce fertility (but not fecundity) mostly resulting from two pairs producing large numbers of non-viable eggs. The LOEC for fecundity and fertility in F_2 was 1.9 µg/L. In this treatment, fecundity was reduced from 20.3 eggs per pair each day to 14 while the mean number of fertile eggs in controls was 17.6 which was reduced to 6.9 fertile eggs per pair each day. Interestingly, these means for exposed breeding pairs were similar to the reproductive output in the same 1.9 µg/L treatment of F_1; however, the control mean in F_2 was higher than in F_1, presumably providing greater statistical power in the analysis of the data.

The overall LOEC for the MMT was 0.07 µg/L;; however, this was based on an increase in the hepatic vitellogenin concentration in adult XX fish. The magnitude of this increase was very modest and did not exhibit the classical dose-response: exacerbation of the response with increased concentration of the test agent. The hepatic vitellogenin concentration was also increased in adult XY fish in both F_1 and F_2, but only in the highest treatment, *i.e.*, 1.9 µg/L.

The secondary endpoints again corroborate the expectation of *o,p'*-DDT acting as an ER agonist. The numbers of anal fin papillae (SSC) were lowered by more than 50% and the concentration of hepatic vitellogenin was increased in XY subadult fish in the 1.9 µg/L treatment. The effect on SSC in subadults was lack of any observable papillae in the adult fish suggesting that this observation was due to a developmental delay. In addition, most of the XY fish exposed to 1.9 µg/L *o,p'*-DDT across both F_1 and F_2 generation and at both developmental stages had testicular oocytes while in adults in F_2 and subadults in F_1 had increased testicular oocytes as well. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 51-64.

4.1.6 4-chloro-3-methylphenol (US EPA/MED)

The full research MMT was performed with 4-chloro-3-methylphenol (CMP) as the test agent. In work done at MED, CMP binds to both recombinant human and rainbow trout (*Oncorhynchus mykiss*) ERs and induces rainbow trout vitellogenin mRNA at about the same concentrations as overt toxicity (Schmieder et al., 2004).

No effects were observed in the entire MMT that would be attributed to EDC exposure except for the possibility of vitellogenin gene expression in XX adult fish in F_0 only in which a small increase (~2X of controls) was observed starting at 88 µg/L of CMP. Despite this, reproduction was not decreased in any of the generations, and of the three significant effects documented (out of close to 50 different measurements), none were consistently observed across generations or

life stages. In addition, no indication of toxicity was observed via histopathology. Therefore, this MMT indicated that the LOEC for CMP was greater than 345 µg/L.

Post-MMT, MED performed a short-term test to establish acute lethality data for medaka to CMP to determine whether higher concentrations than 345 µg/L of CMP cause overt toxicity/mortality or EDC-related impacts. Newly hatched fish were exposed for 35 days to CMP at concentrations of 300, 600, 1200, 2400, and 4800 µg/L. Survival as the measurement of toxicity was monitored throughout the exposure and vitellogenin gene expression, as the measure of EDC effects, was quantified at the end of the exposure. The 7-day and 14-day median lethal concentrations to 50% of the fish tested (*i.e.*, LC_{50} values) were calculated as 2840 µg/L and 1090 µg/L CMP, respectively. At the end of the 35-day exposure, only fish from control, 300 µg/L, 600 µg/L, and 1200 µg/L survived. There was no change in the copy number of vitellogenin mRNA in livers of either XX or XY fish indicating CMP is overtly toxic, causing mortality, before it would induce EDC-related effects. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 65-80.

4.1.7 Tamoxifen (NIES/Japan)

Tamoxifen is classified as a selective ER modulator (SERM) as it has multiple actions both as an ER antagonist and in specific tissues/species as an ER agonist (Shang and Brown, 2002). After binding to the ER, tamoxifen has been shown to act as both agonist and antagonist in various species including the medaka, in which previous tests have shown increased plasma vitellogenin in males but decreased plasma vitellogenin in females (Chikae et al., 2004; Sun et al., 2007a).

The MMT completed by NIES in Japan testing the effects of tamoxifen was an abbreviated MMT protocol that went from F_0 adults to F_2 subadults.

Fecundity was reduced by over 50% upon exposure to 20 µg/L tamoxifen, the highest concentration tested. The AOP related endpoints were typically affected at a LOEC of 1.3 µg/L in subadults. These included increased vitellogenin and decreased SSC in XY fish. Similar to other studies (Chikae et al., 2004; Sun et al., 2007a), tamoxifen actually reduced vitellogenin in female fish, specifically starting at 1.3 µg/L. The induction of vitellogenin in males was fairly modest, a few times that of controls but not approaching the same vitellogenin concentration as normal female fish. The maximum change in SSC was seen in F_2 where the number of anal fin papillae was reduced from about 80 per XY fish to approximately 25. These same endpoints had slightly higher LOECs (from 1 to 2 treatment levels) in the adult life stage, due either to true differences in the sensitivity of the life-stage, or to random variation causing statistical, but not true biological differences in the response. Also note, the subadult sampling were not done at the same point in development, F_1 at 54 dpf and F_2 at 65 dpf, which had considerable impact on control values for these parameters.

In addition to the AOP (ER agonist; SERM) responsive endpoints, tamoxifen exposure reduced growth at 54 dpf (F_1) by about 30% in both XX and XY fish, but by less than 20% in XX fish only at 65 dpf (F_2). It is unclear from the data whether this difference is reflective of delayed maturation, where the fish evidently grow to the same size regardless of exposure, or growth is irreversibly affected by exposure to tamoxifen. Unfortunately, growth data, which would have

provided insight into this question was not recorded for adults. And, histopathology, which would have provided information both about potential EDC impacts and non-EDC toxicity at the organ/tissue level, was also not assessed. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 81-94.

4.2 Results with AR-agonists

4.2.1 17β-Trenbolone (US EPA/MED)

17β-trenbolone acts as a potent agonist of androgen receptors in a variety of systems including fish where it has been shown to reduce fecundity, vitellogenin, and generally induce masculinization of both external SSCSSC and gonad phenotypes (Ankley et al., 2003; Orn et al., 2006; Sone et al., 2005; Wilson et al., 2002). In a 21-day exposure of medaka, the LOEC for changes in SSC and decreased vitellogenin was 365 ng/L and 39.7 ng/L, respectively (Seki et al., 2006). The MMT completed by EPA-MED testing the effects of trenbolone was done with the full research MMT protocol.

The LOEC for reproduction, which occurred in $F_{1,}$ was 32 ng/L. At this concentration fecundity was reduced to zero. In addition, 13 ng/L 17β-trenbolone seemed to reduce fecundity (control fecundity equaled 25 eggs per pair-day versus ~11 eggs per pair-day in the treatment); however, it was not statistically significant. No reduction in reproduction was observed in F_2 (the highest concentration present was 13 ng/L), and the mean fecundity in the 13 ng/L treatment was equal to controls (~24 eggs per pair-day) and much higher than in the same treatment in F_1 (~11 eggs per pair-day).

The LOECs across the entire trenbolone MMT in the AOP-related endpoints ranged from 32 to 84 ng/L, predominately affecting XX fish. All the findings supported the expected response to an AR agonist, *i.e.*, the presence of anal fin papillae (the measure of SSC) in XX fish, decreased vitellogenin gene expression in XX fish, and most dramatically, XX fish presenting complete male phenotypes including functional testes. In the 84 ng/L treatment, essentially all XX fish appeared male for all measurements of phenotype: gonad phenotype, copy number of the vitellogenin gene, and SSC. In addition at 32 ng/L of 17β-trenbolone, the same effects were observed at intermediate magnitudes except gonad phenotype in which, interestingly, intersex gonads were never observed, only complete sex reversal. However, intermediate effects in gonad phenotype were observed in a slightly different manner. For instance in F_1 subadults, 16 of 16 XX fish exposed to 84 ng/L 17β-trenbolone had testes, while in the 32 ng/L treatment, only 7 of 17 XX fish had testes.

The observed pathologies further supported the AOP of 17β-trenbolone. First there were very few pathologies found in XY fish. Second, many of the pathologies observed in XX fish could be ascribed to an AR agonist exposure. For instance oocyte atresia and decreased yolk formation were found in XX subadults and adults at mild to moderate severities. There were also pathologies observed in the kidneys of XX fish at high prevalence and severity including tubule epithelial hypertrophy and eosinophilia. Ironically, 17*β*-trenbolone starting at 32 ng/L so completely caused sex reversal in the XX individuals that the pathologist observed no

pathologies in the gonads of these animals, in spite of the fact that they were XX fish with testes, albeit, histologically normal testes.

In general, the F_0 generation was less sensitive than the F_1. Because there were no viable fish available in the 32 and 84 ng/L treatments of F_2, no effects were observed in any of the measurements. In addition, the LOECs obtained at the F_1 subadult sampling were lower than those of the F_1 adult sampling. However, whether this was a result of a greater sensitivity at this life stage or a result of the statistical impact of greater sample size per replicate at the subadult sampling is difficult to determine from the data.

The suite of biological impacts on the fish within the MMT could be linked across levels of biological organization to the well documented mode of action of trenbolone as an AR agonist. At the population/organismal level, fecundity was reduced starting at 32 ng/L. This presumably reflects the effect trenbolone had on the phenotype of individual XX medaka. Starting at the 32 ng/L level, profound effects were observed on the phenotype at the individual/tissue level of organization, were individual XX fish developed into phenotypic males. This was observed across multiple levels of organization including molecular (vitellogenin), tissue (external SSCs and pathology), and organ (gonad phenotype) all tethered back to the molecular initiating event, *i.e.*, trenbolone binding and activating the AR. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 95-112.

4.3 Results with Steriodogenesis Inhibitor

4.3.1 Prochloraz (US EPA/ MED)

Prochloraz is a fungicide in which its efficacy is based on inhibition of sterol biosynthesis in target organisms thereby limiting target organisms ability to synthesize cell walls. Because of the conserved nature of the pathways involved, prochloraz also inhibits CYP19 aromatase in fish which catalyzes the conversion of testosterone to 17β-estradiol both in the ovary and the brain (Hinfray et al., 2006; Monod et al., 1993; Skolness et al., 2011). In addition, there is evidence that prochloraz acts as an AR antagonist. In the 21-day fathead minnow reproductive assay, prochloraz produces a suite of biological effects characteristic of an aromatase inhibitor and an AR antagonist including alterations in various steroid levels, reduced vitellogenin in females, and lower cumulative fecundity at 0.1 mg/L prochloraz (Ankley et al., 2005). The MMT completed by EPA-MED testing the effects of prochloraz was done with the full research MMT protocol.

The LOEC for fecundity was 41 µg/L in F_0 (the highest concentration tested); 25 µg/L in F_2, and no LOEC was determined at the highest remaining concentration (25 µg/L) in F_1. Comparing the mean fecundity in the 25 µg/L treatments of F_1 versus F_2, the means were very close: ~4 eggs per pair-day in F_1 and ~6 eggs per pair-day in F_2. However, only the 25 µg/L treatment in F_2 was significant because the difference between the exposed group and the appropriate control was different: ~10 eggs in F_1 and ~20 in F_2 as a result of lower than normal fecundity in the F_1 control breeding pairs. In fact, three of six breeding pairs in F_1 did not spawn. It is anticipated that the increased replicate structure of the medaka reproduction test (MRT), proposed as the

Tier 2 fish test (see section 5), would mitigate the reduction in statistical power caused by the random distribution of non-spawning breeding pairs throughout test.

The LOECs across the entire prochloraz MMT of the AOP descriptive endpoints ranged from 5 to 41 µg/L, affecting both XY and XX fish. Prochloraz affected growth in relatively low concentrations but the decreases observed in the subadult samples were 1) not very large in absolute terms, and 2) were not present in the adults. In fact, adults in F_0 were actually larger than controls after exposure to prochloraz. In addition, there were decreased copies of the vitellogenin gene in XX fish, which can be plausibly linked to the inhibition of aromatase, resulting in a decrease in circulating 17β-estrogen that in the end lowered hepatic vitellogenin expression. Lastly, there was a modest reduction of anal fin papillae in F_1 subadult XY fish, but this was not observed in F_2 and even more interesting, it was not observed in adults of all three generations.

Many of the pathologies observed in prochloraz MMT were not present consistently through generations or between life stages. Therefore, interpretation of the histopathology data was not as straightforward as some of the other MMTs. However, some generalities can be made. While prochloraz at the concentrations tested did not seem to have much of an impact on the tissues and organs of XY fish, it did seem to cause EDC-related pathologies in the ovaries of XX fish (*i.e.*, decreased yolk and follicular hyperplasia or hypertrophy). As more information is gathered about pathologies associated with exposure to steriodogenesis inhibitors, a clearer picture of the patterns of pathology may appear that would allow more comprehensive interpretation of histopathology datasets like the one generated in the prochloraz MMT.

Regardless, taken together, the MMT protocol was able to detect effects of prochloraz exposure on the reproduction, the population level parameter. In addition, the MMT provided information regarding the presumptive AOP that, while not as conclusive as say an ER agonist, does contribute to a WOE approach that would describe prochloraz as a steriodogenesis inhibitor. A comprehensive report including graphs and tables of the data can be found in the Appendix H: MMT Summary Reports on pp. 113-133.

4.4 Results with AR Antagonist

4.4.1 Vinclozolin (US EPA/ MED)

Vinclozolin and its degradants (M1 and M2) are anti-androgens that, via primarily through hydrolysis, also persist in the environment (the half-lives of M1 and M2 are over 180 days) (Bayley et al., 2003; Makynen et al., 2000). Vinclozolin is a dicarboximide fungicide that is commonly used on fruits, vegetables, ornamental plants, and turf grasses, and therefore is an environmentally relevant EDC. The presumptive AOP initiated by vinclozolin is binding with the androgen receptor, but not initiating the transcription of androgen-dependent genes (Martinovic et al., 2008).

Vinclozolin is readily hydrolyzed in the exposure water to M1 and M2 (Kelce et al., 1994; Wong et al., 1995) which are more potent anti-androgens. Because of this, the characteristics of the

exposure system delivering toxicant to the aquaria (*i.e.*, flow rates, residence time of toxicant, *etc*.) can quantitatively change the exposure concentrations of the metabolites. Therefore, it might be difficult to compare different aquatic bioassays with vinclozolin, especially if they have different exposure parameters.

In F_0, fecundity decreased in the 253 µg/L treatment starting in the second week of the reproduction assessment which is the third week after the start of exposure to vinclozolin. The mean for the entire three week reproduction assessment was just 7 eggs spawned per day in 253 µg/L, compared to 24 spawned eggs per day in controls. Statistically significant reductions in the numbers of spawned eggs were also seen in all the other treatment groups (17, 33, 70, and 136 µg/L) in the last week of the assessment, but these reductions were not great in absolute terms: controls at 24 eggs per day and all the treatments at approximately 15 spawned eggs per day. For comparison, in a 21-day test measuring reproduction in fathead minnows, vinclozolin reduced cumulative egg production by about a third at 60 µg/L.

In F1, there were no effects of vinclozolin exposure at the concentrations tested. The mean fecundities observed were 25, 26, 20, 29 (only 1 pair), and 26 in the controls, 17, 33, 70, and 136 µg/L, respectively. In addition, there was almost no reproduction in the 253 µg/L treatment;; however, there was only a single breeding pair in this treatment.

There was not an abundance of evidence linking vinclozolin to its AOP as an anti-androgen. The only effect on vitellogenin gene expression was increased copy number in XY subadults exposed to 136 µg/L (the highest concentration present) in F_2 only. In this group, very high measurements were made in a few individuals accounting for the statistical significance. Possibly linked to the AOP, there was a consistent decrease in anal fin papillae in XY subadults with a LOEC of 17 µg/L in F_1 and 33 µg/L in F_2. In these treatments, the SSC was reduced by a 25% and 40%, respectively. Presumably, vinclozolin exposure caused a developmental delay, as the SSCs were similar to control values in adults in both F_1 and F_2.

While growth in subadults was not affected by vinclozolin, there was increased weight and length in some of the adult sampling groups but not all. In F_0, XY fish were heavier, 275 mg in controls versus 350 mg at 33 µg/L;; however, this did not increase with higher concentrations of vinclozolin. However, XY adult fish were not affected in F_1. In addition, XX adult fish were larger (weight) in both F_0 and F_1.

There were few pathologies, fewer that were consistent across generations, and those observed were by-and-large of low severity. The most prominent pathology was tubular epithelial eosinophilia of F_1 subadult kidneys of XY fish which had a LOEC of 17 µg/L and a moderate (3) severity at the same concentration. In F_2, the same pathology at the same severity was present at the LOEC of 70 µg/L. However, there was no evidence of vinclozolin-induced changes to the gonad phenotype in the entire MMT.

It needs to be noted that there were four F_1, XY fish in the 33 µg/L treatment (the highest treatment with XY fish available for histology) that had normal ovaries, no anal fin papillae, and high vitellogenin levels (equal to normal XX levels). In addition, there were three F_2 XY fish in the 136 µg/L treatment (again the highest treatment with XY fish available for histology) also

had the same phenotypes, *i.e.*, normal ovaries, no anal fin papillae, and high vitellogenin levels. There are two explanations for these observations. First, DNA contamination into XX samples from XY fish during the fin clip/DMY assessment procedure would result in this very pattern. Second, the data could be appropriate and simply reflect a biological response to vinclozolin exposure; however, this seems unlikely because certain aspects of the complete dataset do not seem to follow the expected pattern from an EDC with a strong effect on the gonad. For instance, there were no instances of XY fish with intersex gonads in the affected treatments or the lower treatments. Additionally, there was not a suite of pathologies in the gonad or liver that would be easily related to an EDC exposure.

While our interpretation of the data points to contamination, it is impossible to make a definitive statement without a re-assessment of DMY. This can be done on the fixed, paraffin-embedded tissue remaining after histology, but we have been unable to successfully extract DNA of sufficient quality to perform the DMY assessment on these samples. We suspect that the picric acid in the Bouin's fixative interfered with DNA extraction. Appendix H: MMT Summary Reports on pp. 134-151.

4.5 Interlaboratory Evaluation of the MMT protocol - Vinclozolin

4.5.1 Background

An important aspect of the validation of a Tier 2 assay within the EDSCP is demonstrating the reliability or reproducibility of the protocol both within and between individual laboratories. To that end, the EPA Office of Science Coordination and Policy (OSCP), via contract, initiated an interlaboratory evaluation of a proposed MMT protocol that involved three distinct laboratories implementing the protocol with vinclozolin as the test agent. These studies allow the evaluation of the comparability of test responses across multiple laboratories and are necessary to fully evaluate the practical transferability of the protocol and quantitative reproducibility of the method. Three laboratories performed the interlaboratory studies and are termed Lab 1, 2, or 3.

Vinclozolin, an AR antagonist, has tested positive in the EDSP Tier 1 fish assay and would therefore be a candidate for EDSP Tier 2 testing. It was also tested in a MMT performed by EPA/MED (see 4.4.1) which provides a fourth independent assessment of the reproducibility of the protocol. Vinclozolin does pose challenges as a test agent in a validation exercise. It readily hydrolyzes in aqueous solution, producing several intermediates which are also potentially active EDCs (especially M1 and M2, (Kelce et al., 1994; Wong et al., 1995)) therefore the total endocrine effect may combine the effects of vinclozolin and its degradates. Adding to the complexity, the relative proportion of vinclozolin and its degradates is time- and pH-dependent. Thus, the residence time of the toxicant water in the dilution and delivery apparatus as well as the exposure aquaria determines the relative proportion of the parent chemical and its degradates. Another challenge posed by vinclozolin is the expected outcomes of an AR antagonist are not as well defined as other AOPs like ER or AR agonism.

A comprehensive report including graphs and tables of the data as well as a summary report provided by the lead contractor can be found in the Appendix C: Vinclozolin Interlaboratory

Report. Excerpts of the interlaboratory summary report authored by the lead contractor are used below and are indicated by the use of *italics*. Where appropriate an EPA perspective is also included.

4.5.2 Protocol

A timeline for the vinclozolin interlaboratory study is illustrated in Figure 4-1. The test spans three generations, starting with adults in F_0 (minimally 350 mg for females and 250 mg for males), going through F_1 from fertilization to 14 wpf, and finally ending the study at 8 wpf of F_2. Fish are sampled as adults of F_1 and as subadults (9 wpf) of F_1 and F_2. Each endpoint is determined in the same manner as previously described for the MMT. The activity through the course of the test is summarized below.

Figure 4-1. Timeline for the vinclozolin interlaboratory study. Three generations (F0, F1, and F2) are represented by horizontal rows; within a row each small block represents the age of the fish in weeks post fertilization (wpf). The total number of weeks of the test is indicated on the bottom of the figure.

Timeline Schematic

F0	**1**	**2**	**3**	**4**	Start with 10 to 15 wk old fish Female wt = > 350 mg; male = >250 mg																									
F1				**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**													
F2																	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**						
Study Wk	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**						

Study Weeks 1-4: Unlike the initial research MMT protocol described above, fecundity and fertility were not measured in F_0 in this later version of the MMT. Consequently, the age of the breeding adults for F_0 was not as tightly specified as it as for F_1 and F_2. The specification follows: The first generation (F0) adult breeding pairs (10 to 15 week post fertilization; females > 300 mg; males > 250 mg) are exposed for four weeks during which time developing oocytes and sperm accumulate the test chemical, associated metabolites, and possible parental factors as defined by the relative treatment levels. The breeding pairs are placed in an exposure system that has six treatment levels including controls with six replicate tanks per treatment level on Study Day 1. Each replicate tank houses a single breeding fish pair for a test total of 36 breeding pairs (6 treatments with 6 replicates per treatment). **Study Week 4:** On Study Day 22, eggs are removed from both the tanks and the females to guarantee an accurate collection the next day. The following day (Study Day 23), eggs are: a) collected and pooled from each treatment, b) systematically distributed to 6 incubators with a maximum of 20 eggs per incubator, and c) incubators are placed in one of two "incubator tanks" set up for each treatment. On Study Day 24, the process is repeated using the second incubator tank per treatment.
Study Weeks 5-6: As eggs hatch on each day, hatchlings are pooled for each treatment and systematically distributed to each of the six replicate tanks within the specific treatment. The appropriate data are collected to calculate the hatching success of each incubator and each

collection day. On Study Day 29, the F0 breeding pairs are sacrificed. On Study Day 32 and 33, the aeration (agitation) of the eggs from the 1st and 2nd days of egg collection, respectively, is turned off.

Study Week 7: On Study Day 46, the number of juvenile fish is culled to 10 per replicate tank for a total of 60 per treatment.

Study Weeks 8-10: The exposure continues with the 10 fish per replicate as they develop into sub-adults. **Study Week 11:** On Study Day 71, a small sample is taken from the caudal fin of each fish to determine the genotypic sex of the individual. This information is used to establish female-male breeding pairs. On Study Days 73 and 74 after the genotypic sex of each fish is determined, a single breeding pair per replicate is randomly established while the 8 (nominal) remaining fish per replicate are sampled for the various endpoints outlined in Table 1, sub-adults.

Study Weeks 12-13: The exposure continues as the sub-adult breeding pairs develop into adults.

Study Weeks 14-17: If the test was started on a Monday then on Tuesday through Friday of each week (Study Days 93-96, 100-103, 107-110, and 114-117), spawned eggs are counted and assessed for fertility. On Monday of each week (Study Days 92, 99, 106, and 113), all eggs, either held by the female or deposited in the tank are removed to allow for accurate counting starting the next day.

Study Week 17 (repeat of Study Week 4): On Study Day 113, eggs are removed from both the tanks and the females to guarantee an accurate collection the next day. The following day (Study Day 114), eggs are; a) collected and pooled from each treatment, b) systematically distributed to 6 incubators with a maximum of 20 eggs per incubator, and c) incubators are placed in one of two incubator aquaria set up for each treatment. On Study Day 115, the process is repeated using the second incubator tank per treatment. **Study Week 18:** On Study Day 120, the F1 breeding pairs are terminated and analyzed for the endpoints in Table 1, adults.

Study Weeks 18-19: As eggs hatch on each day, hatchlings are pooled for each treatment and systematically distributed to each of the six replicate tanks within the specific treatment. The appropriate data are collected to calculate the hatching success of each incubator and each collection day.

Study Week 20: On Study Day 137, the number of juvenile fish is culled to 10 per replicate tank for a total of 60 per treatment.

Study Weeks 21-23: The 10 fish per replicate are maintained and become sub-adults.

Study Week 24: On Study Days 164 and 165, all the fish are sampled for the various endpoints in Table 1, sub-adults.

4.5.3 Chemistry Data

Below is summary of the measured vinclozolin concentrations for each of the laboratories. Vinclozolin degradate concentrations were not measured. Two of the laboratories maintained parent vinclozolin above the nominal concentrations and fairly close to each other; however, the measured concentrations during the Lab 1 MMT were considerably lower than nominal, and more importantly, the concentrations were much lower (~50%) than the other two MMTs. Therefore, it is more appropriate to compare Lab 1 results to one treatment level lower in the Lab 2 and 3 MMTs. In addition, the measured concentrations of vinclozolin in all treatments during the Lab 2 MMT more than doubled around day 100 of the test.

Table 4-5. Dosing summary for the interlaboratory MMTs.

Nominal	Units	Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
Lab 1	µg/L	0	8.0	19	48	120	300
Lab 2	µg/L	0	8.0	19	48	120	300
Lab 3	µg/L	0	8.0	19	48	120	300
Mean Exposure		**Level 0**	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Lab 1	µg/L	<MQL [a]	4.85	12.6	30.9	81.6	197
Lab 2	µg/L	<MQL [b]	11.6	25.2	62.0	155	360
Lab 3	µg/L	<MQL [c]	8.9	21.7	52.3	133	333
% CV		**Level 0**	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Lab 1		NA	0.88	1.70	6.21	10.1	29.2
Lab 2		NA	49.9	36.5	30.8	32.8	24.7
Lab 3		NA	14.6	10.1	10.9	10.4	8.5
% of Nominal		**Level 0**	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Lab 1		NA	61	66	64	68	66
Lab 2		NA	145	133	129	129	120
Lab 3		NA	111	114	109	111	111

[a] Minimum Quantifiable Limit (MQL) = 2.28 µg//L
[b] Minimum Quantifiable Limit (MQL) = 1.14 µg//L
[c] Minimum Quantifiable Limit (MQL) = 5.0 µg//L

4.5.4 Control Data Supporting Validation

Comparison of control data from each of the laboratory studies provides a measure of the natural variation for each of the variables and is summarized in Table 4-6. This analysis indicated that for many endpoints, the control response was not consistent between laboratories. The median responses from F_1 reproductive measurements, F_1, F_2 length and weight measurements were all significantly different ($p < 0.01$) between laboratories. Closer inspection of the results suggest some of the significant inter-lab variability can be attributed to the relatively small size at termination of the Lab 2 F1 and F_2 fish. The small size of Lab 2 F_1 females appears to have contributed to the high inter-lab variability in fecundity. Normalizing F_1 fecundity for the body mass of the female (F_1 total eggs / day / g day 84 female weight) eliminated the significant difference ($p=0.463$). Two endpoints that showed greater consistency, which were not statistically different between labs was F_1 and F_2 anal fin processes and hepatic expression of VTG-1. However, it should be noted that successful measurement of anal fin

processes only occurred for two labs in both the F_1 and F_2 generation, limiting inter-lab comparisons.

Table 4-6. Median Control Response by Chemical and the Kruskal-Wallis Test

Variable	Lab 1	Lab 2	Lab 3	Kruskal-Wallis p-value
F_1 Total Eggs/Day	27.1	10.02	34.71	0.005
F_1 Total Eggs/Day/ day 84 female weight	170.4	121.5	169.4	0.463
F_1 Viable Eggs/Day	21.73	9.68	33.67	0.005
F_1 Proportion Viable	0.757	0.9614	0.97086	0.044
F_1-Female Day 80/84 Weight (g)	0.161	0.089	0.207	0.002
F_1 Female Day 80/84 Length (mm)	--	19.799	26.625	0.01
F_1 Female Day 84 Total Processes	7.8	0	--	
F_1-Female Day 80/84 copies VTG/ng RNA	22289	1195	17470	0.076
F_1 Male Day 80/84 Weight (g)	0.127	0.107	0.199	0.001
F_1 Male Day 80/84 Length (mm)	--	21.78	26.01	0.004
F_1 Male Day 84 Total Processes	26.02	14.73	--	0.423
F_1 Male Day 80/84 copies VTG/ng RNA	2.25	27	1850	0.075
F_1 Female Day 119 Weight (g)	--	0.192	0.419	0.006
F_1 Female Day 119 Length (mm)	--	27.08	32.4	0.006
F_1 Feale Day 119 Total Processes	0	0	--	
F_1 Male Day 119 Weight (g)	--	0.227	0.389	0.004
F_1 Male Day 119 Length (mm)	--	29.38	32.33	0.005
F_1 Male Day 119 Total Processes	73.5	53.33	--	0.025
F_2 Female Termination Weight (g)	0.211	0.034	0.226	0.002
F_2 Female Termination Length (mm)	--	15.4	27.3	0.004
F_2 Female Termination Total Processes	10.1	--	1.43	0.40
F_2 Female Termination copies VTG/ng RNA	23650	0.5	43273	0.103
F_2 Male Termination Weight (g)	0.193	0.028	0.222	0.002
F_2 Male Termination Length (mm)	--	14.7	28.1	0.004
F_2 Male Termination Total Processes	52.9	--	59.1	1.0
F_2 Male Termination copies VTG/ng RNA	1.22	2.50	1.75	0.912

There are two issues that are plausible explanations for the control parameter variation between laboratories. First, as mentioned above, is the small size of F_1 and F_2 fish in the Lab 2 MMT. Certainly these fish were not developing like the fish in the other two labs or with what would be expected to be a normal growth rate. Because of this, all control parameters of the Lab 2 MMT fish would not be expected to meet minimum performance criteria. Also, most of the statistical differences between MMT control fish are a result of Lab 2 fish being so much different than Lab 1 and Lab 3 fish. Second, as mentioned in the interlaboratory summary report found in Appendix C: Vinclozolin Interlaboratory Report, there was a high rate of unexplained genotype to phenotype mismatch not only in control fish but across all treatments. Given the prevalence

of the mismatches, and that both *XX-male*, and *XY-female* mismatches were observed, points to technical problem in the determination of the genetic sex of the fish. This may have been due to problems with sample/label fidelity during tissue collection and processing. Errors in determining the genetic sex of individual fish would likely skew many of the control parameters, as all of them show some level of sexual dimorphism.

With respect to the issues identified above, EPA believes that the control data do indeed support the reproducibility of the protocol as long as genotypic sex data (DMY) are obtained and acceptable growth of control fish is achieved. At this time, these are presumed to be technical issues involving the appropriate rearing of the animals and sample handling procedures. However, the apparent flaws in the data from the interlaboratory also further highlight the need for ongoing improvement in the detailed methods for accomplishing the protocol and specific training leading to more effective transfer of technical skills required to perform the MMT protocol successfully.

4.5.6 Summary of Lab 1 Vinclozolin MMT

In this MMT, vinclozolin did not lower the fecundity rate or the fertility (either the number of fertile eggs or percent fertile eggs) of the breeding pairs in any of the treatments of either F_0 or F_1. The F_0 reproductive assessment was not part of the original study plan, but during the course of the F_0 loading phase, fecundity data were collected for six consecutive days and fertility data were collected for only two consecutive days. In F_1, both fecundity and fertility were assessed, according to the protocol, for 21 consecutive days.

The only affected endpoint that presumably would reflect an EDC exposure was SSC (anal fin papillae). XY fish, across both F_1 and F_2 and in both subadults and adults, had fewer papillae than controls. Because genetic sex of individual fish was unreliable, analysis of treatment effects on SSC had to be by comparisons to other phenotypic sex markers (*i.e., either the gonad phenotype or the secondary sex phenotype*), and to be completely accurate, these fish are referred to as "males" or "females", instead of the preferred assignment of "XX" or "XY". Vinclozolin-treated subadult and adult male fish, as defined by the gonad or SSC phenotype, in both F_1 and F_2, had fewer anal fin papillae than controls. It was not possible to observe changes in phenotypic females, as increases in the number of anal fin papillae would define an individual as a phenotypic male. However, in the phenotypic males, there was a decrease in the number of papillae in response to vinclozolin exposure with a LOEC of 31 µg/L, and continuing in both the 82 and 197 µg/L treatments with similar effect magnitudes. The control males had a mean of 77 papillae while in these three treatments (31, 82, and 197 µg/L), the mean number of papillae were 59, 53, and 52, respectively. In F_1 subadult males, vinclozolin at a concentration of 197 µg/L reduced the mean number of anal fin papillae from 32 as seen in controls to a mean of 20 per fish. In F_2, the LOEC was 5 µg/L and the reduction of papillae consistently decreased in a dose-dependent manner to the highest concentration (197 µg/L). For unknown reasons, the mean number of papillae in F_1 control males was much lower than seen in the male controls of F_2 (32 versus 73). The mean in males for each of the vinclozolin exposed groups of F_1 was actually similar (or often lower) to the corresponding treatment group of F_2. Therefore, the LOEC and the apparent increased sensitivity of F_2 may have been a result of the further developed fish, as

measured by SSC, in F_2 as compared to the F_1 fish. A comprehensive report including graphs and tables of the data as well as a summary report can be found in the Appendix C: Vinclozolin Interlaboratory Report.

4.5.7 Summary of Lab 2 Vinclozolin MMT

In this MMT, vinclozolin did not lower the fecundity rate or the fertility (either the number of fertile eggs or percent fertile eggs) of the breeding pairs in any of the treatments of F_1, where both fecundity and fertility were assessed, as per the protocol, for 21 consecutive days. The mean fecundities observed were 10, 10, 8, 9, 7, and 7 in the controls, 12, 25, 62, 155, and 360 µg/L, respectively. In addition, the fertility was high across all the treatments.

In the other endpoints, vitellogenin gene expression was induced in XX fish of F1 only in the highest treatment tested (360 µg/L) from a mean of 830 copies per ng of total RNA to 8910 copies. Growth (weight) was increased by vinclozolin exposure in XX fish of both generations (F_1 and F_2) with a LOEC of 12 µg/L (the lowest treatment), and in XY fish in F_1 only with a LOEC of 62 µg/L.

4.5.8 Summary of Lab 3 Vinclozolin MMT

In the primary endpoint of the MMT, reproduction, no changes in fecundity or fertility were observed upon exposure to vinclozolin. In F_0, fecundity and fertility was based upon counts of pooled eggs. Therefore the box plots presented represent inter-day variance of the entire treatment and not inter-breeding pair variance. To avoid confusion, the individual data, daily counts of whole treatments, were not plotted on the graph for F_0. While there was not a statistically significant reduction in fecundity or fertility, there was a trend of lower fertility with vinclozolin exposure. In F_1, there were no effects of vinclozolin exposure on either fecundity or fertility at the concentrations tested. The mean fecundities observed were 35, 28, 34, 33, 28, and 37, in the control, 9 µg/L, 22 µg/L, 52 µg/L, 133 µg/L, and 333 µg/L treatments, respectively. The fertility was high across all the treatments ranging from 92% to 98%.

Of the other endpoints, anal fin papillae were reduced in XY fish in F_2 subadults only. Unfortunately, there was a mistake in sample handling with the F_1 subadult anal fins so that no counts of anal fin papillae were possible. Vinclozolin reduced the mean number of papillae in the XY subadult fish of F_2 at all concentrations tested (9 – 333 µg/L). The mean number of papillae was 57 in the control and was reduced to a mean of 39, 42, 40, 33 and 19 papillae in the 9 µg/L, 22 µg/L, 52 µg/L, 133 µg/L, and 333 µg/L treatments, respectively.

4.5.9 EPA Perspective on the Interlaboratory Results

Overall the interlaboratory was a success insomuch as all three labs did not see decreased reproduction at the concentrations of vinclozolin tested. This is also unfortunate for the very same reason. Other endpoints were not affected by vinclozolin with the exception of SSC in which was, when available for assessment, negatively responsive to vinclozolin exposure, and in

the Lab 2 MMT only, sporadic increases in weight and vitellogenin were observed. As a note, SSC was also decreased in the vinclozolin MMT performed by EPA/MED. So while few significant effects were observed in the three interlaboratory MMTs, there was consistency in the observation of effects on the SSC. In addition, the control data suggest the protocol is reproducible even though most of the measurements had statistical differences between the three labs. Recall, one lab had unusually poor control growth that seemed to cause developmental delay in control fish, which, in turn resulted in poor control reproduction. When those data are removed, there was considerable consistency between the data of the other two laboratories.

Unfortunately, the previously discussed issues with the interlaboratory data cannot simply be ignored, and they highlight that this is a relatively complicated procedure that even with experience is not a trivial undertaking. Many of the directives within the protocol are derived from extensive experience at EPA/MED lab, and without *a priori* knowledge, should be followed without deviation. Otherwise, there is a relatively high risk of generating poor quality data with high variances. The consequence is that it is difficult to discern treatment-related effects with all the endpoints, but especially on reproduction. Low reproduction in controls seriously compromises the statistical power of the MMT to detect decreases in fecundity or fertility. In addition, knowing the genetic sex (DMY) of each fish is an incredibly powerful tool for analysis of the data. To maximize the statistical power for each endpoint requires this information. Proficiency, both in DNA preparation and in the determination of the presence or absence of DMY, must be achieved prior to implementation of the MMT protocol.

The suite of expertise required to perform a MMT is broad, and there are many nuances that need to be recognized by performing labs. The limitations in the interlaboratory trial data are not indicative of low protocol reproducibility, but rather to an inadequate transfer of the technical details necessary to successfully execute the MMT.

5 OPTIMIZATION OF THE MMT AND THE PROPOSED MEDAKA REPRODUCTION TEST (MRT) PROTOCOL

5.1 Primary MMT Protocol Changes for development of MRT Protocol

The outcomes of the various MMT protocols with several different chemicals performed at the EPA MED lab and by the NIES and CERI laboratories in Japan, have provided enough information to provide guidance for the development of a medaka reproduction test (MRT) for use as a fish test in Tier 2 of the EDSP. Two major changes are proposed, *i.e.,* an increase in the number of replicates per treatment for evaluating effects on reproduction, and terminating the test after the embryos hatch in F_2. Other proposed changes include minimizing the collection of endpoint data from F_0, and evaluating pathology in only the F_1 adults sampled after the assessment of reproduction. Figure 5-1 presents the proposed timeline and endpoint data for the MRT. The test conditions for the proposed MRT protocol are presented in Table 5-1. Endpoint measurements are essentially identical to those employed in the full MMT except that the sagittal sectioning plane is recommended for the pathology assessment. Comparisons of various aspects of the research MMT and the proposed MRT are presented Table 5-2. The rationale for the recommended changes in the MMT for the proposed Tier 2 fish test of the EDSP, are presented

in the following sections. Detailed protocols of the Proposed Medaka Reproduction Test are presented in Appendix E: MRT Protocol.

Figure 5-1. Proposed Medaka Reproduction Test (MRT) for Tier 2 The timeline and endpoints in the MRT. The duration of the in-life exposure is 19 weeks and spans 1 full generation (F_1). The F_0 generation serves as a loading phase for the F_1 embryos. The test is terminated after the F_2 embryos hatch.

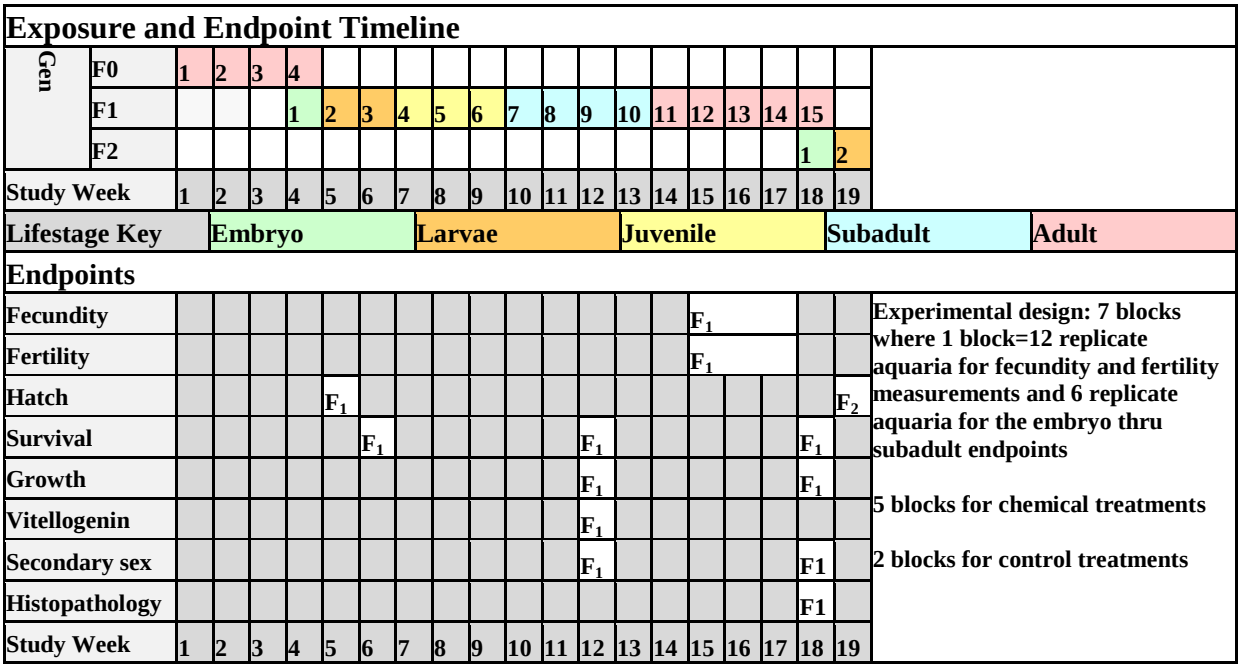


Table 5-1. Test Conditions for the Proposed MRT

Recommended species	Medaka (*Oryzias latipes*)
Test type	Continuous flow-through
Water temperature	25.5 ± 1°C
Illumination quality	Fluorescent bulbs (wide spectrum and ~150 lumens)
Photoperiod	16 h light, 8 h dark
Loading rate	F_0- 2 fish/replicate; F_1 and F_2- initiated with maximum of 20 larvae/replicate eventually culled to 2 fish/replicate
Test replicate chamber size	18x9x15 cm
Test replicate solution volume	~1.8 liters
Volume exchanges of test solutions	16 volume replacements/day (20 ml/min flow).
Age of test organisms at initiation	F_0-11 weeks post-fertilization, F_1 and F_2 continuously from fertilization
No. of organisms per replicate	F_0- 2 adults; F_1 and F_2- initiated with maximum of 20 larval fish eventually culled to 2 adults
No. of treatments	6 (5 toxicant treatments plus 1 control, and, a solvent control if needed)
No. replicates per treatment	6 per treatment for each toxicant treatment; 12 for control treatment; replication structure doubled for fecundity assessment
No. of organisms per test concentration	F_0- 24 for each test concentration and 48 for controls; F_1 and F_2- start with a maximum of 120 and eventually cull to 48 for fecundity assessment.
Feeding regime	Fish were fed 24-hour old brine shrimp (*Artemia* spp.) nauplii at the following rates:

Time (post-	Brine Shrimp (mg dry	Time (post-	Brine Shrimp (mg dry weight/fish/day)

hatch)	weight/fish/day)	hatch)	
Day 1	0.5	Day 12	4.2
Day 2	0.5	Day 13	4.5
Day 3	0.6	Day 14	4.8
Day 4	0.7	Day 15	5.2
Day 5	1.0	Day 16-19	5.6
Day 6	1.3	Week 4	7.7
Day 7	1.7	Week 5	9.0
Day 8	2.2	Week 6	11.0
Day 9	2.8	Week 7	13.5
Day 10	3.5	Week 8-15	22.5
Day 11	4.2		

Aeration	None unless DO reaches <60% saturation
Dilution water	Filtered (5 micron), sterilized fresh surface water, or appropriately conditioned well water.
Exposure period	Trans-generational test: F_0 from 11-15 wpf, F_1 and F_2 from 0-15 wpf
Primary endpoints	Survival: F_1 from 1-3 wpf and from 3-15 wpf Growth : length, weight, (F_1 at 9 wpf and 15 wpf Hatch: $F_1$and F_2 only Anal Fin Papillae: F_1 assessed at 9 wpf and 15 wpf Fecundity: F_1 assessed daily for 21 days, 12-14 wpf Fertility: F_1 assessed daily for 21 days, 12-14 wpf Vitellogenin: F_1 only at 9wpf
Optional endpoints	Time to first spawn for F_1 generation.
Test acceptability	Dissolved oxygen $\geq$60% of saturation; mean temperature of 25.5 ± 1°C; ≥70% survival in control treatments of each generation; successful reproduction in at least 65% of control females with mean daily fecundity in the control treatment of at least 15 eggs.

Table 5-2 Comparison of the Research MMT to the Proposed Tier 2 Fish Reproduction Protocol

	Research MMT	Proposed MRT	Rationale
Replication structure	6 replicates/trtmnt	12 replicates/trtmnt 24 control replicates	Statistical power for reproductive analysis
Generations	2 ½ (F_0, F_1, partial F_2)	1 ½ (F_0, F_1, F_2 hatch only)	Reduce test time, chemical amount, cost, risk of test failure
			F_0 for dose-level evaluation and to expose developing gametes with test chemical
F_0 endpoints	Reproduction, Growth, Survival, Vtg, SSC, Pathology	Reproduction – qualitative only	Adequate dose-level evaluation
		Survival	Adequate dose-level evaluation
		No other endpoints measured	F_0 < sensitive than F_1 and F_2
F_2 endpoints	Hatch, Growth, Survival, Vtg, SSC, Reproduction, Pathology	Hatch only	Ensure adequate time for deposition of test agent to egg.

		No other endpoints measured	Substantial and consistent sensitivity differences between F_1 and F_2 were not observed for the chemicals tested
F_1 endpoints	Hatch, Growth, Survival Subadult: Vtg, SSC, pathology	Hatch, Growth, Survival Subadult: Vtg, SSC	Increased replication allows evaluation of biomarker endpoints with good statistical power
	Adult: Reproduction, Pathology	Adult: Reproduction, Pathology	Increased replication increases statistical power for reproduction and pathology evaluation
		Adult: Reproduction	Three weeks continuous (21 consecutive days) vs. 5 weeks (T,W, Th, FF only). Better for statistics and test duration
Temperature	26	25.5	Maximize development rate; minimize prevalence of XX males
Pathology	F_0 adult F_1, F_2 subadult and adult	F_1 adult	Increased power for pathology analysis, pathology correlated with pair reproduction outcomes

5.1.1 Increased Replication for the Assessment of Treatment-Effects on Reproduction

The reproductive endpoints of fecundity and fertility are important endpoints for ecotoxicology assessments, due to their implicit direct relationship to effects at the population-level. To assess treatment effects on adult reproduction, the full MMT protocol employed 6 replicate aquaria for each test treatment, including the controls. Fecundity data from the full MMTs had relatively high variance. Thus, with only 6 breeding replicates per treatment, the statistical power for detecting significant changes in fecundity was quite low. Some of the modifications of the protocol were implemented to ostensibly increase the power of the protocol either by increasing baseline fecundity or by reducing the variance in egg production. Such modifications included increasing the test temperature, increasing growth with feeding rates or by increasing the development time before assessing reproduction. These modifications were only marginally successful.

After running a number of MMTs and evaluating the reproduction data, it became clear that the statistical power for detecting significant effects on fecundity in the full MMT was quite low. The trends observed in the dose-responses for fecundity suggested that fecundity was being affected at treatment levels lower than the statistical LOEC. In some cases, the fecundity had to be reduced to nearly zero to be significantly different from the control fecundity values. It

appeared that the most reliable method to increase the statistical power of the test to detect treatment effects on reproduction was through increasing the number of breeding pairs.

To determine the number of replicates required a power analysis was performed using the reproductive performance of control fish from several of the MMTs run at the MED laboratory (Appendix I: Power Analysis). MMT fecundity data have characteristics that complicate the calculations of statistical power. For example, the data are not transformable to normally distributed data; they are measured repeatedly over time with several replicates in each treatment; and days without reproduction (*i.e., where the number of spawned eggs is zero)* are randomly distributed in the data set. Because of these characteristics, a numerical solution for power was calculated using Monte Carlo simulation methods.

In order to design the Monte Carlo simulation, various assumptions were made including that the between aquarium within treatment mean fecundity is normally distributed, the within pair fecundity follows a Poisson distribution (typical of count data), the control fecundity does not change during the reproductive assessment, and finally the rate at which individual pairs produce zero eggs is constant. This last assumption is important because the experience of the MED laboratory has been that a certain percent of breeding pairs will not successfully spawn for unknown reasons, and this can be daily, weekly, or for the entirety of the assessment.

Using these assumptions and the control data from previous MMTs, the following parameters were calculated for the Monte Carlo simulation: the population mean = 21 eggs and the population variance = 49. Also the rate at which the fecundity of a pair was zero can be separated into three categories; one day, one week (greater than one day but less than 8 days), and always (greater than 8 days). These were estimated to be 8/91 for one day, 5/91 for one week, and 3/91 for all days.

The data were simulated in the R programming language incorporating the assumptions above to generate a fecundity dataset of control plus 5 pre-assigned reductions in fecundity (*i.e.*, treatments) of 20%, 30% 40%, 50% and 70% reduction. This process was repeated 1000 times for each pair and percent reduction combination. Next, all of the simulated data were analyzed by the repeated measures ANOVA of the StatCHARRMS program. Two different replicate structures were analyzed: 1) the number of replicates is equal in all treatments including the controls (balanced design), and 2) based upon statistical recommendations from the OECD (OECD, 2006), the number of control replicates was double the number of replicates in each chemical treatment (2:1 design). The 2:1 design had substantially greater power and is the only analysis presented here.

Graphic results are illustrated in Figure 5-2. Note that at 6 replicates (12 replicates in the control), there is a small but noticeable probability of not detecting a reduction of 50%, about a 75% probability at detecting a reduction of 40%, less than 50% chance at detecting a reduction of 30%, and very little chance at detecting a reduction of 20%. At 12 replicates (24 replicates in the control) 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%.

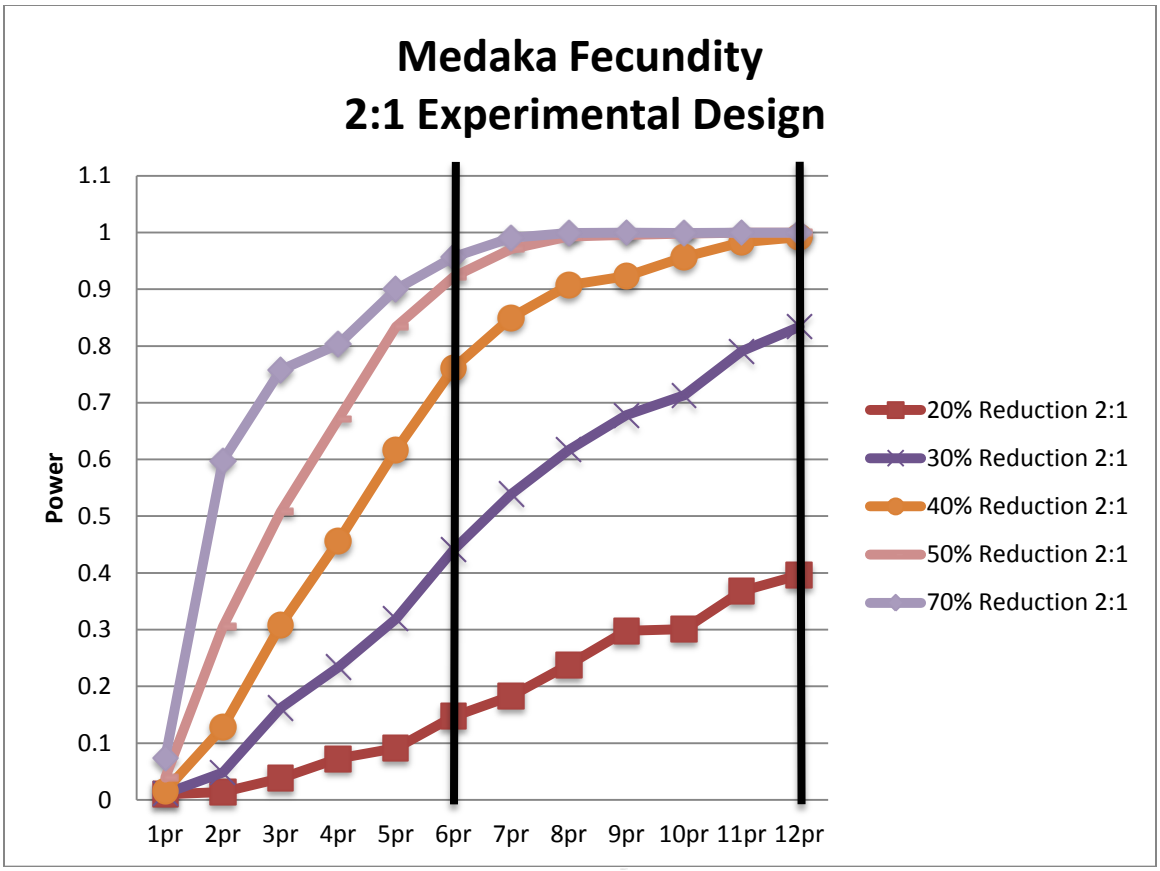


Figure 5-2. Monte Carlo simulation of the statistical power to detect proportionate reductions in medaka fecundity with different numbers of replicate breeding pairs. The experimental design represented is the 2:1 design where control replicates are double of the number of replicates in the chemical treatments. The power is the probability of detecting a treatment effects compared to the controls at an α-level of 0.05.

During the discussions on revising the replicate structure of the MMT, consideration was given not only to the power analysis, but also to the possibility of mortality, especially in the control replicates. It has been the experience of MED researchers that a very small percentage of the adults, irrespective of treatment, may die, and in addition, 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.

Reproduction Results from Two Octylphenol MMTs: 6 vs. 12 Replicate Breeding Pairs

To evaluate the effect of increasing the replication structure of the MMT, two tests (designated PTOP-02 and PTOP-03) were run at the EPA MED lab with the weak ER agonist 4-*tert*-octylphenol. PTOP-02, which employed 1 breeding pair in each of 6 replicate aquaria per treatment, was started late in 2007. PTOP-03, which was started late in 2012, employed 12 replicate aquaria per treatment, except for the control treatment which had 24 replicates. The chemical concentrations were nearly identical for the two tests (see Table 5-3). Reproductive assessments were similar except that for PTOP-02 fecundity and fertility were measured 4 days per week for 5 weeks and for PTOP-03 the measurements were taken for 21 consecutive days.

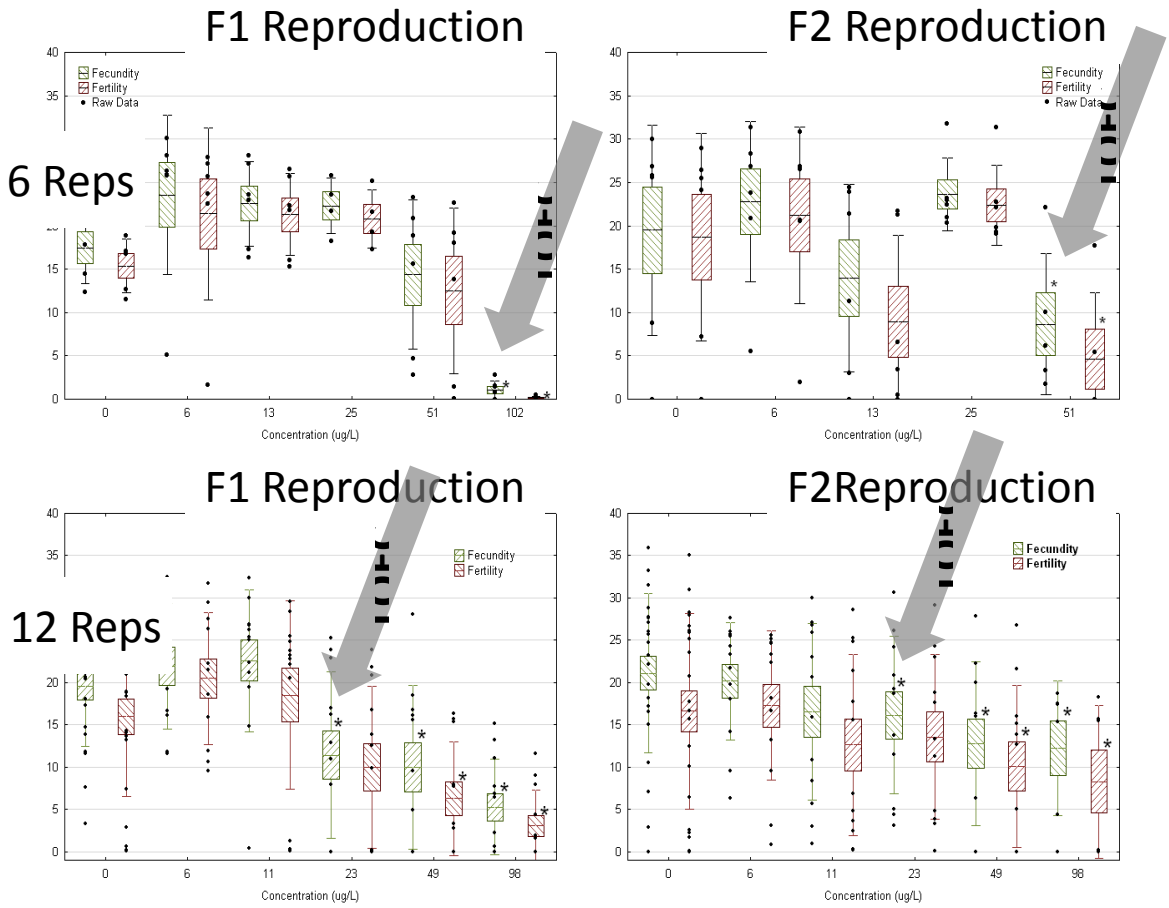
There were no treatment effects on F_0 reproduction in PTOP-02 (Appendix H: MMT Summary Reports on pp. 16-38) and reproduction endpoints were not measured in PTOP-03.

Table 5-3 Comparative Treatment Concentrations for two 4-*t*-Octylphenol MMTs

Treatment	PTOP-02 Conc (µg/L) and (S.D.)	PTOP-03 Conc (µg/L)*
1	0 (0)	0
2	6 (0.6)	6
3	13 (1.5)	11
4	25 (2.6	23
5	51 (5.6)	49
6	102 (8.5)	98

*: S.D not calculated for this test yet

Figure 5-3. Reproduction Results from 6- and 12- Replicate Test Protocols



The primary effect of octylphenol on reproduction is to reduce the fecundity and not the fertility of treated fish. Note also that the control means for fecundity are similar in F_1 and F_2 for both

tests and range from 17 to 21 eggs/pair/day. The LOEC for F_1 fecundity in the 6 rep design was 102µg/L which reduced the fecundity to nearly 0. Because of this, there were insufficient survivors to populate that treatment in F_2. The LOEC for F_2 fecundity in the same test was 51µg/L. There were no significant differences between the mean fecundity at any particular treatment level across generations (*i.e.*, F_1 control fecundity of 17 vs. F_2 control fecundity of 19.5; or F_1 51µg/L fecundity of 14 vs. F_2 51µg/L fecundity of 8).

In the 12 replicate test, the LOEC for fecundity was 23µg/L for both the F_1 and F_2 generations, one treatment level lower than the LOEC for the 6 replicate test. Note also that the reduction in fecundity observed at the LOEC was only about 7.5 eggs/pair/day in F_1 and 5 eggs/pair/day in F_2. In summary, doubling the number of replicate breeding pairs increased the statistical power of the test to detect significant treatment effects on fecundity.

The 12 replicates per treatment design also has increased power due to the fact that there are two blocks of 12 replicates for the control treatment. There are a total of 84 replicate aquaria in the reproductive phase of the test; ((5 chemical treatments) x 12 replicates + (2 control blocks x 12 replicates) = 84). While this may seem excessive, this number of replicates is necessary due to the highly variable nature of fish reproduction. Furthermore, this design provides additional insurance for the baseline probability of having non-spawning pairs within the test population and for the inevitable loss of replicates due to technical accidents. The difference in the statistical power for fecundity in the MRT with 12 breeding replicates for each of the chemical treatments and 24 replicates in the controls is presented in Table 5-4. Note that in the MMT a 40% reduction in fecundity will be detected with a probability of 66 %, and a 28% probability of detecting a 30% reduction in fecundity. In contrast, with the MRT the probability of detecting 40% and 30% reductions in fecundity are 99% and 83% respectively.

Table 5-4 Statistical Power Comparison for Fecundity in the MMT and the proposed MRT

Fecundity Reduction (%)	**MMT**	**MRT**
	Balanced design 6 replicates	**2:1 design 12 replicates**
20	0.17	0.40
30	0.28	0.83
40	0.66	0.99
50	0.83	1
70	0.91	1

When considering the number of fish employed in the two different test designs, it is important to note that there are actually fewer fish used in the 12-replicate MRT design. First, the initial phase of the protocol employs 6 replicates per treatment. And, the initial number of embryos placed in incubators has been reduced from 30 to 20. Then, when starting the reproductive phase of the test at 9 wpf, 4 fish per replicate are used to set up two breeding pairs rather than the 2 fish for the single pair per replicate employed in the used in the 6-replicate design of the MMT. The increase in statistical power afforded by the 12 replicate design is accomplished by reallocating fish from the 9 wpf sampling into the analysis of treatment effects on reproduction. The concomitant reduction in the statistical power of the non-reproductive endpoints (*i.e.*, 6 fish per

replicate in the MRT compared with 8 per replicate in the MMT) is relatively small. Specifically, Table 5-5 presents the statistical power of the MRT to detect changes in the reductions in the biomarker endpoints (*i.e.*, wet weight, length, vitellogenin and SSC) in fish sampled at 9 wpf. Note that except for the female vitellogenin endpoint, the power to detect treatment effects in these endpoints in the 9 wpf samples is similar or higher than the power to detect effects in fecundity. Thus the statistical power of the endpoints appears to be reasonably balanced in experimental design of the proposed MRT (Appendix I: Power Analysis).

Table 5-5. Statistical Power of Non-reproductive Endpoints in Proposed MRT

Statistical power of MRT to detect proportionate differences in non-reproductive endpoints compared with control values (α=0.05)						
	10%	**20%**	**30%**	**40%**	**50%**	**Trend**
	Wet Weight Reduction					
Female	0.14	0.67	1.00	1.00	1.00	1.00
Male	0.10	0.53	0.95	0.99	1.00	1.00
	Length Reduction					
Female	0.99	1.00	1.00	1.00	1.00	1.00
Male	0.99	1.00	1.00	1.00	1.00	1.00
	Secondary Sex (reductions in anal fin papillary processes)					
Male	0.01	0.25	0.87	0.99	1.00	1.00
	Vitellogenin Reduction					
Female	0.01	0.09	0.59	0.99	1.00	1.00
	Vitellogenin Increase					
	2X	5X	10X	20X	50X	Trend
Male	0.10	0.62	0.93	1.00	1.00	1.00

Comparison of Animal Use between the MMT and MRT

An evaluation of the number of fish that are used in the research MMT protocol compared with the proposed MRT indicates that even though the number of replicate breeding pairs is doubled in the chemical treatments and is 4x high in the controls, the total number of fish sampled and culled is less in the MRT; 2,520 fish in the MRT compared with 2,952 in the MMT. Furthermore, if one discounts the portion of newly hatched, as suggested by the OECD guidelines for fish embryo toxicity tests, eleutheroembryos which are not continued to the juvenile and adult stages in this calculation, then there are far fewer fish employed in the MRT; 558 juvenile and/or adult fish in the MRT compared with the same 2,952 juvenile and adult fish used in the research MMT.

Estimates of person-hours required to perform the MMT and MRT

Two critical changes in the MRT compared with the MMT are the recommended increase in the number of breeding replicates employed and, the completion of the in-life potion of the test after embryo hatch in F_2. The increase in replication requires an increase in the human resources necessary to perform the MRT, especially during the assessment of fecundity and fertility. In contrast, the decrease in the exposure time reduces the necessary human resources to perform the

test. A comparison of the time required to perform the MRT and the MMT is presented in Table 5-6.

Table 5-6. Human-Resource Estimates for In-life Portions of MMT and MRT (hours)

Resource Category	MMT	MRT
Technician Hours	1197	939
Biologist Hours	424	297
Chemist Hours	590	382
Total Hours	2211	1618

Note that even with the replicate increase, the estimate of the total human resource time required to run the in-life portion of the MRT is about 25% less than the MMT.

5.1.2 In-life Exposure Complete After F_2 Hatch

The context of this discussion is whether the data from the F_2 generation added significant value to the understanding of treatment effects on the test population. A summary of the data from all the MMTs performed by EPA or Japanese labs that included assessments of fecundity in both F_1 and F_2 is presented in Table 5-7. With 17β-estradiol the LOEC for fecundity was lower by 1 treatment level in F_1 compared to F_2. In one of the three tests with 4-*t*-octylphenol, the LOEC in F_2 was 1 treatment level lower and in the other two the LOECs were the same in both generations. Note that the LOEC in the 2012 octylphenol test, which had 12 breeding pairs per treatment, was 1 treatment level lower than the previous test run by EPA, and that the LOECs were the same in F_1 and F_2. The LOEC for the Japanese run o,p′-DDT test was not determinable with a *greater-than* 1.9 µg/L value for F_1 and 1.9 µg/L in F_2. Lastly, the LOECs for fecundity in F_1 and F_2 for the 17β-trenbolone test run by EPA were equal. In summary, for the six tests using three different chemicals where effects on fecundity were measured in both the F_1 and F_2 generations, the LOEC was lower in F_1 in one test; lower in F_2 in two tests, and the same in three tests. In all cases, when the LOECs were different across generations, it was by one treatment level. In the 2012 4-*t*-octylphenol test with 12 replicate breeding pairs, the LOECs were the same in both generations and lower than the LOECs measured in the other two MMTs with octylphenol. An MMT with more replication in the assessment of reproductive endpoints, and without an assessment of reproduction in F_2 (*i.e.*, the proposed MRT) appears to detect effects on fecundity as well or better than the information obtained in both generations of the research MMT. Not only was the LOEC lower with 12 breeding pairs, but the LOEC in F_1 and F_2 were the same. Additionally, the mean fecundity values in the control treatments and the LOEC treatments were quite similar across generations. Finally, differences of the mean fecundity levels in a particular treatment across generations were not statistically significant.

Table 5-7. LOECs for Fecundity in F1 compared with F2 in MMTs

Chemical	Laboratory	Fecundity LOEC	
		Data through F1	**Data through** F2
17β-Estradiol	CERI/Japan	28 ng/L	>28 ng/L
4-*t*-Octylphenol	US EPA/MED-2007	102 µg/L	51 µg/L
4-*t*-Octylphenol	NIES/Japan	100 µg/L	100 µg/L (>50)
4-*t*-Octylphenol	US EPA/MED-2012 (12 reps)	23 µg/L	23 µg/L
o,p′-DDT	CERI/Japan	>1.9 µg/L	1.9 µg/L
17β-Trenbolone	US EPA/MED	32 ng/L	32 ng/L

Comparisons of the lowest LOECs for the biomarker endpoints obtained from the subadult samples in 6 MMTs is presented in Table 5-8. Note that the LOECs for the most sensitive biomarker endpoints were the same in F_1 and F_2 except for one case, *i.e.*, the US-run test with octylphenol, were the LOEC for anal fin papillae was one treatment level lower in F_2 than in F_1.

Table 5-8. LOECs for Biomarker endpoints in F1 compared to F2 in MMTs

Chemical	**Laboratory**	**Biomarker LOEC**		**Endpoint**
		Data through F_1	**Data through F_2**	
17β-Estradiol	CERI/Japan	2.8 ng/L	2.8 ng/L	Papillae
4-*t*-Octylphenol	US EPA/MED	25 µg/L	13 µg/L	Papillae
4-*t*-Octylphenol	NIES/Japan	6.25 µg/L	6.25 µg/L	Liver vitellogenin
o,p′-DDT	CERI/Japan	0.07 µg/L	0.07 µg/L	Liver vitellogenin
Tamoxifen	NIES/Japan	1.3 µg/L	1.3 µg/L	Liver vitellogenin
17*β*-Trenbolone	US EPA/MED	32 ng/L	32 ng/L	Papillae and Liver vitellogenin

Occasionally, there is a LOEC for an endpoint in the F_2 generation that is lower than the LOEC for the same endpoint in F_1. In all cases, the LOEC is different by only one treatment level. One of the requirements for making valid comparisons of differences in the treatment effects among generations is that the test conditions are the same in both generations. It is very difficult to actually achieve this requirement, especially with respect to the chemical exposure concentrations. Furthermore, in those cases where chemical concentrations are not the same across generations, it is essentially not possible to discern whether the difference between the LOECs are due to differences in test organism sensitivity or to differences in the chemical exposure at a critical window in fishes life history.

When considering using multiple generations in a test protocol it is appropriate to consider whether the additional data provided by each generation increases the quality of the risk assessment, and whether the test protocol is likely to be completed successfully. In general, our experience the longer a test continues, the greater the risk of incidents resulting in test failure. The causes of failure include exposure and/or dilution pump malfunctions, malfunctions in temperature control systems, disease, and, perhaps the most common problem -- maintaining consistent treatment concentrations throughout the test. As the test progresses, heterotrophic microbes which use the test chemical as an energy source, are selected for until they thrive and

proliferate in the exposure system. These microbes metabolize the test chemical with subsequent reductions in the treatment concentrations. To deal with this, the exposure system needs to be periodically sterilized and, adjustments to the dilution system must be made frequently to accommodate reductions in treatment concentrations. These adjustments are less arduous or fewer in the shorter test protocol.

5.1.3 Effect-endpoints Not Measured in F_0

Results from full MMT tests indicate that the LOEC for any particular endpoint in F_0 is never lower than the LOEC from either F_1 or F_2 for that endpoint. In a few MMTs the F_0 LOEC was the same as the LOEC in F_1 or F_2, (see Table 5-9). Consequently, we recommend that it is not necessary to measure endpoints in F_0.

Table 5-9. Comparative LOEC values for measurement endpoints from several research MMTs across generations. Green colored cells indicate the lowest LOEC among the three generations for that particular endpoint. Greater-than values indicate the LOEC for that particular generation was higher than the highest concentration that still had fish. Purple cells indicate that the LOEC is higher than the highest concentration tested.

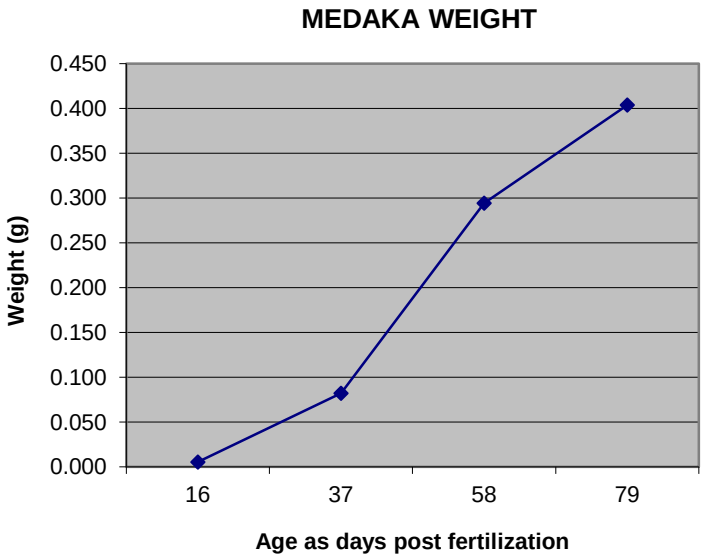
	Endpoint	F_0	F_1	F_2
E_2	SSC	>84 ng/L	28 ng/L	>28 ng/L
	Vitellogenin	28 ng/L	28 ng/L	28 ng/L
	Fecundity	>84 ng/L	28 ng/L	>28 ng/L
4tOP MED	Weight-XY	>102 µg/L	51 µg/L	>51 µg/L
	Weight-XX	102 µg/L	13 µg/L	>51 µg/L
	SSC	>102 µg/L	102 µg/L	13 µg/L
	Vitellogenin	51 µg/L	51 µg/L	51 µg/L
	Gonad Phenotype	102 µg/L	13 µg/L	51 µg/L
	Fecundity	>102 µg/L	102 µg/L	51 µg/L
DDT	SSC	>1.9 µg/L	>1.9 µg/L	>1.9 µg/L
	Vitellogenin	1.9 µg/L	0.69 µg/L	0.22 µg/L
	Gonad Phenotype	>1.9 µg/L	1.9 µg/L	0.69 µg/L
	Fecundity	>1.9 µg/L	>1.9 µg/L	1.9 µg/L
trenbolone	SSC	84 ng/L	84 ng/L	>13 ng/L
	Vitellogenin	>84 ng/L	32 ng/L	>13 ng/L
	Gonad Phenotype	>84 ng/L	84 ng/L	>13 ng/L
	Fecundity	84 ng/L	32 ng/L	>13 ng/L
prochloraz	Weight	41 µg/L	>25 µg/L	>25 µg/L
	Fecundity	41 µg/L	>25 µg/L	25 µg/L

5.1.4 Measure Subadult Endpoints at 9 wpf

At ORD/MED, the subadult sampling had been done, depending on tests, at times between 44 and 59 dpf. Within this time window, medaka are growing and developing rapidly enough that

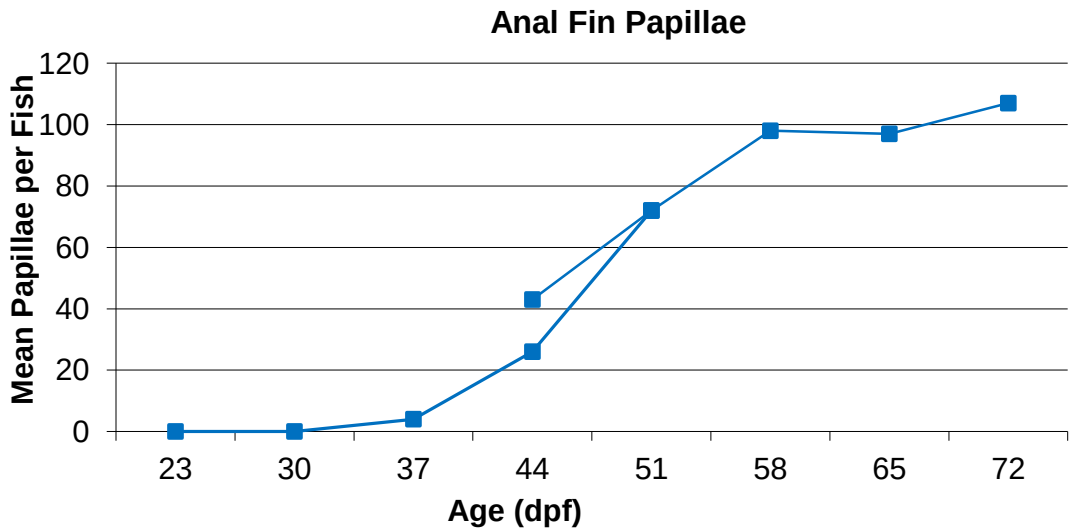
shifting the subadult sampling time a week earlier or later has dramatic impacts on the data collected. For instance, the medaka growth curve is very steep at this time (Figure 5-4).

Figure 5-4. Medaka Growth-Curve in MMT and MRT protocols



Concurrent with the rapid growth observed, the development of SSC (i.e. anal fin papillae) occurs around 40 dpf. Therefore by sampling subadults at 9 wpf, males will have high numbers of anal fin papillae (Figure 5-5). In addition, development of the gonad is dramatic during this phase of development as well, and there is a substantial increase in the production of vitellogenin. In typical MMTs, the mean number of copies of vitellogenin per ng of total liver RNA was 2×10^5 at 6 wpf and 5×10^6 at 7 wpf. Another case in point, the mean liver vitellogenin concentration in subadult control females of the F1 generation in the tamoxifen MMT performed by NIES/Japan was approximately 40 ng/mg liver while the mean of the subadult females of F2 in the same MMT was approximately 1350 ng/mg liver. The difference between the subadult samplings of F1 and F2 in the tamoxifen MMT was that in F1 the sampling was done at 54 dpf while in F2 the subadult sampling was done at 65 dpf, a difference of 11 days.

Figure 5-5. Time-course of Anal-fin Papillae Development



Based upon the information gathered, the recommendation is to have the subadult sampling take place during the 9th wpf so that the fish will have had a chance to development to produce sufficiently high mean values with low variance in all the measured parameters.

5.1.5 Test Temperature 25.5 ° C

Appropriate test temperature and the allowable variance both within an aquarium and between aquaria have been discussed at bilateral meetings between US EPA and Japan Ministry of the Environment. The recommended temperature and variance throughout the test is 25.5°C ± 0.5°C. Brief excursion from the mean by individual aquaria must not be more than 2°C. Replicates within a treatment must not statistically different from each other, and treatments within the test must not be statistically different from each other. This specification is based upon 1) the historical range of temperatures used for toxicology testing with medaka, 2) the temperature(s) for optimal reproduction, and 3) the need to minimize the occurrence of XX males.

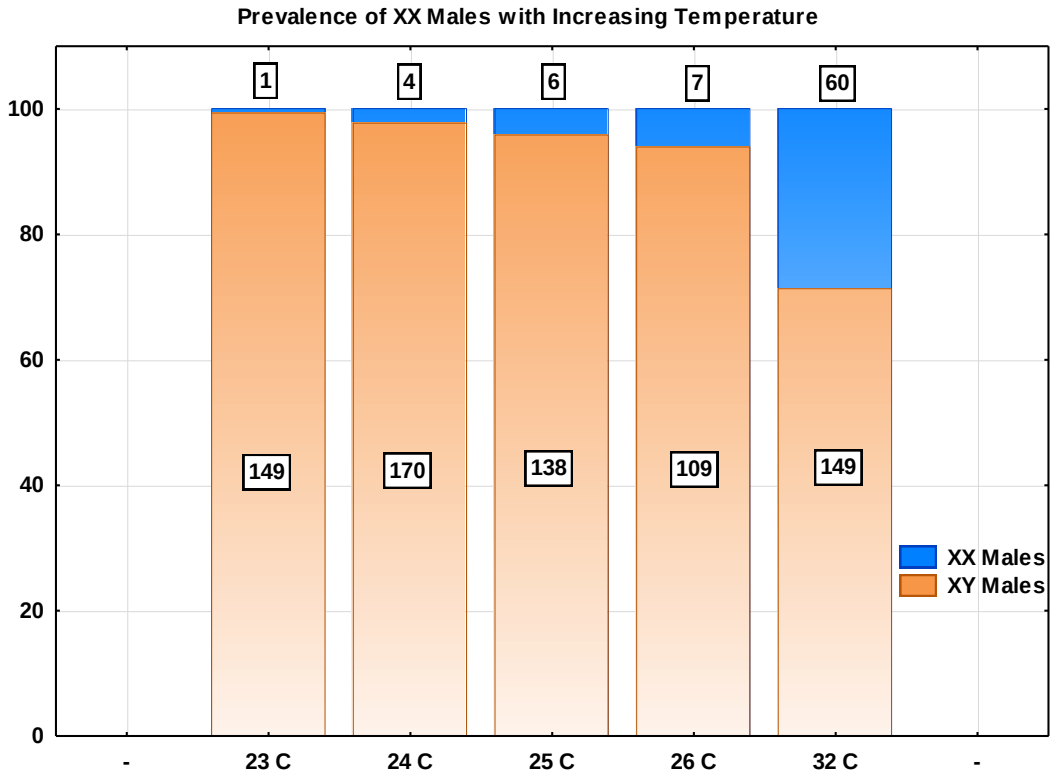
Table 5-9 shows the fecundity of control fish in a MMT run at 24°C and one run with the suggested specification. The higher control fecundity should result in greater statistical power for this vital endpoint.

Table 5-9. Control Fish Fecundity at 24 and 25.5 ° C.

	MMT at 24°C			MMT at 25.5°C		
	F0	F1	F2	F0	F1	F2
Fecundity (mean/pair-day)	26.3	13.9	26.4	37.8	32.5	39.1
Standard Deviation	1.4	6.2	1.6	2.9	2.6	2.8

It has been documented in the literature (Nanda et al., 2003; Sato et al., 2005; Selim et al., 2009) and observed in our medaka culture that a small number of unexposed XX medaka will spontaneously develop into normal phenotypic males. One factor that increases the likelihood of this is increased temperature. Temperatures reported in the published literature typically report effects at 32°C, a temperature not considered for the MRT. Thus, we initiated an ancillary study to evaluate appropriate temperature to recommend for the MRT protocol. Figure 5-6 shows how differences in temperature during development affect the prevalence of XX males in the population. Note that higher temperatures cause an increase in the prevalence of XX males even within a series of temperatures that seem to be appropriate for the MRT, *i.e.*, 24 – 26°C. The recommended temperature of 25.5° C for the MRT balances the competing objectives of maximizing fecundity while minimizing the prevalence of XX males in controls.

Figure 5-6. Prevalence of XX Males with Increasing Temperature. The bars show the proportion of phenotypic males that are either XX or XY out of 100%. The number boxes are the actual count of each genotype.



5.1.6 Pathology Analysis of Adult Fish Only

Another aspect of the 12 replicate test design is that the sample size for pathology analysis of adult fish in the MRT is double compared to the 6 replicates in the MMT. This difference is substantial. It provides pathology results for up to 12 fish per gender per treatment compared to the MMTs 6. Also, because the pathology is done on adult fish sampled after the reproduction evaluation, it provides information that can be correlated with the reproductive performance of each specific breeding pair. This is in contrast with the subadult pathology analysis which does

not allow these correlations to be made. Evaluation of treatment-related effects on gonadal development such as sex-reversal, and the prevalence testicular oocytes, is possible in either adult or subadult specimens, however, only the adult samples provide data that can be used to correlate pathology with reproduction. The increased sample sizes for adult pathology in the 12 replicate design increases both the statistical power and the relevance of the pathology endpoint compared with the 6 replicate test design. With increased replication the sample size for evaluating pathology is also increased which increases the statistical power of pathology endpoints, especially with the development of the RSCABS statistical procedure. Furthermore, the pathology analysis of the adult breeding pairs can be directly coordinated with their fecundity/fertility outcomes.

6 CONCLUSIONS

The outcomes of the various MMT trials have provided enough information to recommend a medaka reproduction test (MRT) for use as the fish test in Tier 2 of the EDSP. Two major changes from the MMT are proposed, *i.e.,* an increase in the number of replicates per treatment for evaluating effects on reproduction, and terminating the test after the embryos hatch in F_2. Other proposed changes include minimizing the collection of endpoint data from F_0, and evaluating pathology in only the F_1 adults sampled after the assessment of reproduction.

Overall, the authors conclude that both the MMT and the MRT are transferable methods and are capable of adequately characterizing potential disruption of the endocrine system by putative endocrine disrupting chemicals. However, the MRT is recommended as the preferred EDSP Tier 2 test method for fish because it is less resource intensive with improved statistical power, appears to be as sensitive, and is better able to ensure consistent findings when performed routinely by testing laboratories.

APPENDICES

Appendix A: Criteria for Validation of Assays in the EDSP (2007)

Appendix B: Detailed Review Paper

Appendix C: Vinclozolin Interlaboratory Summary Report

Appendix D: MMT Protocol

Appendix E: MRT Protocol

Appendix F: Pathology Guidance

Appendix G: Statistics

Appendix H: MMT Summary Reports

Appendix I: Power Analysis

REFERENCES

Adkins-Regan, E., Banerjee, S.B., Correa, S.M., Schweitzer, C., 2013. Maternal effects in quail and zebra finches: Behavior and hormones. Gen Comp Endocrinol.

Adkins-Regan, E., Ottinger, M.A., Park, J., 1995. Maternal transfer of estradiol to egg yolks alters sexual differentiation of avian offspring. Journal of Experimental Zoology 271, 466-470.

Almasi, B., Rettenbacher, S., Muller, C., Brill, S., Wagner, H., Jenni, L., 2012. Maternal corticosterone is transferred into the egg yolk. Gen Comp Endocrinol 178, 139-144.

Ankley, G., Jensen, K.M., Kahl, M.D., Durhan, E.J., Makynen, E.A., Cavallin, J.E., Martinovic, D., Wehmas, L.C., Mueller, N.D., Villeneuve, D.L., 2010a. Use of chemical mixtures to differentiate mechanisms of endocrine action in a small fish model. Aquatic Toxicology.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010b. Adverse outcome pathways: a conceptual framework to support exotoxicology research and risk assessment. Environmental Toxicology and Chemistry 29, 730-741.

Ankley, G.T., Jensen, K.M., Durhan, E.J., Makynen, E.A., Butterworth, B.C., Kahl, M.D., Villeneuve, D.L., Linnum, A., Gray, L.E., Cardon, M., Wilson, V.S., 2005. Effects of two fungicides with multiple modes of action on reproductive endocrine function in the fathead minnow (Pimephales promelas). Toxicol Sci 86, 300-308.

Ankley, G.T., Jensen, K.M., Makynen, E.A., Kahl, M.D., Korte, J.J., Hornung, M.W., Henry, T.R., Denny, J.S., Leino, R.L., Wilson, V.S., Cardon, M.C., Hartig, P.C., Gray, L.E., 2003. Effects of the androgenic growth promoter 17-beta-trenbolone on fecundity and reproductive endocrinology of the fathead minnow. Environ Toxicol Chem 22, 1350-1360.

Ankley, G.T., Kahl, M.D., Jensen, K.M., Hornung, M.W., Korte, J.J., Makynen, E.A., Leino, R.L., 2002. Evaluation of the aromatase inhibitor fadrozole in a short-term reproduction assay with the fathead minnow (Pimephales promelas). Toxicol Sci 67, 121-130.

Bayley, M., Larsen, P.F., Baekgaard, H., Baatrup, E., 2003. The effects of vinclozolin, an anti-androgenic fungicide, on male guppy secondary sex characters and reproductive success. Biol Reprod 69, 1951-1956.

Chikae, M., Ikeda, R., Hasan, Q., Morita, Y., Tamiya, E., 2004. Effects of tamazifen, 17-a-ethynylestradiol, flutamide, and methyltestosterone on plasma vitellogenin levels of male and

female Japanese medaka (Oryzias latipes). Environmental Tozicology and Pharmacology 17, 29-33.

Cripe, G.M., Hemmer, B.L., Goodman, L.R., Fournie, J.W., Raimondo, S., Vennari, J.C., Danner, R.L., Smith, K., Manfredonia, B.R., Kulaw, D.H., Hemmer, M.J., 2009. Multigenerational exposure of the estuarine sheepshead minnow (Cyprinodon variegatus) to 17beta-estradiol. I. Organism-level effects over three generations. Environmental Toxicology and Chemistry 28, 2397-2408.

Cripe, G.M., Hemmer, B.L., Raimondo, S., Goodman, L.R., Kulaw, D.H., 2010. EXPOSURE OF THREE GENERATIONS OF THE ESTUARINE SHEEPSHEAD MINNOW (CYPRINODON VARIEGATUS) TO THE ANDROGEN, 17 beta-TRENBOLONE: EFFECTS ON SURVIVAL, DEVELOPMENT, AND REPRODUCTION. Environmental Toxicology and Chemistry 29, 2079-2087.

Denny, J.S., Spehar, R.L., Mead, K.E., Yousuff, S.C., 1991. Guidelines for culturing the Japanese Medaka, Oryzias latipes, US EPA/600/3-91/064.

Denslow, N., Sepúlveda, M., 2007. Ecotoxicological effects of endocrine disrupting compounds on fish reproduction, The Fish Oocyte. Springer, pp. 255-322.

Edmunds, J.S., McCarthy, R.A., Ramsdell, J.S., 2000. Permanent and functional male-to-female sex reversal in d-rR strain medaka (Oryzias latipes) following egg microinjection of o,p'-DDT. Environ Health Perspect 108, 219-224.

EDSP, 2007. Validation of Screening and Testing Assays Proposed for the EDSP. Draft version 5.4. U.S. Environmental Protection Agency Endocrine Disruptor Screening Program, available at http://www.epa.gov/scipoly/oscpendo/pubs/edsp_validation_paper_v%205.4.pdf 21 pages.

EDSTAC, 1998. Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report, http://www.epa.gov/endo/pubs/edspoverview/finalrpt.htm.

Foran, C.M., Metcalfe, C.D., Metcalfe, T.L., Benson, W.H., 2000. Influence of exposure history on vitellogenin induction in medaka: a case of 'estrogen memory'? Marine Environmental Research 50, 196.

FRN (Federal Register Notice), 1998. Endocrine Disruptor Screening Program. Federal Register. August 11, 1998. 63(154):42852-42855.

Gray, M.A., Metcalfe, C.D., 1999. Toxicity of 4-tert-octylphenol to early life stages of Japanese medaka (Oryzias latipes). Aquatic Toxicology 46, 149-154.

Grim, K.C., Wolfe, M., Hawkins, W., Johnson, R., Wolf, J., 2007. Intersex in Japanese medaka (Oryzias latipes) used as negative controls in toxicologic bioassays: a review of 54 cases from 41 studies. Environ Toxicol Chem 26, 1636-1643.

Gronen, S., Denslow, N., Manning, S., Barnes, S., Barnes, D., Brouwer, M., 1999. Serum vitellogenin levels and reproductive impairment of male Japanese Medaka (Oryzias latipes) exposed to 4-tert-octylphenol. Environ Health Perspect 107, 385-390.

Hayward, L.S., Wingfield, J.C., 2004. Maternal corticosterone is transferred to avian yolk and may alter offspring growth and adult phenotype. Gen Comp Endocrinol 135, 365-371.

Henriksen, R., Groothuis, T.G., Rettenbacher, S., 2011. Elevated plasma corticosterone decreases yolk testosterone and progesterone in chickens: linking maternal stress and hormone-mediated maternal effects. PLoS One 6, e23824.

Hinfray, N., Porcher, J.M., Brion, F., 2006. Inhibition of rainbow trout (*Oncorhynchus mykiss*) P$S%) aromatase activities in brain and ovarian microsomes by various environmental substances. Comparative Biochemestry and Physiology Part C: Toxicology & Pharmacology 144, 252-262.

Hirai, N., Nanba, A., Koshio, M., Kondo, T., Morita, M., Tatarazako, N., 2006. Feminization of Japanese medaka (*Oryzias latipes*) exposed to 17β-estradiol: effect of exposure period on spawning performance in sex-transformed females. Aquat Toxicol 79, 288-295.

Hirshfield, M.F., 1980. An experimental analysis of reproductive effort and cost in the Japanese medaka Oryzias latipes, Ecology, 61:282-292.

Holdway, D.A., Hefferman, J., Smith, A., 2008. Multigeneration assessment of nonlyphenol and endosulfan using a model Australian freshwater fish, Melanotaenia fluviatilis. Environmental Toxicology 23, 253-262.

Kelce, W.R., Monosson, E., Gamcsik, M.P., Laws, S.C., Gray, L.E., 1994. Environmental hormone disuptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. Toxicology and Applied Pharamacology 126, 276-285.

Kiparissis, Y., Metcalfe, T.L., Balch, G.C., Metcalfe, C.D., 2003. Effects of the antiandrogens, vinclozolin and cyproterone acetate on gonadal development in the Japanese medaka (*Oryzias latipes*). Aquat Toxicol 63, 391-403.

Kirchen, R.V., West, W.R., 1976. The Japanese medaka: its care and development. Carolina Biological Supply Company.

Knorr, S., Braunbeck, T., 2002. Decline in reproductive success, sex reversal, and developmental alterations in Japanese medaka (Oryzias latipes) after continuous exposure to octylphenol. Ecotoxicol Environ Saf 51, 187-196.

Koger, C.S., Teh, S.J., Hinton, D.E., 1999. Variations of light and temperature regimes and resulting effects on reproductive parameters in Medaka (Oryzias latipes). Biology of Reproduction 61, 1287-1293.

Korner, W., Spengler, P., Bolz, U., Schuller, W., Hanf, V., Metzger, J.W., 2001. Substances with estrogenic activity in effluients of sewage treatment plants in southwestern Germany. 2. Biological analysis. . Environmental Toxicology and Chemistry 20, 2142-2151.

Laws, S.C., Carey, S.A., Ferrell, J.M., Bodman, G.J., Cooper, R.L., 2000. Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. Toxicol Sci 54, 154-167.

Lin, F., Wu, J., Abdelnabi, M.A., Ottinger, M.A., Giusti, M.M., 2004. Effects of dose and glycosylation on the transfer of genistein into the eggs of the Japanese quail (Coturnix japonica). Journal of agricultural and food chemistry 52, 2397-2403.

Makynen, E.A., Kahl, M.D., Jensen, K.M., Tietge, J.E., Wells, K.L., Van Der Kraak, G., Ankley, G.T., 2000. Effects of the mammalian antiandrogen vinclozolin on development and reproduction of the fathead minnow (Pimephales promelas). Aquat Toxicol 48, 461-475.

Martinovic, D., Blake, L.S., Durhan, E.J., Greene, K.J., Kahl, M.D., Jensen, K.M., Makynen, E.A., Villeneuve, D.L., Ankley, G.T., 2008. Reproductive toxicity of vinclozolin in the fathead minnow: confirming an anti-androgenic mode of action. Environ Toxicol Chem 27, 478-488.

Metcalfe, T.L., Metcalfe, C.D., Kiparissis, Y., Nimi, A.J., Foran, C.M., Benson, W.H., 2000. Gonadal development and endocrine responses in Japanese medeka (Oryzias Latipes) exposed to o,p'-DDT in water or through maternal transfer. Environmental Toxicology and Chemistry 19, 1893-1900.

Monod, G., De Mones, A., Fostier, A., 1993. Inhibition of ovarian microscomal aromatase and follicular oestradiol secretion by imidazole fungicides in rainbow trout. Marine Environmental Research 35, 153-157.

Nakari, T., Erkomaa, K., 2003. Effects of phytosterols on zebrafish reproduction in multigeneration test. Environmental Pollution 123, 267-273.

Nanda, I., Hornung, U., Kondo, M., Schmid, M., Schartl, M., 2003. Common spontaneous sex-reversed XX males of the medaka Oryzias latipes. Genetics 163, 245-251.

Newsome, C.S., 1980. A multigeneration fish toxicity test as an aid in the hazard evaluation of aquatic pollutants. Ecotoxicology and Enviromnental Safety 4, 362-369.

NIEHS, 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Report 97-3981. Research Triangle Park, NC.

OECD (Organization for Economic Cooperation and Development), 1996. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. Paris, France.

OECD (Organization for Economic Cooperation and Development), 2005. Guidance Document on the Validation of International Acceptance of New or Updated Test Methods for Hazard Assessment. Series on Testing and Assessment Number 34.

OECD, 2006. Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. OECD Series on Testing and Assessment Number 54.

Orn, S., Yamani, S., Norrgren, L., 2006. Comparison of vitellogenin induction, sex ratio, and gonad morphology between zebrafish and Japanese medaka after exposure to 17alpha-ethinylestradiol and 17beta-trenbolone. Arch Environ Contam Toxicol 51, 237-243.

Patyna, P.J., Cooper, K.R., 2000. Multigeneration reproductive effects of three phthalate esters in Japanese medaka (*Oryzias latipes*). Marine Environmental Research 50, 194.

Patyna, P.J., Davi, R.A., Parkerton, T.F., Brown, R.P., Cooper, K.R., 1999. A proposed multigeneration protocol for Japanese medaka (Oryzias latipes) to evaluate effects of endocrine disruptors. Sci Total Environ 233, 211-220.

Quinn, J.M.J., Lavoie, E.T., Ottinger, M.A., 2007. Reproductive toxicity of trenbolone acetate in embryonically exposed Japanese quail. Chemosphere 66, 1191-1196.

Quinn, M.J., Summitt, C.L., Ottinger, M.A., 2006. Effects of androgen disruption by DDE on the development and functioning of the immune system in Japanese quail. Immunopharmacology and immunotoxicology 28, 535-544.

Raimondo, S., Hemmer, B.L., Goodman, L.R., Cripe, G.M., 2009. Multigenerational exposure of the estruarine sheepshead minnow (*Cyprinodon variegatus)* to 17beta-estradiol. II. Population-level effects through two life cycles. Enviromental Toxicology and Chemistry 28, 2409-2415.

SAB-SAP, 1999. Science Advisory Board. EDSP. Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and the FIFRA Scientific Advisory Panel. U.S. Environmental Protection Agency. EPA-SAB-EC-99-013.

Sato, T., Endo, T., Yamahira, K., Hamaguchi, S., Sakaizumi, M., 2005. Induction of female-to-male sex reversal by high temperature treatment in Medaka, Oryzias latipes. Zoolog Sci 22, 985-988.

Schmieder, P., Tapper, M., Denny, J., Kolanczyk, R., Sheedy, B., Henry, T., Veith, G., 2004. Use of trout liver slices to enhance mechanistic interpretation of estogen receptor binding for cost-effective prioritization of chemicals within large inventories. Environmental Science and Technology 38, 6333-6342.

Schwaiger, J., Mallow, U., Ferling, H., Knoerr, S., Braunbeck, T., Kalbfus, W., Negele, R.D., 2002. How estrogenic is nonylphenol? A transgenerational study using rainbow trout (Oncorhynchus mykiss) as a test organism. Aquat Toxicol 59, 177-189.

Schweitzer, C., Goldstein, M.H., Place, N.J., Adkins-Regan, E., 2013. Long-lasting and sex-specific consequences of elevated egg yolk testosterone for social behavior in Japanese quail. Hormones and behavior 63, 80-87.

Segner, H., Navas, J.M., Schafers, C., Wenzel, A., 2003. Potencies of estrogenic compounds in in vitro screening assays and in life cycle tests with zebrafish in vivo. Ecotoxicol Environ Saf 54, 315-322.

Seki, M., Fujishima, S., Nozaka, T., Maeda, M., Kobayashi, K., 2006. Comparison of response to 17 beta-estradiol and 17 beta-trenbolone among three small fish species. Environ Toxicol Chem 25, 2742-2752.

Seki, M., Yokota, H., Maeda, M., Kobayashi, K., 2005. Fish full life-cycle testing for 17β-estradiol on medaka. Environ Toxicol Chem 24, 1259-1266.

Seki, M., Yokota, H., Maeda, M., Tadokoro, H., Kobayashi, K., 2003a. Effects of 4-nonylphenol and 4-tert-octylphenol on sex differentiation and vitellogenin induction in medaka (Oryzias latipes). Environ Toxicol Chem 22, 1507-1516.

Seki, M., Yokota, H., Matsubara, H., Maeda, M., Tadokoro, H., Kobayashi, K., 2003b. Fish full life-cycle testing for the weak estrogen 4-tert-pentylphenol on medaka (Oryzias latipes). Environ Toxicol Chem 22, 1487-1496.

Selim, K.M., Shinomiya, A., Otake, H., Hamaguchi, S., Sakaizumi, M., 2009. Effects of high temperature on sex differentiation and germ cell population in medaka, Oryzias latipes. Aquaculture 289, 340-349.

Shang, Y., Brown, M., 2002. Molecular determinants for the tissue specificity of SERMs. Science 295, 2465-2468.

Skolness, S.Y., Durhan, E., Garcia-Reyero, N., Jensen, K.M., Kahl, M.D., Makynen, E.A., Martinovic-Weigelt, D., Perkins, E., Villeneuve, D., Ankley, G.T., 2011. Effects of a short-term exposure to the fungicidee prochloraz on endocrine function and gene expression in female fathead minnows (*Pimephales promelas*). Aquatic Toxicology 103, 170-178.

Sone, K., Hinago, M., Itamoto, M., Katsu, Y., Watanabe, H., Urushitani, H., Tooi, O., Guillette, J.L.J., Iguchi, T., 2005. Effects of an androgenic growth promoter 17beta-trenbolone on masculinization of Mosquitofish (*Gambusia affinis affinis*). Gen Comp Endocrinol 143, 151-160.

Staples, C.A., Tilghman Hall, A., Friederich, U., Caspers, N., Klecka, G.M., 2011. Early life-stage and multigeneration toxicity study with bisphenol A and fathead minnows (Pimephales promelas). Ecotoxicology and Environmental Safety 74, 1548-1557.

Sun, L., Zha, J., Spear, P.A., Wang, Z., 2007a. Tamoxifen effects on the early life stages and reproduction of Japanese medaka (Oryzias latipes). Environmental Toxicology and Pharmacology 24, 23-29.

Sun, L., Zha, J., Spear, P.A., Wang, Z., 2007b. Toxicity of the aromatase inhibitor letrozole to Japanese medaka (Oryzias latipes) eggs, larvae and breeding adults. Comp Biochem Physiol C Toxicol Pharmacol 145, 533-541.

Uchida, M., Nakamura, H., Kagami, Y., Kusano, T., Arizono, K., 2010. Estrogenic effects of o,p '-DDT exposure in Japanese medaka (Oryzias latipes). Journal of Toxicological Sciences 35, 605-608.

USEPA, 2006. U.S. EPA Peer Review Handbook. EPA/100/B-06/002.

Wilson, V.S., Lambright, C., Ostby, J., Gray, L.E., Jr., 2002. *In Vitro* and *in Vivo* Effects of 17β-Trenbolone: A Feedlot Effluent Contaminant. Toxicol. Sci. 70, 202-211.

Wong, C., Kelce, W.R., Sar, M., Wilson, E.M., 1995. Androgen receptor antagonist versus agonist activities of the fungicide vinclozolin relative to hydroxyflutamide. J Biol Chem 270, 19998-20003.

Yamamoto, T., 1958. Artificial induction of functional sex-reversal in genotypic females of the medaka (Oryzias latipes). J Exp Zool 137, 227-263.

Yamamoto, T., 1975. Medaka (killifish): biology and strains. Yugaku-sha.

Yamamoto, T.O., 1953. Artificially induced sex‐reversal in genotypic males of the medaka (Oryzias latipes). Journal of Experimental Zoology 123, 571-594.

Yokota, H., Seki, M., Maeda, M., Oshima, Y., Tadokoro, H., Honjo, T., Kobayashi, K., 2001. Life-cycle toxicity of 4-nonylphenol to medaka (*Oryzias latipes*). Environ Toxicol Chem 20, 2552-2560.

Zha, J., Sun, L., Zhou, Y., Spear, P.A., Ma, M., Wang, Z., 2008. Assessment of 17alpha-ethinylestradiol effects and underlying mechanisms in a continuous, multigeneration exposure of the Chinese rar minnow (*Gobiocypirs rarus)*. Toxicology and Applied Pharamacology 226, 298-308.

Zhang, X., Hecker, M., Jones, P.D., Newsted, J., Au, D., Kong, R., Wu, R.S.S., Giesy, J.P., 2008. Responses of the medaka HPG axis PCR array and reproduction to prochloraz and ketoconazole. Environmental Science and Technology 42, 6762-6769.