**DRAFT FINAL REPORT**

**on**

# COMPARATIVE EVALUATION OF VITELLOGENIN METHODS FOR MEDAKA AND ZEBRA FISH

**EPA CONTRACT NUMBER 68-W-01-023**
**WORK ASSIGNMENT 2-26**

**July 2003**

**Prepared for**

**LES TOUART, PH.D.**
**WORK ASSIGNMENT MANAGER**
**U.S. ENVIRONMENTAL PROTECTION AGENCY**
**WASHINGTON, DC 20460**

**BATTELLE**
**505 King Avenue**
**Columbus, Ohio 43201**

## LIMITED DISTRIBUTION NOTICE

This document is transmitted in advance of patent clearance. This document is not to be published nor its contents otherwise disseminated or publicly used until approval for such release or use has been obtained from Battelle Intellectual Property Services, Pacific Northwest National Laboratory, Richland, Washington 99352.

## TABLE OF CONTENTS



## LIST OF TABLES

## LIST OF FIGURES

## ACKNOWLEDGMENTS

This document was prepared by the following:

Susan A. Thomas, Valerie Cullinan, Roy Kropp, Ann Skillman, and Michael L Blanton

Battelle's Pacific Northwest Division
1529 West Sequim Bay Road
Sequim, Washington
98382
USA.

This report would not be possible with out the participation of the following laboratories, which donated materials and time in the application of their methods to the samples in this study:

Biosense Laboratories, Norway
Centre for Fish and Wildlife Health, University of Bern, Switzerland
Department of Pathology, Vet. Medicine, Swedish University of Agricultural Sciences, Sweden
EnBioTec Laboratories, Ltd., Japan
Environmental and Symbiotic Sciences, Prefectural University of Kumamoto, Japan
Institute of Biology, University of Southern Denmark, Denmark
Los Angeles County Sanitation District, USA
National Institute for Environmental Studies (NIES), Japan
Notox Safety & Environmental Research, the Netherlands
Phylonix Pharmaceuticals, Inc., USA
Unité d'Evaluation des Risques Ecotoxicologiques (INERIS), France.

The authors gratefully acknowledge review comments from Les Touart (U.S. Environmental Protection Agency [EPA]), Gary Ankley (EPA Mid-Continent Ecology Division [EPA-MED]).

## 1.0 INTRODUCTION

The U.S. Environmental Protection Agency (EPA) is implementing an Endocrine Disruptor Screening Program (EDSP) composed of a battery of Tier 1 screening assays and Tier 2 tests. An international effort is also underway to develop and coordinate screens and tests appropriate for use in investigating potential endocrine disrupting chemicals. The Organization for Economic Cooperation and Development (OECD) has established an Endocrine Disrupter Testing and Assessment task force (EDTA) to oversee the coordination of this effort. One of the Tier 1 assays under development is a short-term screening assay designed to detect substances that interact with the estrogen and androgen systems of fish. It is thought that the inclusion of the fish screening assay in Tier 1 is important because estrogenic and androgenic controls on reproduction and development in fish may differ significantly from that of higher vertebrates, such that mammalian screening methods may not identify potential endocrine disruptor chemicals (EDCs) in this important class of animals. The measurement of a biochemical marker, vitellogenin (VTG), in oviparous vertebrates is generally agreed to be a good indicator of estrogenic and antiestrogenic effects, and it is proposed as one of several endpoints in the fish screening assay. VTG is a phospholipoglycoprotein precursor to egg yolk protein that normally occurs in sexually active female oviparous fishes, but can be induced to occur in males in response to estrogenic substances. Different methods are available to assess VTG induction in fishes, such as measurement of the VTG protein with enzyme-linked immunosorbant assays (ELISA) or matrix-assisted laser desorption/ionization-mass spectrometery (MALDI-MS), and messenger ribonucleic acid (mRNA) detection. Plasma, liver, and whole body measurements have been proposed. The Validation Management Group for ecotoxicity (VMGeco) of the EDTA recommended that a survey of existing VTG analytical methods be undertaken to assess their relative comparability.

The purpose of this study was to coordinate an interlaboratory comparison of existing ELISA VTG methods for analysis of zebra fish *(Danio rerio*) and medaka (*Oryzias latipes*) for suitability in a routine screening program. This comparison was not intended to be a validation of a given method, but an evaluation across methods to ascertain the qualitative and/or quantitative comparability of the variety of methods currently available.

The objectives of the study were the following:

A. Prepare a standard evaluation of two tissue homogenates (liver and whole body) and plasma taken from each of two species, zebra fish (*Danio rerio*) and medaka (*Oryzias latipes*), to provide a range of VTG concentrations produced in male and female fish that are either exposed or not exposed to an estrogen compound. The series for each species was to be produced with 1) uninduced male, 2) uninduced female, 3) induced male, and 4) induced female fish. In addition to the standard series, a set of positive control samples was to be prepared using uninduced male tissue spiked with purified VTG from the appropriate species. After shipment of samples to participating laboratories, an archive of the standard evaluation series and controls was to be created and maintained.

B. Identify laboratories to participate in the analysis of the standard evaluation series, coordinate transfer of the samples to the participating laboratories, and collect analytical results. Each laboratory was directed to employ the specific analytical technique it routinely uses to measure VTG, to report the results of the analysis, and to provide a detailed analytical protocol.
C. Statistically compare the data derived from the variety of analytical methods applied to the standard series by the participating laboratories, and prepare a final report that presents and discusses the data provided by the study participants, and the variability of the results.
D. Prepare a Quality Assurance Project Plan (QAPP) supporting this task to identify all applicable procedures and quality requirements.

## 2.0 SAMPLE PREPARATION AND HANDLING

The two VTG standard evaluation series were to be prepared from two homogenates (liver and whole body) and from blood plasma of each species of fish.[1] Fish were acquired, a subset of fish was exposed to estrogen, and both exposed and unexposed fish were used to prepare the standard series under an animal care protocol reviewed and approved by the Battelle's Pacific Northwest National Laboratory Animal Care Committee (accredited by the Association for Assessment and Accreditation of Laboratory Animal Care [formerly American Association for the Accreditation of Laboratory Animal Care]). In the preparation of all materials for the standard series, several steps were employed to aid in preserving the integrity of the samples, such as the use of a protein-inhibitor (aprotinin), cold processing, and quick-freezing to stabilize the VTG in the samples.

To generate samples of each tissue with zero to low levels of VTG representing *uninduced* background concentrations for the standard series, a set of adult male and female fish of each species was maintained without exposure to 17 β-estradiol. To generate samples of each tissue with high levels of VTG for the series representing *induced* concentrations, a set of adult male and female fish of each species was exposed to a nominal concentration of 300 ng/L 17 β-estradiol in the laboratory in a 7-day static renewal treatment to stimulate the production of VTG. After a 1-week exposure, when maximal VTG protein levels were anticipated, all of the fish from both the exposed and unexposed sets were sacrificed and processed as necessary to obtain the required tissues, which were then pooled by species, gender, and tissue type, and quick-frozen. At a later date, each composite was quickly thawed and used to prepare the standard series. In summary, the approach outlined in more detail below and in Section 3 resulted in four samples within the series: unexposed (uninduced) male, unexposed female, exposed (induced) male, and exposed female. In addition, a positive control was prepared from unexposed male tissue spiked with a known quantity of purified VTG, as the fifth sample to be included in each analysis.

- One portion of male and female fish from the exposed and unexposed groups was sacrificed for the preparation of whole body homogenate. The fish in each category (e.g.,

[1] However, blood plasma was subsequently deleted from this study.

female-exposed, or male-unexposed, etc.) were individually weighed, then pooled and homogenized. The composites were subsampled to prepare aliquots of each category for the standard series to be provided to each participating laboratory for analysis.

- From another portion of exposed and unexposed male and female fish, blood plasma and liver tissue samples were collected. Due to the small size of zebra fish and medaka, there was a limited amount of material available. Livers from individual fish were pooled and homogenized to create a large composite sample from which subsamples were taken to prepare individual aliquots for the standard series. The plasma from individual fish was pooled to create a composite plasma sample.

- A positive control for each tissue type for each species was prepared from unexposed male tissue spiked with a known quantity of purified VTG from the appropriate species, and was subsampled to prepare a set of aliquots. The purified VTG for each species was purchased from Biosense Laboratories.

- For each of the two fish species, multiple aliquots of each of the four samples in the whole body and liver standard series and of a positive control for each tissue were prepared as described above, and stored at -80°C until required for analysis.

- Shipment was coordinated to the participating laboratories, each of which assessed the level of VTG in the samples by various ELISA methods. Samples were shipped on dry ice in a package that included appropriate chain-of-custody documentation and instructions for storage, sample handling, analysis, and reporting. Instructions and forms were also sent electronically to each lab. The entire activity was carefully documented to ensure that sample integrity was not compromised. One of 11 shipments for ELISA analysis was allowed to thaw after the shipment had been received, and a replacement sample set was provided for analysis, following the standard shipping procedure.

## 3.0 SAMPLE METHODS

Approximately 400 adult zebra fish and 400 adult medaka were used in this study. Both species are small fish, and consequently, limited amounts of liver tissue and plasma can be collected from individuals. Therefore, to generate sufficient tissue for analysis by multiple laboratories, the plan was to collect samples from at least 20 individuals of each gender for each category of tissue type, exposure or nonexposure to estrogen, and species. Because the determination of the sex of zebra fish based on morphology alone is not reliable, sex ratios were confirmed by examination of the gonads after sacrifice of the fish. For the treatment of live fish, the number of male and female zebra fish was estimated based upon an anticipated normal 50:50 sex ratio of the fish (Braunbeck et al. 2003). The medaka were sexed prior to exposure based on external morphology. Table 1 summarizes the schedule of exposure to estrogen for both species; the staggered initiation dates facilitated fish handling. Fish were exposed to 300 ng/L 17 β estradiol in a 7-day static renewal treatment.

**Table 1. Estrogen exposure schedule for zebra fish and medaka**

| Species | Initiation Date | Termination Date (Day 7) | Number of Fish | Sex | Type |
|---|---|---|---|---|---|
| **Zebra fish** | 31 Mar 03 | 7 Apr 03 | 40<br>20 | unknown<br>unknown | whole body<br>WQ[(a)] |
| | 1 Apr 03 | 8 Apr 03 | 20 | unknown | whole body |
| | 2 Apr 03 | 9 Apr 03 | 60 | unknown | plasma/liver |
| | 3 Apr 03 | 10 Apr 03 | 60 | unknown | plasma/liver |
| **Medaka** | 27 Mar 03 | 3 Apr 03 | 20<br>20<br>20<br>20 | male<br>female<br>male<br>male | whole body<br>whole body<br>whole body<br>WQ |
| | 9 Apr 03 | 16 Apr 03 | 60 | female | plasma/liver |
| | 10 Apr 03 | 17 Apr 03 | 60 | male | plasma/liver |
| **Total** | | | **400** | | |

(a) WQ Water quality; these fish were used as control to monitor water quality, not for generation of tissue samples. They were therefore not sacrificed on Day 7.

On Day 7 of exposure, approximately 40 exposed fish of each species were anesthetized with tricaine methane sulfonate (MS-222), examined to determine sex in the case of zebra fish, quick-frozen, and stored at -80°C to be used to prepare whole body induced homogenate (Brion et al. 2002). A set of 60 unexposed fish of each species was similarly collected, sexed, and sacrificed for the whole body uninduced homogenate preparation. Whole body tissue homogenate was prepared as follows: fish were placed in ice-cold ELISA assay buffer in a 1:2 ratio by weight and homogenized on ice with a hand-held, ground-glass homogenizer. After the resulting material was centrifuged, the supernatant was harvested, subsampled to 20-µL aliquots, quick-frozen on liquid nitrogen, and stored at -80°C (Kang et al. 2002).

Similarly, one set of 60 exposed fish and one set of 60 unexposed fish of each species were processed on Day 7 for the liver and plasma standard series. From the sacrificed fish, blood was collected from the caudal vein via microcapillary tubes (Kordes et al. 2002; Van den Belt et al. 2002). Following centrifugation of the plasma, the supernatant was harvested and frozen at -80°C. Following the collection of blood, livers were removed from the same fish. The liver homogenate was prepared by placing the weighed livers in ice-cold ELISA assay buffer in a 1:2 ratio by weight and homogenizing on ice with a hand-held, ground-glass homogenizer. The resulting homogenate was centrifuged, the supernatant was harvested, subsampled to 20-µL aliquots, quick-frozen on liquid nitrogen, and stored at -80°C (Kang et al. 2002).

Positive controls were created from whole body and liver homogenates of unexposed male fish of each species spiked with a known quantity of purified VTG from the corresponding species. All work was conducted on ice to maintain the cold temperature of materials. The spiking rate for each control depended on the actual weight of lyophilized, purified VTG available in each commercially prepared vial purchased from Biosense Laboratories; the specific concentrations were 6.25 µg VTG/mL homogenate for the zebra fish whole body, medaka whole body,

**Table 2. VTG spike concentration in positive controls**

| Tissue | VTG Concentration (µg/mL homogenate) |
|---|---|
| Zebra fish whole body homogenate | 6.25 |
| Zebra fish liver homogenate | 6.25 |
| Medaka whole body | 6.25 |
| Medaka liver | 6.85 |

and medaka liver positive controls, and 6.85 µg VTG/mL homogenate for the zebra fish liver positive control. (Table 2). The resulting positive controls were subsampled to 20-µL aliquots, quick-frozen on liquid nitrogen, and stored at -80°C (Kang et al. 2002). Care was taken in each of the sample preparation steps to collect and process the samples in a timely manner under cold conditions followed by quick-freezing to limit the time from collection to storage and to avoid repeated freezing and thawing.

The total number of samples prepared as aliquots for analysis was determined by the number of participating laboratories. The volume of the aliquots was based on the amount of material available and the analytical requirements; 20 µL was determined to be sufficient and appropriate for all homogenate samples. Cryovials had been selected as the appropriate containers for the aliquotted samples. In addition to the samples for analysis by the participating laboratories, the excess material of every category of the standard series, including blood plasma and positive controls, was placed in long-term storage at -80°C, assigned a unique code, and entered into an archive management system.

Eleven laboratories participated in the VTG ELISA survey for the analysis of the tissues of one or both species. Based on the requirements of each laboratory, a complete set of either zebra fish, medaka, or both was assembled, packaged on dry ice in well-insulated containers, and shipped via Federal Express to the testing facilities of the participating laboratories. Also included in the shipping boxes were chain of custody documents, sample receipt questionnaire, and complete instructions/information. To limit the need to thaw and refreeze samples, three aliquots of each sample in the series (e.g., three vials of induced male whole body homogenate) were provided. Because the samples were analyzed fully blind, each vial was labeled with a unique code that did not reveal to the analyst the identity of the samples in the series, nor their corresponding low to high expected concentrations of VTG. Only the fish species and the type of homogenate, either whole body or liver, were defined. The laboratories were instructed to analyze the contents of each sample cryovial in triplicate, and each cryovial contained sufficient volume of material to apply to three wells on an ELISA plate. Figure 1 illustrates diagrammatically the organization of one set of zebra fish samples as they were provided to each laboratory for analysis. One large zip-lock bag labeled by species held three smaller bags containing the whole body standard series and positive control (15 cryovials organized as 5 sets, each consisting of 3 replicate samples), liver homogenate standard series and control (15 cryovials organized as 5 sets, each consisting of 3 replicate samples), and purified VTG (1 vial of commercially prepared, lyophilized VTG of the appropriate species).

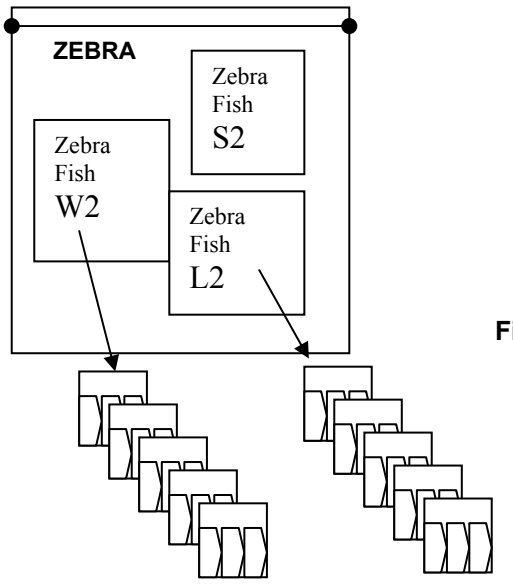



**Figure 1. Diagram of the organization of one set of zebra fish samples for analysis** (codes: S indicates purified VTG, W indicates whole body, L indicates liver, and 2 is the laboratory identification number; each vial was also individually labeled with a unique code)

## 4.0 LABORATORY ANALYTICAL METHODS

Several methods have been developed for the quantification of VTG in blood plasma, liver tissue, or whole body homogenates; they differ in sensitivity, specificity, and technical difficulty. A comprehensive survey of the literature and information from experts in the field of induction of VTG in fish (Battelle 2002) revealed that the technique of ELISA is currently the most widely developed and applied technique, with multiple methods that have been developed for specific application to zebra fish and medaka. The various ELISA methods employ enzyme-linked antibodies and an adsorbent surface to detect specific antigens in solution, typically in one of three general assay formats: competitive, sandwich, and direct assessments. Competitive ELISAs incorporate a step in which the samples and antibody (antibody-capture) or labeled antigen (antigen-capture) are incubated together prior to adding the sample on the test plate. This nonequilibrium design is often used to enhance sensitivity, and it counteracts potential preferential binding (Edmunds et al. 2000). Sandwich ELISAs employ two antibody preparations to detect the antigen. The antigens can recognize different epitopes on the target analyte, thereby providing a large degree of specificity and sensitivity. In a direct antibody-capture ELISA, the sample and standards are adsorbed directly on the surface of the microwell plate. After incubation, the wells are blocked and anti-VTG antibody is added to bind to the VTG attached to the well. As in other ELISAs, subsequent steps culminate in the development of color indicator that is reflective of the amount of antigen present in the sample.

Specific protocols employed by the participating laboratories (Appendix A) were applied to the identical sets of zebra fish and/or medaka homogenate samples supplied for the present study. Although the majority of laboratories used the same method (a commercially available kit), there were four distinct ELISAs applied to zebra fish homogenates, and three to medaka. Further, in the medaka survey, one laboratory applied all three medaka methods using three sets of samples, thereby providing a special opportunity to make a comparison in which variability due to laboratory precision or accuracy could be reduced.

## 5.0 PARTICIPATING LABORATORIES

The final list of participants consisted of 11 laboratories selected because of their previous experience in the measurement of VTG protein, and their willingness and ability to commit to completing the analysis on a volunteer basis. Each laboratory had established protocols in routine use. The 10 laboratories that analyzed VTG in the whole body and liver homogenate standard series and contributed their results to the survey are listed in Table 3, with the methods used. One laboratory was unable to complete the analysis within the required timeline for inclusion in the statistical comparison; therefore, this laboratory's method is described briefly in a footnote to Table 3 and its original data sheets are presented in Appendix B along with the results from the other participants.

## 6.0 DATA ANALYSIS

Data analysis was intended to provide descriptive statistics and plots that allow a general assessment of the objectives of the study. Statistically, the first objective was to determine whether an increasing concentration of VTG was produced by the standard series. This series was represented in order of concentration from low to high by 1) uninduced male, 2) uninduced female, 3) induced male, and 4) induced female fish. The second statistical objective was to determine the analytical results and variation for the set of control and spiked VTG samples for each species. The third statistical objective was to compare the analytical results and variation of each laboratory's analytical method, including the standard and assay used.

Analysis of the data yielded descriptive statistics, including the number of samples, means, standard deviations (SD), medians, first and third quartiles, and the coefficient of variation (CV). Simple linear regression of the ranked average VTG concentration (mean of the within-run analyses) and plots of the analytical results against the concentration series were used to assess the strength of the VTG concentration trend, ignoring the positive control. Tukey's Honestly Significant Difference (HSD) multiple comparison test was conducted on the ranked average VTG concentrations to specifically determine whether neighboring means in the series were significantly different (i.e., the blank mean compared with the uninduced male mean, the uninduced male mean compared with the uninduced female mean, and so on). Linear regression for each laboratory was also conducted on the average VTG concentrations observed for the blank and the uninduced male data. The regression results allow a test of the null hypothesis that the slope equals 0 and provides a measure of the strength of the trend. The multiple comparison testing, which is less powerful than the regression analysis due to the smaller degrees of freedom for testing, provides a test of how quickly differences can be detected in the series. Excel spreadsheet software (Microsoft Excel) and Minitab statistical software (Minitab Inc.) were used for this analysis.

## Table 3. Summary of reporting laboratories and their VTG ELISA methods[(a)]

| Lab ID | Participating Laboratory | ELISA Method | Reference | zebra fish liver MDL | zebra fish whole body MDL | medaka liver MDL | medaka whole body MDL |
|---|---|---|---|---|---|---|---|
| 1 | University of Kumamoto Kumamoto, Japan | Sandwich ELISA[(b)] | Biosense Laboratories (2002, 2003) | 0.49 ng/mL | 0.49 ng/mL | 0.24 ng/mL | 0.24 ng/mL |
| 2 | Biosense Laboratories Bergen, Norway | Sandwich ELISA | Biosense Laboratories (2002, 2003) | determined for two standards using whole body homog (e.g., 151 and 76 ng/mL at 1:300 dilution) | determined for the two STDs using whole body homog; (e.g., 151 and 76 ng/mL at 1:300 dilution) | determined for two standards using liver homog (e.g., 75 and 1.5 ng/mL at 1:300 dilution | detemined for two standards using liver homog (e.g., 75 and 1.5 ng/mL at 1:300 dilution) |
| 4 | EnBioTec Laboratories, Ltd. Tokyo, Japan | Sandwich ELISA, monoclonal antibody[(c)] | EnBio (2002); Nishi et al. (2002) | 8.2 ng/mL (n=8) determined by ELISA (0.82 using EnBio VTG standard) | 16.4 ng/mL (n=8) determined by ELISA (0.82 using EnBio VTG standard) | 9.8 ng/mL (n=8) determined by ELISA (0.49 using EnBio VTG standard) | 9.8 ng/mL (n=8) determined by ELISA (0.49 using EnBio VTG standard) |
| 5 | Notox Safety & Environmental Research Hertogenbosch, The Netherlands | Sandwich ELISA | Biosense Laboratories (2002, 2003) | 0.12 ng/mL determined based on Lab 5's standard | 0.12 to 0.24 ng/mL determined based on Lab 5's standard | 0,4357 ng/mL, from previous experiments | 0,4357 ng/mL, from previous experiments |
| 7 | Prefectural Universität Bern Bern, Switzerland | Sandwich ELISA | Biosense Laboratories (2002) | 0.5 ng/mL, routine using Biosense kit | 0.5 ng/mL, routine using Biosense kit | | |
| 8 | Phylonix Pharmaceuticals, Inc. Cambridge, Massachusetts, USA | Sandwich ELISA | Biosense Laboratories (2002) | 1.999991544 ng/mL, determined by phylonix (n=8) | 1.999991544 ng/mL, determined by phylonix (n=8) | | |
| 9 | Institute of Biology, University of Southern Denmark Odense, Denmark | Direct noncompetitive sandwich ELISA[(d)] | Holbech et al. (2001) | determined for two standards: 16 ng/mL (n=12) and 35.6 ng/mL (n=12) | determined for two standards: 16 ng/mL (n=12) and 35.6 ng/mL (n=12) | | |
| 10 | Department of Pathology, Faculty Veterinary Medicine Swedish University of Agricultural Sciences (SLU) Uppsala, Sweden | Modified direct noncompetitive sandwich ELISA[(d)] | Borg (unpublished)[(e)] | 40 ng/mL determined for Lab 10's standard | 40 ng/mL determined for Lab 10's standard | | |

**Table 3. Contd.**

| Lab ID | Participating Laboratory | ELISA Method | Reference | zebra fish liver MDL | zebra fish whole body MDL | medaka liver MDL | medaka whole body MDL |
|---|---|---|---|---|---|---|---|
| **11** | National Institute for Environmental Studies (NIES) Ibaraki, Japan | Sandwich ELISA, monoclonal antibody; Direct sandwich ELISA, monoclonal and polyclonal antibodies;[f] Sandwich ELISA | EnBio (2002); Nishi et al. (2002)<br><br>Transgenic (2002)<br><br>Biosense Laboratories (2002) | | | 2.0 ng/mL published in protocol<br><br>0.488 ng/mL minimum of working range, published Transgenic | 2.0 ng/mL published in protocol<br><br>0.488 ng/mL minimum of working range, published Transgenic |
| **12** | Los Angeles County Sanitation Districts Whittier, California, | Sandwich ELISA | Biosense Laboratories (2002, 2003) | 0.12 ng/mL<br><br>matrix-specific determination | 0.12 ng/mL<br><br>matrix-specific determination | 0.49 ng/mL<br><br>matrix-specific determination | 0.49 ng/mL<br><br>matrix-specific determination |

a) Laboratory 6, Unité d'Evaluation des Risques Ecotoxicologiques (INERIS), Verneuil-en-Halatte, France, used a competitive binding assay (Brion et al. 2002). Because the results from Laboratory 6 were received too late for inclusion in the comparative statistical analysis, the original data sheets are presented in Appendix B.
b) Sandwich enzyme immunoassay, anti-zebra fish or anti-medaka VTG capture antibody and detecting antibody.
c) Sandwich ELISA, anti-zebra fish VTG monoclonal antibody produced in mouse; mono-mono.
d) Direct, noncompetitive sandwich ELISA, anti-zebra fish lipovitellin, polyclonal antibody produced in rabbit.
e) D. Borg, 2003 (unpublished), *ELISA protocol for the detection of vitellogenin in zebrafish*, Department of Biology, Odense University, Denmark; received as personal communication, 4 July 2003.
f) Direct sandwich ELISA, anti-medaka VTG monoclonal antibody and biotinylated polyclonal antibody; poly-poly.

## 7.0 SUMMARY OF RESULTS

Sets of zebra fish and/or medaka homogenates of liver and whole body were supplied to 11 participating laboratories for VTG analysis; not every laboratory analyzed both species (Table 4). It was intended that one laboratory would also analyze blood plasma from each species; however, due to extenuating circumstances, the particular laboratory was unable to join the study as planned, and plasma was therefore not analyzed. All laboratories contributed their services without compensation, and Battelle made every attempt to accommodate and to assist the laboratories in performing their task.

The majority of the participating laboratories (6 of 11) employed exclusively a commercial enzyme immunoassay (EIA) kit for zebra fish or medaka, which is a sandwich ELISA using specific binding between antibodies and VTG (Biosense 2002, 2003) (see Table 3). One additional lab used the same kit along with two other methods for comparison: a sandwich ELISA using monoclonal antibodies (a commercial kit, Nishi et al. 2002) and a direct sandwich ELISA using monoclonal antibodies and biotinylated polyclonal antibodies (a commercial kit, Transgenic 2002). Still others used a direct noncompetitive sandwich ELISA (Holbech et al. 2001), or a modification thereof (personal communication, Daniel Borg, 2003, Department of Biology, Odense University, Denmark, unpublished *ELISA protocol for the detection of vitellogenin in zebrafish)*. The eleventh laboratory, not included in the statistical comparison, used a competitive binding assay (Brion et al. 2002).

**Table 4. Species analyzed by participating laboratories[a]**

| Lab ID | Participating Laboratory | Zebra Fish | Medaka |
|---|---|---|---|
| 1 | University of Kumamoto | X | X |
| 2 | Biosense Laboratories | X | X |
| 4 | EnBioTec Laboratories, Ltd. | X | X |
| 5 | Notox Safety & Environmental Research | X | X |
| 7 | Prefectural Universität Bern | X | |
| 8 | Phylonix Pharmaceuticals, Inc. | X | |
| 9 | Institute of Biology, University of Southern Denmark | X | |
| 10 | Swedish University of Agricultural Sciences (SLU) | X | |
| 11 | National Institute for Environmental Studies (NIES) | | XXX[b] |
| 12 | Los Angeles County Sanitation Districts (LACSD) | X | X |

a) Laboratory 6, Unité d'Evaluation des Risques Ecotoxicologiques (INERIS), analyzed both zebra fish and medaka tissues; however, it reported its results too late for inclusion in the comparative statistical analysis. The original data sheets presented in Appendix B.
b) NIES requested and analyzed three sets of medaka samples, using three different methods.

Protocols appear in detail in Appendix A and are listed in Table 3 with a brief description along with the method detection limit (MDL). Original data sheets from all the participating laboratories are found in Appendix B. A full and detailed presentation of results, including text, data tables, and figures, comprises Appendix C, and the descriptive statistics are available in Appendix D, E, and F for the within-run VTG results, intralaboratory results, and intra-assay results, respectively.

## Zebra Fish

Nine of the participating laboratories analyzed zebra fish tissue homogenates. A summary of the codes indicating concentration, standards, and assay method for zebra fish statistical analysis is presented in Table 5. The *concentration codes* 1 through 5 are based on the exposure history and sex of the fish used to generate the samples in the series, plus the positive control. The *standard codes* identify the two different calibration standards. However, in the text that follows, the term "homologous" refers to the standard routinely employed by the individual laboratory with respect to the assay in use, and is prepared or purchased by the laboratory; the term "purified" refers to the standard that was supplied by Battelle to the participating laboratories—that is, the purified VTG standard commercially prepared by Biosense Laboratories. The four *method* or *assay code* numbers define the particular commercial kits or unique method employed. These groupings were used to analyze the variability of the reported results.

Not all laboratories used both standards: six of the nine laboratories that analyzed zebra fish homogenates employed the two standards, as requested; two used only the purified zebra fish VTG supplied by Battelle,[2] and one used only its own homologous standard (Table 6).

[2] The two laboratories commented that the purified standard was identical to that they typically used with the method employed for this study (see original data sheets, Appendix B).

**Table 5. Summary of codes for relative VTG concentration, standards, and ELISA methods for zebra fish**

| Description | Code |
|---|---|
| **CONCENTRATION** | |
| Uninduced Male | 1 |
| Uninduced Female | 2 |
| Induced Male | 3 |
| Induced Female | 4 |
| Positive Control | 5 |
| **STANDARD** | |
| Homologous | 1 |
| Purified | 2 |
| **ZF-METHOD AND REFERENCE[(a)]** | |
| 1 (Biosense 2003) | 1 |
| 2 (EnBio 2002; Nishi et al. 2002) | 2 |
| 3 (Holbech et al. 2001) | 3 |
| 4 (Borg 2003, unpublished) | 4 |

a) ZF Zebra Fish; see description of assay methods in Table 3.

**Table 6. Summary of standards employed by participating laboratories analyzing zebra fish**

| Lab ID | Participating Laboratory | Homologous standard | Purified zebra fish standard[(a)] |
|---|---|---|---|
| **1** | University of Kumamoto | X | X |
| **2** | Biosense Laboratories | X | X |
| **4** | EnBioTec Laboratories, Ltd. | X | X |
| **5** | Notox Safety & Environmental Research | | X[(b)] |
| **7** | Prefectural Universität Bern | | X[(b)] |
| **8** | Phylonix Pharmaceuticals, Inc. | X | X |
| **9** | University of Southern Denmark | X | X |
| **10** | Swedish University of Agricultural Sciences | X | |
| **12** | County Sanitation Districts of Los Angeles County | X | X |

a) Commercially prepared, purified VTG supplied by Battelle to participants.
b) The standard supplied by Battelle was the same as that used by the laboratory; accordingly, the laboratory elected to employ only one standard, providing one set of data for its sample set.

### *Trends of Standard Series: Measured VTG Concentrations*

A goal of the study was to generate a range of concentrations of VTG in male and female liver and whole body homogenates, beginning with zero to low levels in uninduced males, increasing in the following order: uninduced male<uninduced female<induced male<induced female zebra fish VTG concentrations. The VTG concentration in the precisely spiked positive control, which was 6.25 µg (=6250 ng) VTG/mL liver and whole body homogenate for zebra fish, could have been expected to fall between the zero-to-low uninduced male and the uninduced female levels. In consideration of all of the reported data, the general trend for the liver samples

observed for each laboratory, averaged over antibodies, standards, and assays, differed from the expected increasing concentration series of VTG specifically in the similarity of VTG levels in uninduced female and induced males (Figure C4 in Appendix C). The whole body homogenate samples were more consistent with expected results with respect to the order of induced females with VTG levels greater than those for induced males in over 71% (5 of 7 labs) of the reported data sets; however, the VTG concentrations in whole body tissue of induced and uninduced females were reversed with respect to the expected gradient in over half (4 of 7 labs) of the laboratories' data sets reporting for this analysis (Figure C5 in Appendix C.) The divergence from the expected series points out the need for further examination of the potential sources of variation. This analysis included all reported results, and additional analysis was conducted based on laboratory and type of method to examine this variability.

### *Difference in Standards*

A critical aspect of the performance of an analytical method is the use of standards and controls in an assay. Two standards were used for the present study: *purified* standard was purchased and supplied to the participating laboratories by Battelle; the *homologous* standard was the standard routinely used by each laboratory with their preferred method. The percentage difference between the average replicate VTG concentrations obtained with the homologous (H) and the purified (B) standard was calculated as (H-B)/B X 100%. Negative values represented greater VTG concentrations obtained with the purified standard. Six of the nine labs analyzing zebra fish homogenates used both standards; two used only the purified standard because it was identical to their homologous standard; and one used only its homologous standard. In general, the differences were relatively small for all concentrations for zebra fish (ZF) Methods 1 and 3, which represent a sandwich ELISA and a direct noncompetitive sandwich ELISA, respectively, but quite large in the results of ZF-Method 2, a monoclonal-antibody-based sandwich ELISA, for all concentrations (see Tables C9-C12 and Figures C6-C10 in Appendix C; Appendix F). VTG concentrations were about 700% to 800% lower for the purified than for the homologous standard.

There was a wide range of values measured in both liver and whole body samples analyzed in the various laboratories with calibration to the two different standards; the variability depended primarily on the analytical method and secondarily on the number of laboratories running the analyses. VTG values tended to be highest for the ZF-Method 1 and lowest for ZF-Method 2 for liver. For example, for the homologous standard, VTG concentrations measured by ZF-Method 1 for liver homogenate were 24 times to 3277 times greater than those measured by ZF-Method 2. Similarly, for the purified standard, the concentrations were 98 to 2099 times greater measured by ZF-Method 1 than by 2. The great variability between methods is illustrated by the detection of VTG in uninduced males: not detected by ZF-Method 2, but measured at 136,117 ng/mL by ZF-Method 1.

For whole body homogenate, homologous standard, the measured VTG values were highest for ZF-Method 4, a modified direct noncompetitive sandwich ELISA, from 13 to 19,951 times higher than those measured by ZF-Method 2. For the purified standard, the values were generally

highest for ZF-Methods 1 and 3, which were 16 to 17,589 times greater than those measured by ZF-Method 2.

Multiple laboratories applied the same, commercially available method (ZF-Method 1), and a high interlaboratory variability (CVs from 63% to 229% for liver; 26% to 82% for whole body) could be seen in the values attained with the homologous standard, with a surprisingly high variability (the highest) for the positive controls for the liver homogenate, and in the uninduced male samples for whole body. For the purified standard, the highest variability was likewise for the positive controls in liver, but for induced male samples in whole body homogenate samples.

### *Within-Run Variability*

The three analytical replicates provide a measure of the within-run variability (Table C7 and Figures C2, C3 in Appendix C; Appendix D). When all of the methods were applied to the full range of samples, both the zebra fish liver and whole body showed a wide range of variability of absolute values (CVs 0% to 72.9%; mean 9.4%, and 0% to 76.1%, mean 8.6%, respectively) among analytical replicates, but a consistency of trend. However, because 75% of the CVs were <10% for both homogenates, and both tended to be <30% for uninduced females and induced males, this comparison indicates relatively low variability in the expected mid- and higher VTG concentrations, but a somewhat higher variability in bottom ranges. The latter can be graphically seen to be explained by low outlier values for concentration Codes 1 and 2 for Labs 4 and 5 for the liver (Figure 2a), and for Codes 1 and 3 in particular for Lab 4 in whole body (Figure 2b). Otherwise, the results of the balance of the laboratories' results appear to hold relatively tight grouping, especially at the higher values. This plot also emphasizes the general trend of measured VTG values, which follows the expected concentration curve of the standard series. The detection of low-level VTG is a critical component of a method for use in a screening assay to detect the induction of the VTG protein; the zebra fish results appear to exhibit sufficient consistency within that range for most of the laboratories.

### *Intra-Assay Variability*

The analysis of sample triplicates provided a measure of intra-assay variability (Table C8 in Appendix C; Appendix E). The range of CVs was broad for both liver and whole body homogenates (0.3% to 173.2%), but 75% of the intra-assay CVs were <29% for both types of tissue. This level of intra-assay variability indicates that when a sample is provided to multiple laboratories employing a variety of methods (the results using multiple standards are also included in this sample set) a quarter of the methods provide a moderately high degree of variability in absolute VTG measured value when replicate samples are analyzed, but three quarters showed low variability. The general trends (in contrast with the focus on absolute VTG quantities) tracked the expected concentration series fairly well. To further examine this type of variability, which is critical to the application of ELISA to a screening assay, the data were further examined by individual laboratory.

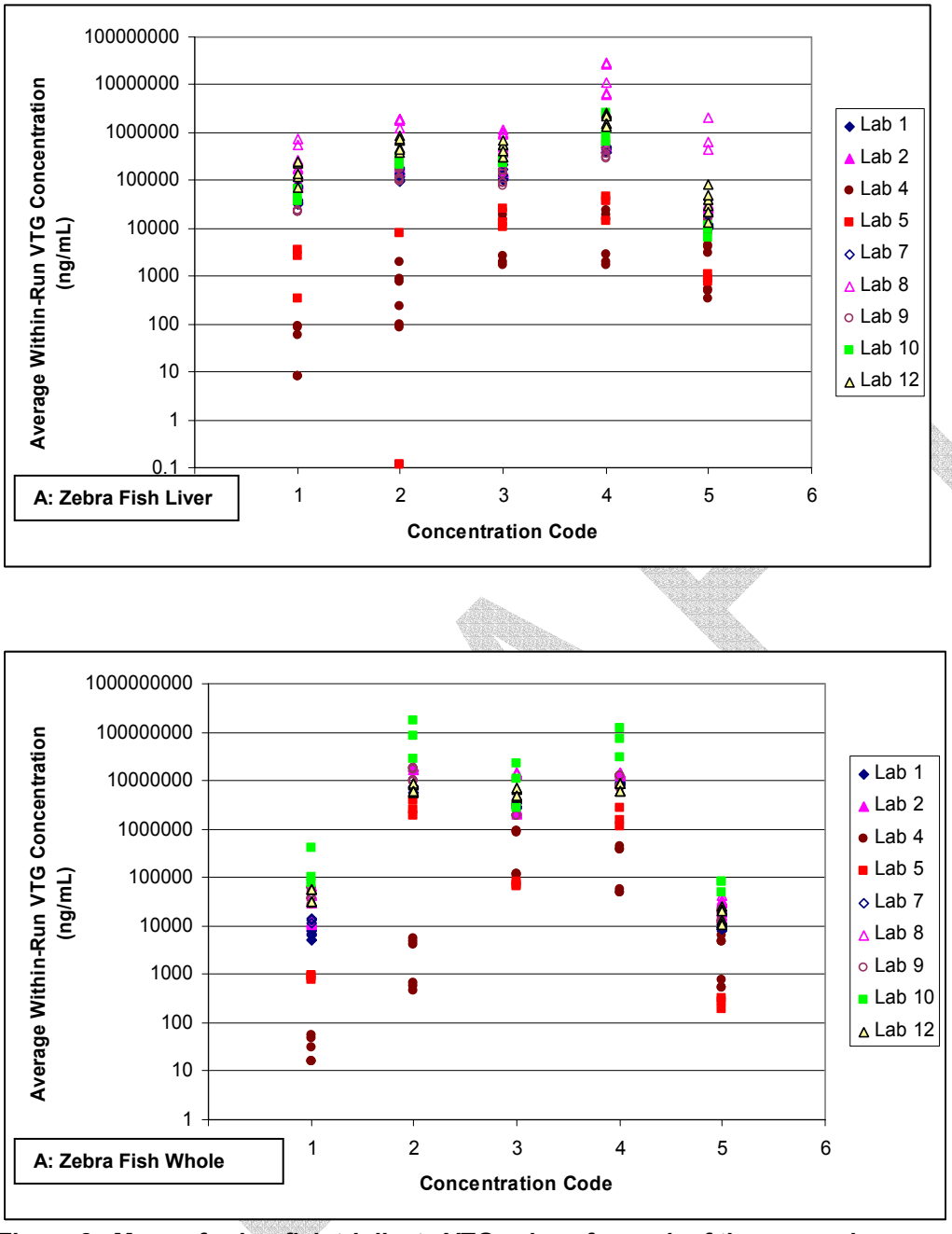



**Figure 2. Mean of zebra fish triplicate VTG values for each of three samples per concentration in the standard series, plotted by laboratory: (a) liver; (b) whole body** (Code 1 = Uninduced Male; Code 2 = Uninduced Female; Code 3 = Induced Male; Code 4 = Induced Female; Code 5 = Positive Control)

### *Comparison by Method*

The ranked average (natural log-transformed) VTG concentrations of the samples were compared using Tukey's HSD to determine whether methods vary in their ability to detect differences between pairs of treatment concentrations (see Tables C13-C16 in Appendix C). For the liver samples and homologous standard, the direct noncompetitive sandwich ELISA (ZF-Method 3) failed to detect differences among any of the six possible pairwise comparisons; the other three methods did only slightly better, each distinguishing one of the six pairs. In contrast, for whole body samples, all four methods detected differences in VTG values in at least four of the six possible pairwise comparisons and could distinguish differences between uninduced males, which represent the low end of the concentration spectrum, and the three other treatments, and between uninduced females and induced males, which represent the middle ranges. One significant difference was detected by ZF-Method 1 between uninduced males and induced females, which should represent the maximum spread of concentrations, zero to high, respectively, for liver samples using the purified standard. With the purified standard, ZF-Methods 2 and 3 failed to detect differences among any of the pairwise comparisons for liver samples.

For whole body samples and homologous standard, all three analytical methods detected differences in VTG values for at least three of the six pairwise comparisons. All three methods distinguished uninduced males from the other treatments, again with focus on the low end of the concentration range. Only ZF-Method 2 could detect the difference between uninduced females and induced females, and none of the methods saw a difference between induced males and induced females at the high end of the concentration gradient. The strongest pattern that emerges from the detailed outcomes in this test is the contrast between liver and whole body tissue results.

### *Within-Method 1 Comparison—Multiple Laboratories Applying the Same Method*

Four laboratories used ZF-Method 1, the sandwich ELISA with anti-zebra fish VTG capture antibody and detecting antibody, using the homologous standard for liver and whole body (Tables C17, C18 in Appendix C). The detection of differences among treatment concentrations depended first on the laboratory conducting the analysis, and next on the sample type and standard used. For liver homogenate, three laboratories distinguished the concentration differences for five of the six possible pairwise comparisons. No laboratory detected the expected midrange values (uninduced females, induced males). Three of the four laboratories reported results for whole body samples, but the results were variable: for each of the six pairwise comparisons, at least one laboratory detected a significant difference.

For the purified standard for liver homogenate, the results were similar to those above. Six laboratories used the same method (ZF-Method 1), and the detection of differences depended on the same factors in the same order as in the case of homologous standard use. Five of the six laboratories failed to detect differences for all of the six possible pairwise comparisons (Table C19 in Appendix C); one laboratory detected significant differences in two pairs. Only four of the six laboratories reported results for all pairwise comparison for whole body, and the ability to detect pairwise differences varied considerably (Table C20 in Appendix C). No laboratory detected

differences between uninduced and induced females, but did detect significant differences between three other pairs.

### *Positive Control Results*

Each lab analyzed a positive control precisely spiked with purified zebra fish VTG, the concentration of which was unknown to the analysts. As prepared, the concentration of the positive controls was 6.25 µg/mL (=6250 ng/mL) liver and whole body homogenate (see Tables C21-C23 in Appendix C). Although the within-laboratory variability was relatively low (75% of CVs were below 45% for liver samples and below 20% for whole body), the interlaboratory variability, and that among laboratories and methods was high in the percentage recovery of the spiked concentration from liver samples. For example, percentage recovery from liver homogenates was 56% to 20,073% for the homologous standard, and from 7% to 6647% for the purified. The variability for whole body analysis was lower, but nonetheless, >100% recovery was seen in most samples. It appears that the outcome of tests could be influenced by the choice of method in particular, as well as by the performing laboratory and other factors considered here.

## Medaka

Six laboratories analyzed medaka tissue homogenates by three different methods; one among these participants analyzed three sets of medaka samples by all three methods, bringing the number of data sets to eight, and providing a unique opportunity for a comparison of assays that reduced interlaboratory variability. The medaka analyses were similar to and concurrent with performance of those of zebra fish tissues, and two of the ELISA methods followed the same or parallel protocols specifically tailored to medaka. See Table 7 for the codes assigned to categorize relative VTG concentrations in the standard series, the two standards used for calibration, and the three methods applied to medaka. The methods are described and associated with specific participants in Table 3 and presented in full in Appendix A. Not all laboratories used both standards: five of the six laboratories that analyzed medaka homogenates employed the two standards, as requested; one used only the purified medaka VTG supplied by Battelle (Table 8).

### *Trends of Standard Series: Measured VTG Concentrations*

It was anticipated that male fish that were not induced by exposure to estrogenic compounds would provide minimal levels of VTG in their tissues, and that uninduced female fish, induced male, and induced female fish would generate increasing levels of VTG in their respective systems. In consideration of all reported results, the general trend for liver samples fit the expected serial increase in general: uninduced male<uninduced female<induced male<induced female medaka VTG concentrations (Figure C13 in Appendix C). (The VTG concentration in the precisely spiked positive control, which was 6.85 µg [=6850 ng] VTG/mL liver homogenate, and 6.25 µg [=6250 ng] VTG/mL whole body homogenate for medaka [Table 2], could have been expected to fall between the zero-to-low uninduced male and the uninduced female levels.) Exceptions to this consistent performance were as follows: three of the eight analyses from six laboratories

**Table 7. Summary of codes for relative VTG concentration, standards, and ELISA methods for medaka**

| Description | Code |
|---|---|
| **CONCENTRATION** | |
| Uninduced Male | 1 |
| Uninduced Female | 2 |
| Induced Male | 3 |
| Induced Female | 4 |
| Positive Control | 5 |
| **STANDARD** | |
| Homologous | 1 |
| Purified | 2 |
| **M-METHOD AND REFERENCE[a]** | |
| 1 (Biosense 2003) | 1 |
| 2 (Transgenic 2002) | 2 |
| 3 (EnBio 2002; Nishi et al. 2002) | 3 |

a) M medaka; see description of assay methods in Table 3.

**Table 8. Summary of standards employed by participating laboratories analyzing medaka**

| Lab ID | Participating Laboratory | Homologous standard[a] | Purified medaka standard[b] |
|---|---|---|---|
| **1** | University of Kumamoto | X | X |
| **2** | Biosense Laboratories | X | X |
| **4** | EnBioTec Laboratories, Ltd. | X | X |
| **5** | Notox Safety & Environmental Research | | X[c] |
| **11** | National Institute for Environmental Studies | X | X |
| **12** | Los Angeles County Sanitation Districts | X | X |

a) The standard routinely employed by the individual laboratory with respect to the assay in use.
b) Commercially prepared, purified VTG supplied by Battelle to participants.
c) The standard supplied by Battelle was the same as that used by the laboratory; accordingly, the laboratory elected to employ only one standard, providing one set of data for its sample set.

found VTG concentrations in induced female livers to be lower than those of induced males. This is contrary to the logic of the standard series; however, it will be valuable to explore further the possible explanations, based perhaps on condition of the fish—for example, the medaka females were reported in the laboratory notebook to be actively laying eggs at the time of exposure—or possibly on dose-related response to estrogen exposure. The exposure concentration varies from study to study; here, 300 ng/L; Nishi et al. (2002) used 10 ng/L (medaka); Brion et al. (2002) tried a range from 0.1 µg/L to 100 µg/L (zebra fish); and Holbech et al. (2001) used 10 ng/L exposure (zebra fish). For the whole body results, the pattern was as would be expected for six of the eight reports: the VTG concentration sequence was uninduced female<induced males<induced females. These results appear to be a more promising indication of utility of medaka over zebra fish for screening of VTG induction.

### *Difference in Standards*

The comparison of data by standard used offers a perspective on method specificity and standardization in a screening assay. The percentage difference between the average replicate VTG concentrations obtained with the homologous (H) and the purified (B) standard was calculated as (H-B)/B X 100%. Negative values represented greater VTG concentrations obtained with the purified standard. Five of the six laboratories employed both standards; one used only the purified medaka standard. This comparison showed relatively small differences in the results generated with use of the two standards for all three medaka methods. For both liver and whole body homogenates results, there was generally <60% difference (see Tables C28-C31 and Figures C13-C19 in Appendix C; Appendix F).

The VTG values measured in both liver and whole body samples for the homologous standard varied primarily with analytical method, and secondarily on the number of laboratories conducting the analyses. For both tissue homogenates, VTG values tended to be highest for the M-Method 1, which represent a sandwich ELISA with anti-medaka VTG capture antibody and detecting antibody, for all concentrations, and lowest measured by M-Method 2, which is a monoclonal and polyclonal-antibody-based direct sandwich ELISA.

For liver, M-Method 1's measured concentrations were four to seven times greater than those by M-Method 2, and 1.4 to 9 times greater than those measured by M-Method 3, which is a monoclonal-antibody-based sandwich ELISA. When the purified standard was used, VTG values measured in liver and whole body samples varied widely by analytical method, primarily, and secondarily by number of laboratories conducting the analyses or by laboratory, respectively. For both tissue homogenates, M-Method 1 yielded higher VTG concentration measures (≤7% for liver, ≤13% for whole body) than did either of the other methods.

Multiple laboratories applied the same, commercially available medaka methods (M-Methods 1, 3), and a high interlaboratory variability (CVs from 62% to 126% for liver; 63% to 195% for whole body) could be seen in the values attained with the homologous standard. The highest variability was seen for the positive controls for liver, and for uninduced male in whole body homogenate. Similarly for the purified standard with liver and whole body homogenate, the CVs ranged from 10% to 108% (liver) and 12% to 184% (whole body), with the highest variability occurring in positive controls for liver, and in induced male samples for whole body.

### *Within-Run Variability*

The three analytical replicates provide a measure of the within-run variability (Table C26 and Figures C11,C12 in Appendix C; Appendix D). When all of the methods were applied to the full range of samples, both the medaka liver and whole body showed a wide range of variability (CVs 0.8% to 100.4%; mean 9.1%, and 04% to 125.7%, mean 11.9%, respectively) among analytical replicates. However, because 75% of the CVs were <14% for both homogenates, and both tended to be <30% for uninduced females and induced males, this comparison indicates relatively low variability in the expected midrange VTG concentrations, but a somewhat higher variability in the top and bottom ranges. However, because 75% of the CVs were <10% for both

homogenates, and both tended to be <30% for uninduced females and induced males, this comparison indicates relatively low variability in the expected mid- and higher VTG concentrations, but a somewhat higher variability in bottom ranges. The variability can be graphically seen in Figure 3: most of the values are grouped tightly for the within run results for each laboratory, particularly for the medaka liver samples. The exception is for a single low, outlier result from Laboratory 12 in Code 4 the liver samples (Figure 3a). In the whole body samples, there is a wider range for Laboratory 5's Code 1 results (two moderately high outlier values) (Figure 3b). This plot also emphasizes the general trend of measured VTG values, which follows the expected concentration curve of the standard series particularly well in the medaka liver samples, and nearly as well in the whole body.

### *Intra-Assay Variability*

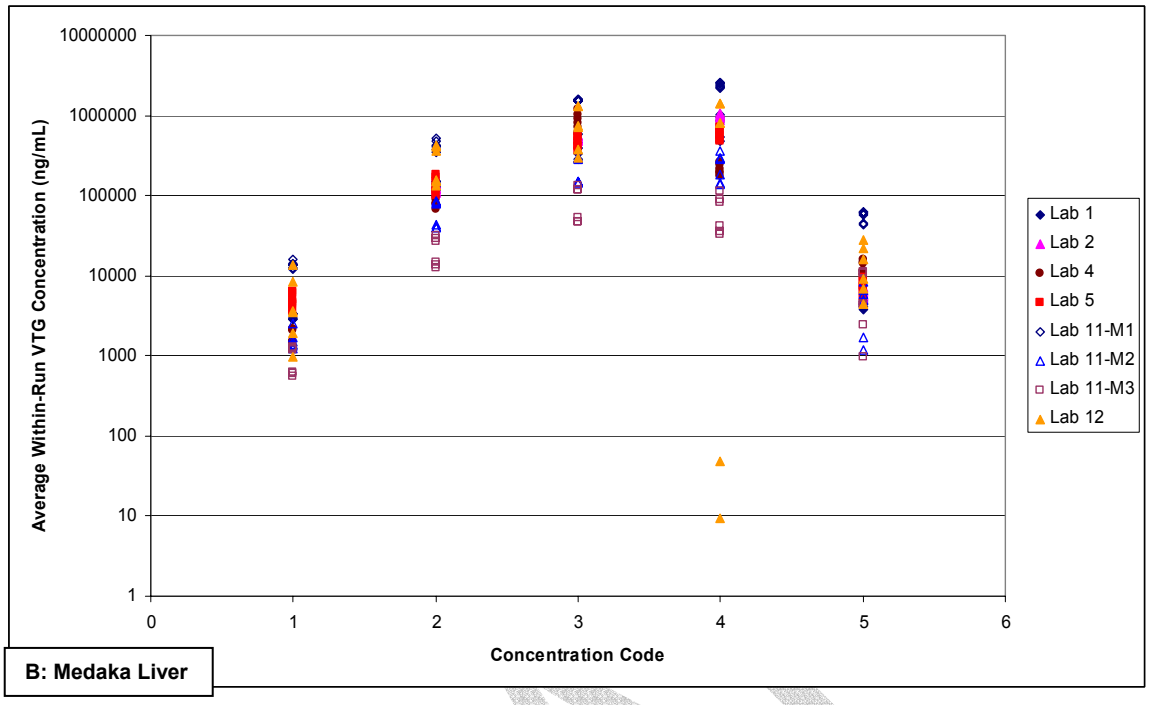
The analysis of sample triplicates provided a measure of intra-assay variability (Table C27 in Appendix C; Appendix F). The range of CVs was broad for both liver and whole body homogenates (0.7% to 158.8%), and 75% of the intra-assay CVs were <41% for both types of tissue. This level of intra-assay variability follows a trend parallel to that of zebra fish results, indicating that when a sample is provided to multiple laboratories employing a variety of methods (the results using multiple standards are also included in this sample set) the methods provide a relatively high degree of variability when replicate samples are analyzed. To further examine this type of variability, explored here by analysis of all results, the data were further examined by individual laboratory.

### *Comparison by Method*

The ranked average (natural log-transformed) VTG concentrations of the samples were compared using Tukey's HSD to determine whether methods vary in their ability to detect differences between pairs of treatment concentrations (see Tables C32-C35 in Appendix C). For the liver and whole body samples and both standards, the results were identical: M-Methods 1 and 3 could detect differences on the low end of the concentration gradient, namely, between the lowest VTG concentration group, the uninduced males, and each of the following: uninduced females, induced males, and induced females. These are important distinctions for discovery of induction in males, as would be applied in endocrine disrupter screening programs. None of the methods could detect differences at the high end to distinguish induced females from induced males, and in the midrange, uninduced females from induced males and induced females.

### *Between-Method Comparison—Single Laboratory Applying Multiple Methods*

Laboratory 11 used all three methods, allowing a comparison of the methods under relatively consistent conditions (Figures C20, C21 and Tables C36, C37 in Appendix C). For both sample types, the VTG concentrations averaged over standards were the highest measured by M-Method 1, and lowest by M-Method 3, and the variability was small (most CVs <20%). There were no differences between the two standards by any analytical methods for the measurement of



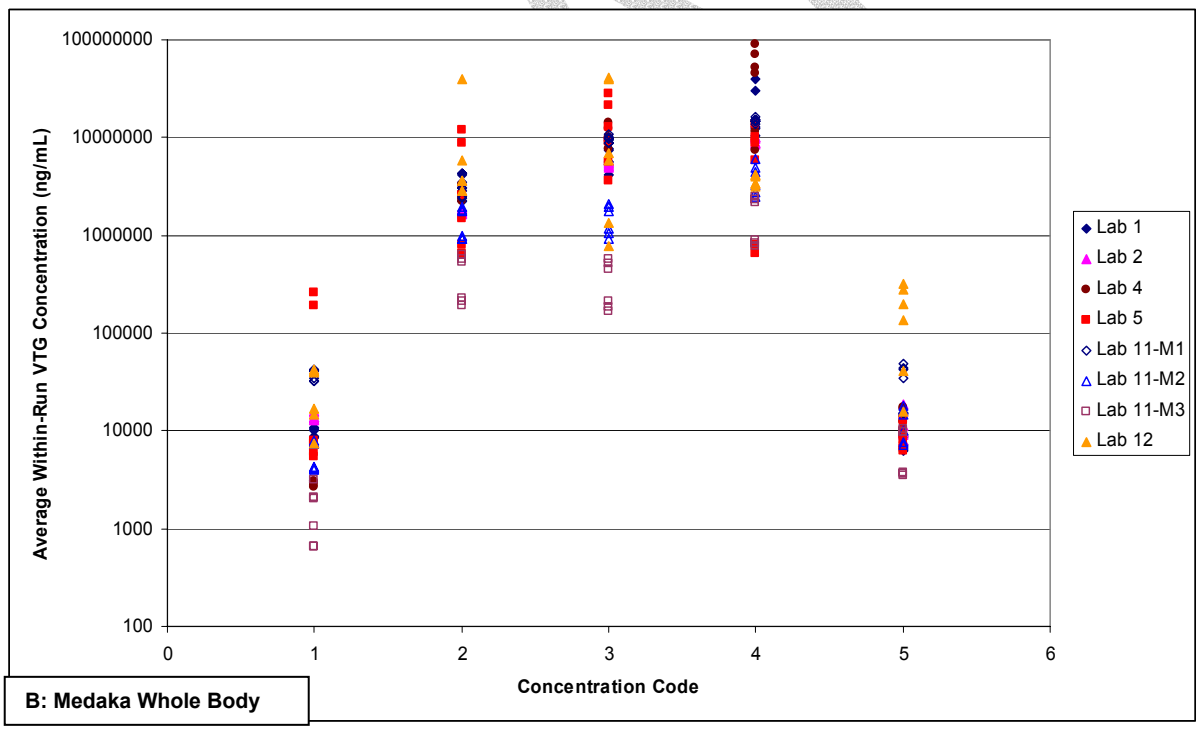




**Figure 3. Mean of medaka triplicate VTG values for each of three samples per concentration in the standard series, plotted by laboratory: (a) liver; (b) whole body** (Code 1 = Uninduced Male; Code 2 = Uninduced Female; Code 3 = Induced Male; Code 4 = Induced Female; Code 5 = Positive Control)

VTG in liver homogenates. In contrast, for whole body, the VTG concentrations did not vary between standards for M-Method 1, varied for some concentrations in the standard series by M-Method 2, and differed for each and every concentration by M-Method 3. The pairwise comparisons analyzed by Tukey's HSD on natural log-transformed VTG levels showed that all three methods could detect differences in liver and whole body VTG levels for at least four of the six possible pairs for the homologous standard, and for five of the six possible pairs for the purified standard. This comparison demonstrates the potentially large degree of variability among laboratories that could account for some of the observed disparity of results.

***Within-Method Comparison—Multiple Laboratories Applying the Same Methods***

Two analytical methods, M-Method 1 and M-Method 3 were exercised by more than one laboratory in the medaka study, providing the opportunity to evaluate within-method variability using the homologous standard for liver and whole body (Tables C38-C43 in Appendix C). Five laboratories used M-Method 1 to determine VTG concentrations in liver and whole medaka samples using the homologous standard. The ability of the assay to detect differences among the treatment concentrations depended on the laboratory performing the analyses, the sample type, and the standard that was used. Four of the five laboratories that reported liver results for all six comparisons, and all five that reported whole body results failed to detect differences between uninduced females and either induced males or induced females, and between induced females and induced males in the middle and high range of the concentration gradient. For liver, four of five laboratories' results distinguished uninduced males from induced males, and for whole body, all five could distinguish not only induced males, but also uninduced females and induced females from the uninduced males, at the low range of the gradient.

In comparison, when purified standard was used, the ability to discern differences in VTG concentrations in both liver samples depended on the laboratory performing the analysis, the sample type, as well as the standard used. Notably, one laboratory detected significant differences in VTG concentrations only between uninduced males and induced males, at the low concentration range, and the list of failed detection corresponds to that resulting from the homologous standard, above. All six laboratories that used M-Method 1with purified standard to analyze whole body samples detected differences in the same pairs as the whole body, homologous standard results listed above: between uninduced males and uninduced females, between uninduced males and induced males, and between uninduced males and induced females (Table C41 in Appendix C). All six laboratories similarly failed to detect differences between uninduced females and induced males, between uninduced females and induced females, and between induced males and induced females.

Two laboratories (4 and 11) used M-Method 3 to analyze VTG concentrations in medaka. Whether the homologous or the purified standard was used, both laboratories detected differences both in liver and in whole body VTG concentrations for five of the six possible pairwise comparisons (Tables C42, C43 in Appendix C). In the case of both standards applied to whole body analysis, Laboratory 11 did not detect a difference in VTG concentration between uninduced females and induced males, in the midrange of the concentration gradient. Nonetheless, this is the most positive result for any of the methods with respect to within-method variability.

***Positive Control Results***

Each lab analyzed a positive control precisely spiked with purified medaka VTG, the concentration of which was unknown to the analysts. The VTG concentration in the precisely spiked positive control was 6.85 µg (=6850 ng)VTG/mL liver homogenate, and 6.25 µg (=6250 ng) VTG/mL whole body homogenate for medaka (Table 2). Although the within-laboratory variability was relatively low, 75% of CVs were below 42% for liver samples and below 41% for whole body, the interlaboratory variability, and that among laboratories and methods was moderate (lower than the comparable zebra fish figures) in the percentage recovery of the spiked concentration from liver samples (Tables C44, C45 in Appendix C). For example, percentage recovery from liver homogenates was 52% to 863% for the homologous standard, and from 76% to 882% for the purified. The variability for whole body analysis was greater, from >100% up to 1873% (homologous) and up to 3418% (purified) recovery (Table C46 in Appendix C). The variability was in the orders-of-magnitude range for the positive control test results. It appears that the outcome of tests could be influenced primarily by the choice of method, and secondarily by the performing laboratory and other factors considered in these statistical comparisons.

## 8.0 DISCUSSION

Zebra fish and medaka are among the fish under consideration to be test species for endocrine disrupter research. Both species are well characterized, sensitive to exposure to hormones and other endocrine disrupting compounds, and suitable for laboratory handling. The present study surveyed existing ELISA methods that are currently available to detect the protein, VTG, in liver and whole body homogenate of zebra fish and medaka. It was initially intended that blood plasma would also be analyzed by one laboratory. Although the plasma samples were collected and preserved, circumstances precluded the participation of the particular group that could have conducted the analysis, and plasma was therefore not included in the survey.

Methods routinely performed by the participating laboratories were applied to a standard series of samples of one or both species of fish. Each series consisted of tissue homogenates in four categories by gender and treatment of fish expected to yield a gradient of VTG concentration from low (zero) to high. In addition, a positive control was prepared from unexposed male tissue spiked with a known quantity of purified, species-specific VTG as the fifth sample to be included in each analytical series. All of the samples were provided blind-coded with respect to the concentration series and control, but clearly identified as zebra fish or medaka, liver or whole body homogenate. Because the samples were blind-coded and represented a potentially wide range of concentrations, it required multiple dilutions of the sample to ensure a response within the working range of the assays. Accordingly, a significant investment of time and resources was needed, which was kindly donated by the participating laboratories to aid in reaching the goals of the study. The contributions of the participants are gratefully acknowledged.

The study was not intended to validate a given method, protocol, system, or technique, but rather, it was meant to survey the current methods and to discern the relative variability among them. The results obtained from the use of the particular method, by circumstance, by a statistically valid number of laboratories should not be used to assess the strength or weakness of this method

compared with others within or outside of the present survey. Rather, it should be assumed that the variability encountered in results would be found with the application of any one of the methods in this study by multiple laboratories. The use of trade names, identification of laboratories, and methods described in this do not constitute endorsement by EPA or Battelle Memorial Institute.

A standard series of liver and whole body homogenate samples representing a range of VTG concentrations in male and female zebra fish and medaka, and a set of positive controls were generated for this study. A sample repository/archive was created and maintained. Sets of aliquots representative of the standard series and positive controls were prepared for shipping to 11 laboratories around the world. Along with a commercially prepared, purified VTG standard for each species, documentation, and information, and they were shipped in coordination to the participating labs. The integrity of the samples was maintained at all times by proper cold-temperature control (maintained at -80°C), both in storage at the preparatory laboratory and in transit in special, well-insulated boxes to the participating laboratories, and by instruction, during storage and use at the participating laboratories. Ten of the laboratories returned results that were included in the statistical comparison; the eleventh laboratory was unable to complete the analysis in time for inclusion in this analysis, but the original data sheets and protocol are nonetheless presented in the appropriate appendices of this report.

Nine of the laboratories analyzed zebra fish homogenates, and six laboratories analyzed medaka; however, the number of data sets for medaka was brought to eight, because one of the participants analyzed three sets of medaka samples by three different methods. Four ELISA methods were applied to zebra fish homogenates, and three to medaka. The methods could be generally grouped as various sandwich ELISAs based either on monoclonal, polyclonal, or both types of antibody; and two direct, noncompetititve sandwich ELISAs using anti-zebra fish lipovitellin and polyclonal antibody. The eleventh laboratory, the results of which were not included in the statistical analysis, used a competitive binding assay. Three of the ELISA methods were commercially available kits; the others were unique.

The statistical analyses in this report address the within-run variability, the intra-assay variability based on the mean triplicate result, and the general trend of the ELISA VTG results associated with the standard evaluation series of fish liver and whole body homogenates. This series was represented by 1) uninduced male, 2) uninduced female, 3) induced male, and 4) induced female zebra fish or medaka, respectively. In addition to the standard series, a set of positive control VTG results are summarized. The distribution of CVs of the resulting triplicate mean VTG concentrations are summarized for a given concentration, laboratory, and standard; across laboratories, standards, and assays for a given concentration; and by method for a given laboratory, standard, and concentration.

There was substantial variation in the reported quantitative ELISA results in this study for both fish species. However, in spite of a range in the absolute values measured, the trends for the concentration series values in both liver and whole body of both species tracked the expected values fairly well and were generally consistent (of similar slope) within each series, namely, liver or whole body for each species (Figure 4). The curving slopes of the medaka liver most precisely

track the standard series expected slope, with the lowest VTG value for uninduced males (Code 1), increasing with each category to peak at induced females (Code 4), and to drop lower again for the positive control (Code 5). The "M"-shaped tendency of the slopes in the whole body curves for both species would indicate that in these cases, the uninduced females exhibited higher VTG concentrations than did induced males, and that there was a small, or no difference between the values in females, whether induced or not. There are indeed biological factors that could account for these results: for example, the females of both species were laying eggs at a high rate during the period of the experiment, and could therefore have had elevated VTG levels in any case, or have been more resistant to effects from exposure to estrogen-like compounds in the environment (Irv Schultz, personal communications, July 2003). Brion et al. (2002) reported substantial difference in VTG levels (from 3.97±2.7 μg/mL to 442.5±180 μg/mL) of unexposed females depending on their state of maturity.

The degree to which one must focus on the absolute value of the VTG measurements measured by ELISA must be tempered by the understanding that there are sources of variation within the tests that can influence the precision of quantification. For example, as the natural degradation of VTG occurs in the samples, the large molecule may be broken down into several smaller components, and the VTG antibody might detect all of the "pieces" and count each as a molecule of VTG, resulting in higher than expected values. Similarly, the VTG antibody could cross-react with other proteins in the sample, and detect them as VTG when they are not actually the specific target.

There was a wide range of within-run variability, but 75% of the coefficients of variation were low, <10% and <14% for zebra fish and medaka, respectively. That is, it was typically one of the methods out of four that accounted for some outlier values that increased the coefficient of variation, whereas the others showed very low variability, and their within-range values were closely clustered. The overview of within-run variability shown in Figures 2 and 3 above (see 7.0 Results) graphically demonstrates this observation, and confirms the fit to the general trend of concentrations in the standard series.

The range of intra-assay variability among the methods in the study was also broad for both zebra fish and medaka, with coefficients of variation almost three times higher (<30% for zebra fish, and <41% for medaka) than those in the within-run tests, but still moderate. Although the absolute VTG concentrations measured by the various methods varied by orders of magnitude, the objective of applying ELISA methods to detect the standard series of concentrations in liver and whole body homogenates was generally met. In general, the better fit was shown by the medaka applications than by those used for zebra fish, and the medaka liver concentrations most closely represented the expected concentration series as analyzed by all methods for medaka.

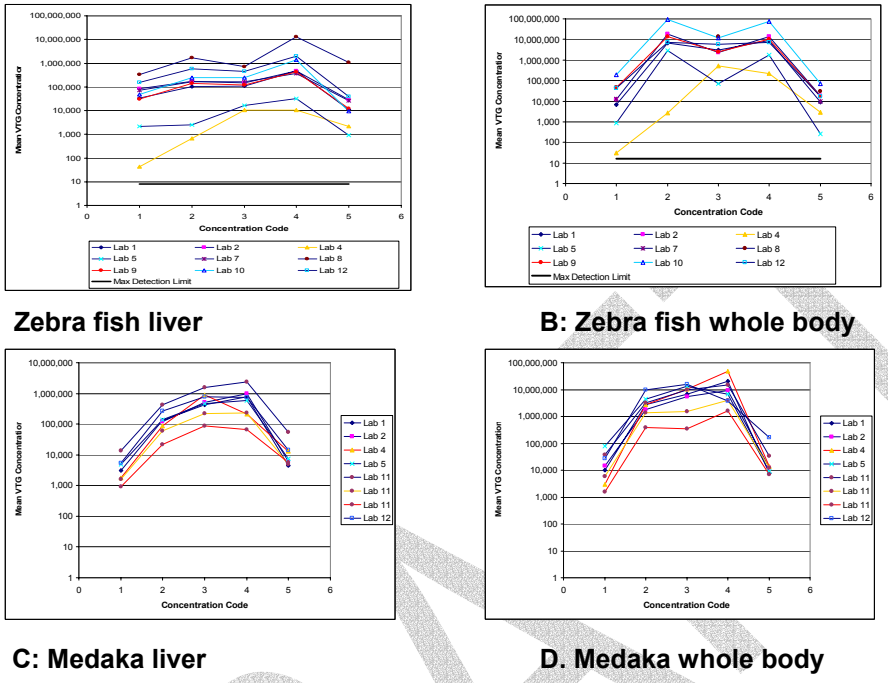



**Figure 4. VTG measurement by laboratory for each concentration code averaged over standard and assay for (a) zebra fish liver; (b) zebra fish whole body; (c) medaka liver; and (d) medaka whole body samples (see Figures C4, C5, C13, and C14 in Appendix C)** (Code 1 = uninduced male; Code 2 = uninduced female; Code 3 = induced male; Code 4 = induced female; Code 5 = positive control)

A recent EPA study compared VTG methods for fathead minnow (*Pimephales promelas*) whole body homogenate and blood plasma (EPA Work Assignment WA 2-19, Battelle 2003), and reported that the trend of the standard series followed the expected increasing levels of VTG for all methods in the comparison, even though there were large variations in the actual quantities of VTG measured. That is, one method consistently reported higher actual values than the others; however, the methods under consideration varied more dramatically in the types of antibodies used (one method used carp instead of fathead minnow antibody, for example), than did the methods used in the present study. In the present zebra fish and medaka study, there were more participating laboratories that used the same method for comparison; up to six laboratories applied a single method for zebra fish, for example.

The comparison of data by standard offers a perspective on method specificity and standardization in a screening assay. The percentage difference between the average replicate VTG concentrations obtained with the homologous and the purified standard was calculated. The comparison showed little variation between the two for medaka for all three methods applied; however, for zebra fish, the differences were small for all concentrations for two methods, but quite

large in the results of a third method. In the latter, VTG concentrations were about 700% to 800% lower for the purified than for the homologous standard.

The comparable, commercially available sandwich ELISA kit used by the majority of laboratories for both species yielded the highest VTG values. For example, for zebra fish, VTG concentrations measured by this method were up to 3277 times greater than those measured by another method for liver. The most dramatic example is perhaps the measurement of VTG in uninduced male zebra fish by different methods: VTG was undetected by one method, but was measured as 136,117 ng/mL by another, the aforementioned commercial ELISA kit. Nonetheless, there was a high degree of variability in the results from the multiple laboratories that applied the same method, both for zebra fish and for medaka. For both the purified and homologous standard in the zebra fish and medaka liver homogenate, the highest variability was seen in the positive controls, and for whole body, in the uninduced male (zebra fish) or induced male (medaka) samples. This demonstrates that particularly for zebra fish, the quantification of VTG is not absolute, but rather, depends primarily on the method used for measurement.

Statistical comparison by method to determine whether the various ELISAs vary in their ability to detect differences between pairs of treatment concentrations yielded a moderately strong distinction between liver and whole body homogenates in the zebra fish. With one method, differences in the low end of the concentration spectrum were more readily discerned for liver samples used with the homologous standard, whereas with the purified standard, only the highest and lowest values were significantly distinguished. For whole body samples, all four methods could detect differences in VTG values in at least two-thirds of the concentration pairs, and could distinguish between uninduced males, which represent the low end of the concentration spectrum, and the three other treatments that represent the middle concentration range. Analysis of whole body samples used with purified standard found that all three methods distinguished uninduced males from the other treatments, again with focus on the low end of the concentration range.

In medaka, for the liver and whole body samples and both standards, the results were identical: two methods could detect differences on the low end of the concentration gradient, namely, between the lowest VTG concentration group, the uninduced males, and each of the other three treatment concentrations. These are important distinctions for discovery of induction in males that would be the focus in endocrine disrupter screening programs, for example. None of the methods could detect differences at the high end to distinguish induced females from induced males, nor in the midrange, to distinguish uninduced females from induced males and induced females.

Because one laboratory used all three methods for medaka analysis, a comparison of the methods under relatively consistent conditions could be made. For both liver and whole body samples, very high VTG concentrations averaged over the two standards were obtained by application of one method, and very low concentrations by another method. There were no differences shown by any analytical method when one or the other of the two standards was used for the measurement of VTG in liver homogenates. In contrast, for whole body, the VTG concentrations did not vary between standards for one method, varied for some concentrations in the standard series by a second method, and differed for each and every concentration by a third

method. All three methods could detect differences in liver and whole body VTG levels for at least four or five of the six possible pairs in the pairwise comparison using both standards. Altogether, the one-laboratory three-method comparison demonstrates by contrast the potentially large degree of variability among laboratories that could account for some of the disparity of results reported here.

Two analytical methods were exercised by more than one laboratory in the medaka study, providing the opportunity to evaluate within-method variability. For example, five laboratories used the same method to determine VTG concentrations in liver and whole body medaka samples using the homologous standard. The ability of the assay to detect differences among the treatment concentrations depended primarily on the laboratory performing the analyses, secondarily on the sample type and by the standard that was used. Two laboratories used a second method for analysis, by which both laboratories detected differences both in liver and in whole body VTG concentrations for five of the six possible pairwise comparisons, regardless of which standard was used. Again, the ability of the assay to detect differences among the treatment concentrations depended primarily on the laboratory performing the analyses, and next on the sample type and the standard that was used. Nonetheless, this is the most positive result for any of the methods with respect to within-method variability. For zebra fish, up to six laboratories used a single method for liver and whole body, using one or the other standard. As in the previous comparisons, detection of differences among treatment concentrations depended first on the laboratory conducting the analysis, and next on the sample type and standard used.

In both the zebra fish and medaka studies, a positive control precisely spiked with purified VTG was analyzed. The concentration, which was toward the lower end of the concentration gradient of the standard series, was unknown to the participating laboratories. As an example, although the within-laboratory variability in the percentage recovery of the spiked concentration from liver samples was relatively low, the variability among laboratories and among methods was moderate (medaka) to high (zebra fish). The variability was in the range of several orders of magnitude for these positive control test results. However, similar results are seen in other studies. For instance, Brion et al. (2002) reported that in a similar ELISA study, the control group yielded VTG concentrations that were in some fish less than the practical detection limit for the assay (i.e., <40 ng/mL), whereas in others, as high as 560 ng/mL. The relative increase of VTG in males with exposure to an estrogen compound, however, is a clearly marked trend, in spite of the variability among controls, up to an 18,000-fold increase in VTG for exposed males Brion et al.'s study (2002). Holbech et al. (2001) saw a 200-fold increase in VTG levels in males exposed for 7 days to a nominal concentration of ethinylestradiol, compared with levels in controls. Of course, the compound and the concentration thereof used for the exposures, along with other experimental conditions and techniques vary from one study to another, and differ from the specifications of the present study. The increases seen among induced compared with uninduced males in the present study were typically in the range of one to more than five-hundredfold among medaka, and less than one hundredfold among zebra fish. It appears that the quantitative result in the case of positive controls in this study could be influenced or confounded primarily by the choice of method, and secondarily by the performing laboratory and by the selection of standard for calibration, with respect to the two species of fish that were the focus of this survey.

In the categories compared statistically in this study, the sources of the variability could not be precisely defined based on the present results. Due to the design of the study, not all factors could be addressed; as an example, potential confounding factors associated with storage, shipping, and handling of samples were not evaluated. The possible consequences of time and resource restrictions were also not considered. Because all of the participating laboratories donated their labor, materials, and other resources to conduct the analysis, it is possible that some portion of the interlaboratory variability could have grown from consequent limitations—that is, for example, perhaps the number of dilutions that could be run for each sample varied among laboratories, depending on their available time and materials. Differences could arise from any number of other potential sources, such as technical issues in the protocols, or particular characteristics of a laboratory's performance of the tests. There could be concern about the range of method detection limits, or the methods by which they are determined. Other issues could include instability of samples, reagents, or other materials that are stored, or inherent variability among the source populations of fish used in the study, as well as other biological factors affecting the fish, among others. Understanding the factors capable of contributing to the high degree of variability observed in the present survey would be a valuable contribution to scientific progress in this area.

Various researchers have demonstrated the utility of VTG as a biomarker for providing evidence of endocrine disruption in fish. Based on the results of this study, it can be said that most of the laboratories and methods considered are capable of distinguishing changes in VTG levels in zebra fish and/or medaka. However, there are issues still to be resolved before VTG measurements could be used as a reliable tool in screening and testing. It is recommended that greater effort be given to developing specific performance criteria for VTG analytical methods. It is also recommended that a single, standardized protocol for each fish species be used in quantifying VTG in the interlaboratory validation trials.

## 9.0 REFERENCES

Battelle. 2002. *Draft Detailed Review Paper, Fish Screening Assays for Endocrine Disruption.* Prepared for the U.S. Environmental Protection Agency by Battelle Pacific Northwest National Laboratory, Sequim, Washington.

Battelle. 2003. *Draft Final Report on Comparative Evaluation of Vitellogenin Methods.* Prepared for the U.S. Environmental Protection Agency by Battelle Pacific Northwest National Laboratory, Sequim, Washington.

Brion, F., B.M. Nilsen, J.K. Eidem, A. Goksϕyr, and J.M. Porcher. 2002. Development and validation of an enzyme-linked immunosorbent assay to measure vitellogenin in the zebrafish (*Danio rerio*). *Environmental Toxicology and Chemistry* 28(8): 1699-1708.

Braunbeck, T., H. Holbech, L. Norrgren, and G. Petersen. 2003. Zebrafish as a test organism in endocrine disrupter research, pp. 9-14. In *Proceedings of the 3rd International Medaka Symposium, Development of Test Methods with Medaka to Detect Endocrine Disrupting Chemicals*, Session 1:

Each fish's activity of evaluation of endocrine disrupting chemicals, 27-28 February 2003, Okazaki, Japan. Available online: http://www.env.go.jp/chemi/end/medaka_sympo.html.

Edmunds, J.S., R. A. McCarthy, and J.S. Ramsdell. 2000. Permanent and functional male-to-female sex reversal in d-rR strain medaka (*Oryzias latipes*) following egg microinjection of o,p;-DDT. *Environmental Health Perspectives* 108(3): 219-224.

Holbech, H., L. Andersen, G.I. Petersen, B. Korsgaard, K.L. Pedersen, and P. Bjerregaard. 2001. Development of an ELISA for vitellogenin in whole body homogenate of zebrafish (*Danio rerio*). *Comparative Biochemistry and Physiology* Part C 130(2001): 119-131.

Kang, I.J., H. Yokota, Y. Oshima, Y. Tsuruda, T. Oe, N. Imada, H. Tadorokoro, and T. Honjo. 2002. Effects of bisphenol A on the reproduction of Japanese medaka (*Oryzias latipes*). *Environmental Toxicology and Chemistry* 21(11): 2394-2400.

Kordes, C., E.P. Rieber, and H.O. Gutzeit. 2002. An *in vitro* vitellogenin bioassay for oestrogenic substances in the medaka (*Oryzias latipes*). *Aquatic Toxicology* 58(3-4): 151-164.

Nishi, K., M. Chikae, Y. Hatano, H. Mizukami, M. Yamashita, R. Sakakibara, and E. Tamiya. 2002. Development and application of a monoclonal antibody-based sandwich ELISA for quantification of Japanese medaka (*Oryzias latipes*) vitellogenin. *Comparative Biochemistry and Physiology* Part C 132(2002): 161-169.

Transgenic. 2002. Medaka ELISA kit information. Available online: http://www.metocean.co.jp/new/inet/vol2/vol2_information.htm (in Japanese).

Van den Belt, K., P.W. Wester, L.T.M. Van der Ven, R. Verheyen, and H. Witters. 2002. Effects of ethynylestradiol on the reproductive physiology in zebrafish (*Danio rerio*): Time dependency and reversibility. *Environmental Toxicology and Chemistry* 21(4): 767-775.