VALIDATION OF THE AMPHIBIAN METAMORPHOSIS ASSAY AS A SCREEN FOR THYROID-ACTIVE CHEMICALS: INTEGRATED SUMMARY REPORT

ORGANIZATION OF ECONOMIC COOPERATION AND DEVELOPMENT

October 16, 2007

TABLE OF CONTENTS

1 INTRODUCTION 6

Endocrine Disruptor Screening Program *6*
Validation *8*

2 HISTORICAL OVERVIEW OF THE AMA 10

2.1 RELEVANCE OF THE AMPHIBIAN METAMORPHOSIS ASSAY 10
Purpose of the assay *10*
Rationale for the assay *11*
Scientific bases for the assay *11*
2.2 OVERVIEW OF AMPHIBIAN METAMORPHOSIS AND POINTS OF POTENTIAL DISRUPTION 12

3 PROTOCOL DEVELOPMENT 16

3.1 RATIONALE FOR THE USE OF *X. LAEVIS* IN THE AMA 16
Test Species *16*
3.2 SELECTION OF THE DEVELOPMENTAL PERIOD AND DURATION OF EXPOSURE FOR THE AMA 17
3.3 RATIONALE FOR THE CORE ENDPOINTS IN THE AMA 18
Developmental Stage *18*
Hind Limb Length *18*
Body Length and Wet Weight *19*
Thyroid Gland Histology *19*
3.4 RATIONALE FOR EXPOSURE SCENARIOS IN THE AMA 20
Number of Animals on Test and Replicates *20*
Exposure System *20*
Chemical delivery *21*
3.5 STATISTICAL ANALYSES 21
Special Issues for Statistics *22*
3.6 DATA INTERPRETATION 22
Potential Endpoint Outcomes: *23*

4 OPTIMIZATION OF THE TEST METHOD 24

4.1 OECD PHASE 1 STUDY 24
Overview of Methods *25*
Data Analysis *26*
Overview of Results – Control Organism Performance *26*
Overview of Results – PTU *28*
Overview of Results – T4 .... *33*
4.2 MULTI-CHEMICAL STUDY AND METHOD DEMONSTRATION – USEPA 38
Objectives *38*
Overview of Methods *39*
Quality Assurance *41*
Overview of Results – Controls *41*
Overview of Results – PTU *42*
Overview of Results – Methimazole *43*
Overview of Results – T4 .... *44*
Overview of Results – Dexamethasone *45*
Overview of Results – Phenobarbital *46*
Overview of Results – PCN *47*
Quality Assurance *47*
Endpoint Sensitivity *48*
Sensitivity of Exposure Protocol *48*
Conclusions – Multi-Chemical Study and Method Development *49*

5 INTER-LABORATORY EVALUATION OF THE AMA (PHASE 2) 50

5.1 STUDY DESIGN 51
Statistical Analyses *51*
5.2 PERFORMANCE OF THE CONTROLS 52
Developmental Stage *52*
Growth Metrics *55*
5.3 SUMMARY OF RESULTS – SODIUM PERCHLORATE 55
Mortality *56*
Growth *56*
Development *58*
Thyroid Gland Histology *58*
5.4 SUMMARY OF RESULTS – T4.... 61
Mortality *62*
Growth *62*
Development *64*
Thyroid Gland Histology *67*
5.5 SUMMARY OF RESULTS – IOP.... 68
Mortality *69*
Growth *69*
Development *70*
Thyroid Gland Histology *71*
5.6 OVERALL ENDPOINT PERFORMANCE 74
Body Length and Weight *74*
Hind Limb Length *75*
Developmental Stage *75*
Thyroid Histology *76*
5.7 SUMMARY OF THE OECD PHASE 1 AND PHASE 2 STUDY RESULTS 76

6 OECD PHASE 3 STUDY – NEGATIVE AND WEAK POSITIVE 77

6.1 STUDY DESIGN 77
6.2 OVERVIEW OF RESULTS – E2 78
6.3 OVERVIEW OF RESULTS – BP-2.... 79

7 ADDITIONAL STUDIES SUPPORTING THE VALIDATION OF THE AMA 80

7.1 STUDIES USING X. LAEVIS.... 81
7.2 STUDIES USING SILURANA (XENOPUS) TROPICALIS 82
7.3 DISCUSSION 82

8 DISCUSSION/SUMMARY OF THE VALIDATION OF THE AMA 83

8.1 REPRODUCIBILITY OF THE ASSAY 83
8.2 RECOMMENDED REFINEMENTS TO THE METHOD 84
Dietary regime *84*
Water iodine level *84*
Dose levels *84*
Stage matching for histopathology *85*
Performance criteria.... *85*
Improved data interpretation criteria *86*
8.3 STRENGTHS AND LIMITATIONS OF THE ASSAY 90

9 REFERENCES.... 92

LIST OF TABLES

Table 1-1. Assays recommended for consideration for the Tier 1 screening battery.................................... 8
Table 1-2. Alternative assays for Tier 1. 8
Table 4-1. Overview of exposure studies conducted by German, Japanese and U.S. laboratories.............. 26
Table 4-2. Test termination summary data for control performance for the OECD Phase 1 pre-validation study. 27
Table 4-3. Summary data for the 21-day, NF stage 51 exposure scenario using PTU in the OECD Phase 1 pre-validation study. 30
Table 4-4. Summary data for the 14-day, NF stage 54 exposure scenario using PTU in the OECD........... 31
Table 4-5. Summary results for PTU from the OECD Phase 1 study. 33
Table 4-6. Summary data for the 21-day, NF stage 51 exposure scenario using T4 in the OECD Phase 1 pre-validation study. 35
Table 4-7. Summary data for the 14-day, NF stage 54 exposure scenario using T4 in the OECD Phase 1 pre-validation study. 36
Table 4-8. Summary data for T4 from the Phase 1 trial. 38
Table 4-9. Overview of nominal exposure concentrations used in the USEPA protocol demonstration and optimization studies. 40
Table 4-10. Primary and secondary endpoints for the USEPA protocol demonstration and optimization studies. 40
Table 4-11. Summary data for PTU for the USEPA protocol demonstration and optimization studies...... 42
Table 4-12. Summary data for PTU from Contractor lab coupled with results from Phase 1. 43
Table 4-13. Summary data for methimazole for the USEPA protocol demonstration and optimization studies. 43
Table 4-14. Summary data for T4 for the USEPA protocol demonstration and optimization studies......... 44
Table 4-15. Summary assay results for T4 coupled with Phase 1 results. 45
Table 4-16. Summary data for dexamethasone for the USEPA protocol demonstration and optimization studies. 45
Table 4-17. Summary data for phenobarbital for the USEPA protocol demonstration and optimization studies. 46
Table 4-18. Summary data for PCN for the USEPA protocol demonstration and optimization studies. 47
Table 5-1. Critical test parameters and conditions for the OECD Phase 2 inter-laboratory studies. 54
Table 5-2. Summary data for gross morphological endpoints from the OECD Phase 2 perchlorate study. 57
Table 5-3. Overall assay results - perchlorate.. 61
Table 5-4. Summary data for gross morphological endpoints from the OECD Phase 2 T4 studies. 63
Table 5-5. Overall assay results in Phase 2 - T4.. 68
Table 5-6. Summary data for gross morphological endpoints from the OECD Phase 2 IOP studies. 70
Table 5-7. Overall assay results - IOP. 74
Table 6-1. Overall assay results - E2. 78
Table 6-2. Overall assay results - BP-2. 80
Table 7-1. Summary of tests conducted with *X. laevis* that include analyses of developmental effects and thyroid gland histology. 81
Table 8-1. Performance criteria for the AMA. 86
Table 8-2. Overall assay results for PTU from the Phase 1 studies and Contractor studies using the refined data interpretation criteria. 87
Table 8-3. Overall assay results for T4 from Phase 1 studies and Contractor lab using the refined data interpretation criteria. 88
Table 8-4. Overall assay results for the multi-chemical study performed by a Contractor for four chemicals using the refined data interpretation criteria. 88
Table 8-5. Overall assay results for perchlorate from Phase 2 using the refined data interpretation criteria. 88
Table 8-6. Overall assay results for T4 from the Phase 2 studies using the refined data interpretation criteria. 89
Table 8-7. Overall assay results for IOP from Phase 2 using the refined data interpretation criteria.......... 89

LIST OF FIGURES

Figure 2-1. The association between thyroid hormone changes and metamorphosis. ... 13
Figure 2-2. Possible mechanisms of thyroid axis activity. ... 15
Figure 5-1. Incidence of core histopathological lesions associated with perchlorate exposure from the OECD Phase 2 study. ... 60
Figure 5-2. Incidence of core histopathological lesions associated with T4 exposure from the OECD Phase 2 study.. ... 66
Figure 5-3. Incidence of core histopathological lesions associated with IOP exposure from the OECD Phase 2 study.. ... 73
Figure 8-1. Flow chart describing decision logic for the interpretation of results obtained with a test substance in the AMA. ... 87

LIST OF ATTACHMENTS

Attachment A - Test Method for the Amphibian Metamorphosis Assay

Attachment B - Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1, Optimisation of the Test Protocol

Attachment C - Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 – Multi-Chemical Interlaboratory Study

Attachment D - Draft Report of the Phase-3 Validation of the Amphibian Metamorphosis Assay

Attachment E - Detailed Review Paper for the Amphibian Metamorphosis Assay

Attachment F - Endocrine Disruptor Screening Program Validation Paper

Attachment G - Power Analysis for Determining Study Design for the Amphibian Metamorphosis Assay

Attachment H - Final Report (Battelle) for the Multi-Chemical Study

OBJECTIVE: The objective of this integrated summary report is to provide a detailed account of the validation process for the Amphibian Metamorphosis Assay (AMA) so that peer reviewers may address the attached charge questions.

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 Act in 1996, 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 (1), the EPA expanded the EDSP using the Administrator's discretionary authority to include the androgen and thyroid hormonal systems as well as wildlife. Following broader international concerns and similar programs, 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 Programme. 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 Amphibian Metamorphosis Assay (AMA) is one of the thyroid-relevant screening assays proposed and is the subject of this report. This report complements the attached supporting materials and is meant to provide the peer reviewers with the necessary information, and/or direction to necessary information, to address the peer review charge questions. It introduces the purpose of the AMA and how it fits into the EDSP, the scientific rationale for the assay, and an historical account of the development and optimization of the assay protocol. It also synthesizes the information gained during the validation process and addresses the advantages and limitations of the AMA based on its biological strengths and weaknesses, practicality, reproducibility, reliability, and protocol transferability.

Endocrine Disruptor Screening Program

To comply with its mandate, the EPA chartered a Federal advisory committee (Endocrine Disruptor Screening and Testing Advisory Committee; EDSTAC) to develop 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, androgen, and thyroid (EAT) hormone axes. The ultimate goal of these assays is to provide input into hazard identification to assess risk of adverse consequences to humans and wildlife (1).

The EPA submitted a proposal of the EDSP for public review and comment (2) as well as peer review by a joint subcommittee of the EPA Science Advisory Board and FIFRA Scientific Advisory Panel (3). A complete description of the program proposal can be found in the Federal Register Notice (4). Briefly, the proposed EDSP 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 will become a battery in which some endocrine axis redundancy is incorporated (e.g. two different assays may cover some similar aspects of the thyroid axis). This redundancy will allow for a weight-of-evidence approach, as recommended by EDSTAC, to determine whether a chemical shall undergo further, more definitive, testing.

The following assays, recommended by EDSTAC, were to undergo validation prior to their inclusion in the Tier 1 screening battery (**Table 1-1**, **Table 1-2**). The recommended assays are meant to detect chemicals that may affect the estrogen, androgen and thyroid hormone axes through any known modes of action.

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.
Androgen receptor binding or transcriptional activation 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
Uterotrophic Assay	An *in vivo* assay to detect estrogenic chemicals.
Hershberger Assay	An *in vivo* assay to detect androgenic and antiandrogenic chemicals.
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.

In addition, EDSTAC recognized there were other combinations of assays that might substitute for some components of the recommended battery and also recommended that EPA validate the following assays as alternatives.

Table 1-2. Alternative assays for Tier 1.

Assay	Reasons for Inclusion
Placental Aromatase Assay	An assay to detect interference with aromatase.
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.
Adult Male	An *in vivo* assay designed 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.

Validation

Validation has been defined as “the process by which the reliability and relevance of a test method are evaluated for a particular use” (5;6).

> *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 (6). For Tier I EDSP assays, relevance can be defined as the ability of an assay to detect 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 (7). The criteria used to guide the validation process for

the AMA were based on the principles of validation developed by the U.S. Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) (5) and the OECD (7). These criteria as stated by ICCVAM (5) 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 inter-laboratory 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 attachment F (8). Although attempts have been made to thoroughly comply with all validation criteria, various *in vitro* and *in vivo* screening assays under consideration for the Tier 1 battery 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 is following a five-part or stage validation process outlined by the ICCVAM (5). The EPA believes that it is essential to recognize that this process was specifically developed for *in vitro* assays intended to replace *in vivo* assays. The fundamental 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 have no suitable *in vitro* substitute. Nonetheless, the stages of the process outlined by the ICCVAM are as follows:

The first stage of the process 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 assay, the context in which it will be used, and the scientific bases upon which the assay's protocol, endpoints, and relevance rest (9). This draft was subsequently finalized by the OECD (10), (attachment E). The DRP reviews the scientific literature for candidate protocols and evaluates them with respect to a number of considerations, such as whether candidate protocols meet the assay's intended purpose, costs, and other practical considerations. The DRP also identifies the

developmental status and questions related to each protocol; the information needed to answer the questions; and, when possible, recommends an initial protocol for the initiation of the second stage of validation, *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. The OECD Phase 1 trial report (attachment B) describes the studies that were performed for standardizing the AMA, study results, the consequent refinement of the protocol, and the initial information on protocol transferability. In the third phase, i*nter-laboratory validation,* studies were conducted in several independent laboratories with the refined protocol. The results of these studies were used to determine inter-laboratory variability and to set or cross-check performance criteria. The report on the inter-laboratory trials for the AMA is provided in attachment C. Inter-laboratory validation is followed by *peer review*, an independent scientific review by qualified experts, and by *regulatory acceptance*, 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 (11).

It should be remembered that even though assays are being validated and peer reviewed individually (i.e., their strengths and limitations are being evaluated as stand alone assays), the Tier 1 assays will, in fact, be used in a complementary battery of screens.

The purpose of this Integrated Summary Report is to provide a historical summary of the development and validation of a standardized protocol for the AMA proposed as an *in vivo* assay for the Tier-1 screening battery. The reasoning and judgments leading to the various studies, and conclusions concerning the strengths and weaknesses of the assay in its current form, are presented.

2 HISTORICAL OVERVIEW OF THE AMA

2.1 Relevance of the Amphibian Metamorphosis Assay

Purpose of the assay

The AMA is a screening assay intended to empirically identify substances which may interfere with the normal function of the hypothalamic-pituitary-thyroid (HPT) axis. The AMA represents a generalized vertebrate model to the extent that it is based on the conserved structure and functions of thyroid systems. It is an important assay in the EDSP screening battery because amphibian metamorphosis provides a well-studied, thyroid-dependent process which responds to substances active within the HPT axis, and it is the only candidate assay for the Tier 1 battery that assesses thyroid activity in an animal undergoing morphological development. It is important to recognize that the AMA is not intended to quantify or confirm endocrine disruption, or to provide a quantitative assessment of risk, but only provide suggestive evidence that thyroid regulated processes may be sufficiently perturbed to warrant more definitive testing.

Rationale for the assay

Clearly, a screen that will generate the information required to identify thyroid active toxicants must be based on what is known about thyroid endocrinology. Thyroid hormone is essential for normal development and for maintenance of normal physiological functions in vertebrates. Delivery of thyroid hormone to tissues and cells is highly regulated during development and in the adult, and it is governed by complex physiological processes involving the HPT axis and peripheral organs/tissues. Environmental factors, such as the presence of specific toxicants, can perturb this web at various points of regulation, inducing a variety of responses that can be captured in toxicological assays.

For an in depth discussion of the rationale for developing the AMA for the EDSP, please refer to the DRP on the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (10) and the DRP on Thyroid Hormone Disruption Assays (12;13). Briefly, the AMA is based on the principle that the dramatic morphological changes that occur during postembryonic development are dependent upon the normal functioning of the HPT axis, and that interference with these processes leads to measurable effects. During tadpole metamorphosis, thyroid hormone (TH) influences virtually every tissue in the animal's body initiating diverse morphological, physiological and biochemical changes that include cell proliferation, differentiation and death. The result is *de novo* organ formation, organ loss, and extensive tissue remodeling. Given the dependence of metamorphosis on TH, and the strict biochemical control under which these processes occur, the timing and character of these processes can serve as experimental endpoints representative of thyroid axis function, and as such, are exploited in the AMA. Additionally, although postembryonic development appears quite different in mammals and most amphibians (direct development vs. metamorphosis), there is a high level of evolutionary conservation of the thyroid system among vertebrates, and the underlying cellular and molecular pathways that control these processes are similar, if not identical. The evolutionarily conserved nature of the vertebrate thyroid system enhances the ability to use an amphibian, particularly anurans, as a general model for evaluating HPT axis interference that can be extrapolated to other vertebrate species.

Scientific bases for the assay

The general morphological, biochemical and molecular architecture of the thyroid axis arose early in vertebrate evolution and has been largely maintained through strong directional (positive) selection. Thus, specific aspects of the thyroid axis are conserved amongst most chordates at multiple organizational levels. The uniform identity of TH structure across vertebrate species suggests that synthetic pathways for TH are also highly conserved. Hence, the structural and functional similarities of the amphibian HPT axis to other vertebrates, particularly as they pertain to organization of follicular cells in the glands, transport mechanisms for iodide, enzymes involved in organification of iodide, and the uptake, cleavage, secretion and action of thyroxine (T4) and triiodothyronine (T3) suggest that a metamorphosing anuran species, such as *Xenopus laevis,* is a suitable vertebrate model system for thyroid disruption. For an in depth discussion on the

scientific bases for using amphibian metamorphosis as a mechanism to evaluate chemical interactions with the HPT axis, please refer to the Detailed Review Paper (DRP) on the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (10) and the DRP on Thyroid Hormone Disruption Assays (12). For a comprehensive discussion on endocrine control of amphibian metamorphosis, please refer to (14) and (15).

Generally, metamorphosis, which is primarily controlled by TH, is a period of substantial morphological change in which an organism alters its mode of living. It occurs in fishes and amphibians but not amniotes (16;17) however, it is developmentally comparable to post-embryonic organogenesis in mammals (18). While most amphibians experience metamorphosis, the AMA focuses on anuran metamorphosis because it has been well characterized, is most dramatic, and selected anuran species are well-suited for laboratory research.

Tata (19) described amphibian metamorphosis as a unique model for studying HPT axis function because it is dependent on the thyroid axis which orchestrates a multi-level, well-understood process resulting in physiological and biochemical changes, selective cell death, and anatomical restructuring in free-living larvae. It is possible that substances such as environmental chemicals, toxicants, natural products, and complex mixtures can alter metamorphosis by interacting with the HPT axis. Within the metamorphosis paradigm, disrupted thyroid processes through chemical interaction can result in measurable morphological and histopathological effects. The ability to measure such effects demonstrates the practical basis for the AMA.

2.2 Overview of amphibian metamorphosis and points of potential disruption

As previously indicated, the OECD (10) describes amphibian metamorphosis in depth, and provides an overview of the endocrinological mechanisms that drive amphibian metamorphosis. Peer reviewers are directed to those sources for additional information (12). Anuran (*Xenopus laevis*) metamorphosis is separated into three distinct periods, premetamorphosis, prometamorphosis, and metamorphic climax (20-22), each of which can be distinguished morphologically using Nieuwkoop and Faber's staging atlas (NF) (23). Premetamorphosis, which is TH-independent, refers to a period of embryonic and early larval development where some morphological changes are initiated including hind limb bud development, and corresponds to NF stages 48-53. Plasma concentrations of TH are very low in premetamorphic tadpoles because significant T4-secretory activity of the thyroid gland has not yet begun. However, premetamorphic tadpoles display a competence to respond to exogenous TH by upregulating TH-responsive gene expression leading to precocious induction of morphological changes (24). Prometamorphosis, characterized by rising concentrations of endogenous TH, is the period where specific morphogenesis, such as differentiation of the toes and rapid growth (elongation) of the hind limbs, as well as initiation of transformation of internal organs, occurs. NF stages 54-59 morphologically represent prometamorphosis. The final period, metamorphic climax, is characterized by a surge of circulating TH that stimulates forelimb development and resorption of the tail, as well as drastic internal transformations at the

organ system, tissue, and biochemical levels. NF stages 60-66 morphologically correspond to metamorphic climax. Please refer to **Figure 2-1** for a generalized scheme depicting the association between thyroid hormone changes and metamorphic stages (adapted from Shi *et al* and LeLoup *et al* (25;26).

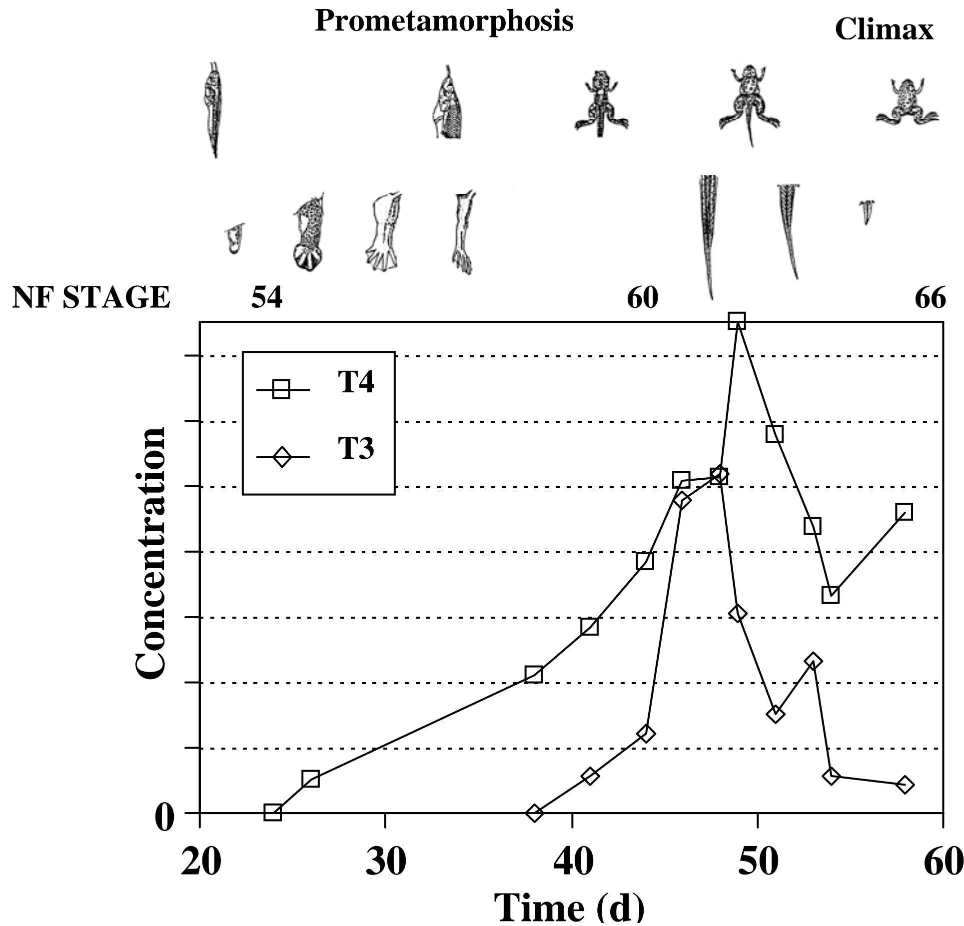


Figure 2-1. The association between thyroid hormone changes and metamorphosis.

Because an intact and appropriately functioning thyroid axis is necessary to progress through metamorphosis, chemicals that alter the structure or function of its constituents have the potential to alter metamorphic processes. Endocrine active compounds could potentially affect the thyroid axis at multiple sites, including: the central nervous system (CNS) (including hypothalamus), the pituitary, the thyroid glands, during TH transport, during TH metabolism and elimination, and in peripheral tissues. Specific modes of actions of thyroid axis disruptors could potentially include perturbation of iodine uptake, TH synthesis (iodination), TH storage, TH release, TH transport, TH activation (deiodination), TH elimination, neuro-endocrine axis regulation, and thyroid receptor (TR)

expression and/or function. **Figure 2-2** depicts peripheral (non-neuroendocrine) mechanisms of thyroid activity interference that could affect progression through metamorphosis. A summary of potential modes/sites of direct endocrine activity on the thyroid axis, in relation to assay endpoints possibly useful in measuring thyroid disruption, is provided in the DRP (10).

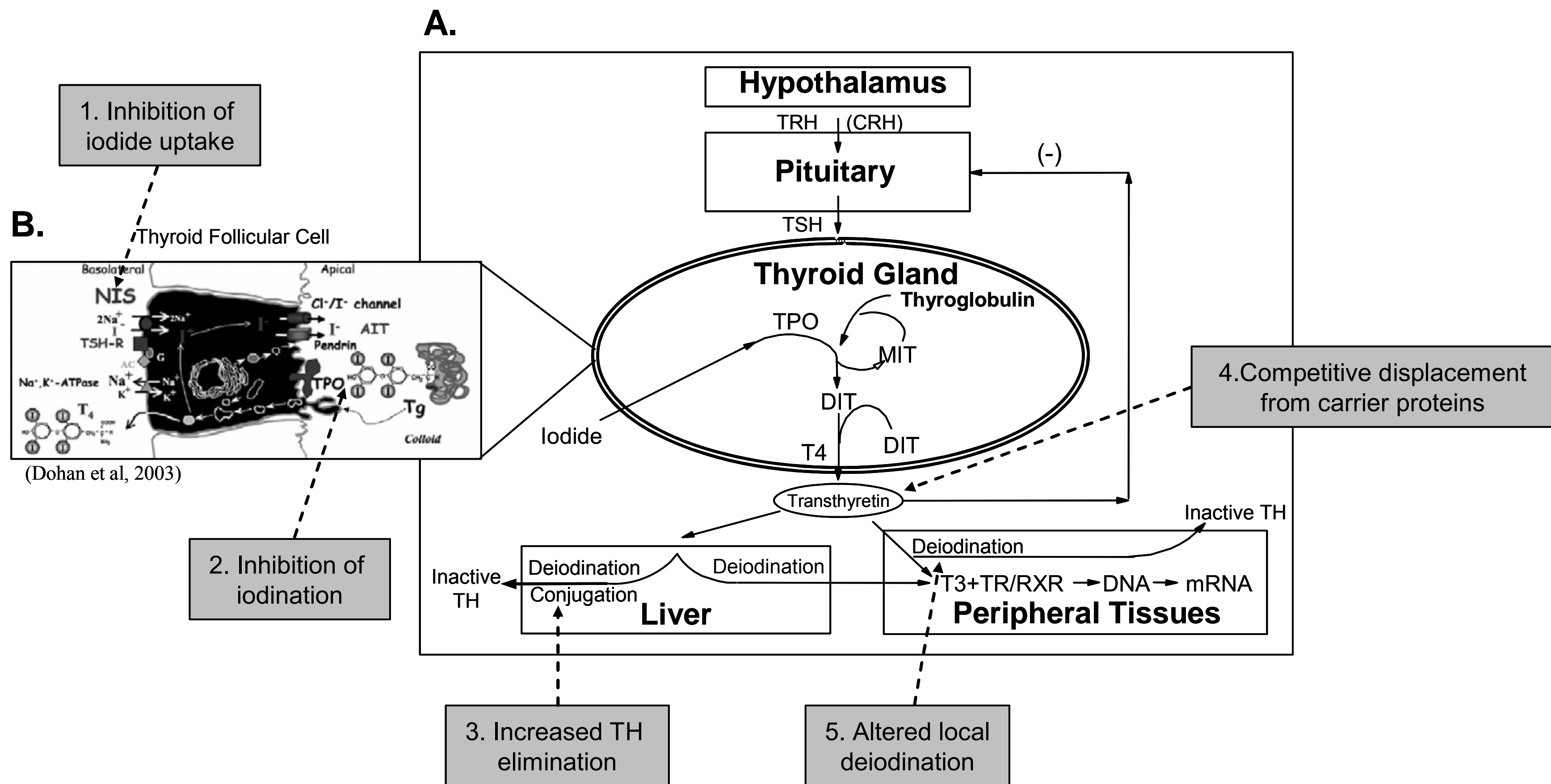


Figure 2-2. Possible mechanisms of thyroid axis activity. The generalized vertebrate HPT axis, including the liver and peripheral tissues (A) and the thyroid follicular cell (B) can be affected at several sites, including: 1. inhibition of iodide uptake at the sodium-iodide symporter (NIS); 2. inhibition of the iodinating activity of thyroid peroxidase (TPO); 3. increased elimination via upregulation of deoidination and conjugation reactions; 4. competitive displacement of TH from carrier proteins in the blood, such as transthyretin; and 5. altered local deiodination at the tissue level.

It should be noted that metamorphosis, and in some cases thyroid function, can be influenced by a combination of other biotic and abiotic factors beyond the realm of chemical stressors (15). These factors include temperature, water availability, crowding, light, diet, and environmental iodine levels (20). Amphibian larvae respond to changes in these factors through exhibiting plasticity in the timing of metamorphosis (27;28). Some factors that inhibit growth when present during premetamorphic stages are also capable of inducing rapid metamorphosis when present during prometamorphosis. These factors include crowding, resource limitation, habitat desiccation, and predation (27;28). Temperature also affects the rate of metamorphosis such that greater temperatures stimulate the rate of metamorphosis (29), whereas lower temperatures slow down metamorphosis (20). The effects of temperature may be due to reduction in TH binding at the tissue level, changes in neuroendocrine control of TH synthesis, or more generalized effects on metabolism (i.e., Q10 effects on enzyme kinetics) (20;30). Overall, it must be understood that the link between the thyroid axis and metamorphosis can be influenced by different external factors and/or combinations of external factors. This phenomenon is not unique to the thyroid system, and external factors are capable of affecting other developmental processes. However, since the AMA is a laboratory assay with strict controls of the various environmental factors mentioned above, these factors are unlikely to introduce significant experimental error.

3 PROTOCOL DEVELOPMENT

3.1 Rationale for the use of *X. laevis* in the AMA

Test Species

In general, *Xenopus laevis* was chosen as the appropriate test species for the AMA because it is routinely cultured in laboratories worldwide, it is easily obtainable through commercial suppliers, and there is a wealth of scientific information on the physiology of metamorphosis in this species. Reproduction is easily induced in this species throughout the year using human chorionic gonadotropin (hCG) injections. The resultant larvae can be routinely reared to selected developmental stages, in large numbers, to permit the use of stage-specific test protocols. The developmental rate for *X. laevis* is relatively rapid compared to the two ranid species commonly used in biological and toxicological research, *R. pipiens* and *R. catesbeiana*, minimizing test duration and costs. In terms of thyroid dependent post-embryonic development, more is known about *X. laevis* than any other anuran species. Genetic information regarding *X. laevis* is more extensive than with other anurans, especially in the area of the thyroid axis, where numerous publications have detailed the genes and the genetic program involved in the process of metamorphosis. Finally, the information on the biochemical and metabolic control of TH in this species is well-developed, and includes information on all of the typical HPT modulators as well as peripheral tissue enzymes, such as the deiodinases, which ultimately control the local and downstream effects of TH.

3.2 Selection of the developmental period and duration of exposure for the AMA

In order to maximize assay sensitivity and the ability for the AMA to detect thyroid axis activity, several considerations were evaluated to determine the appropriate metamorphic stages and duration of exposure during which the assay is performed. Two primary developmental periods were considered in the exposure regime - pre/prometamorphosis (NF stages 51 to 54) and metamorphic climax (stages 61 to 66). The EDSTAC originally recommended the use of a frog tail resorption protocol – a type of metamorphic climax assay. There was some question whether this type of assay would be sufficiently sensitive to thyroid-active compounds because effects on metamorphic climax proper (NF stages 61-66) seemed unlikely since metamorphic climax occurs too rapidly to be modulated by endocrine active substances. Additionally, high levels of circulating endogenous TH could also reduce the sensitivity of the model system to respond to TH synthesis disrupters during this metamorphic period. Therefore, an EPA Office of Research and Development (ORD) laboratory (MED, Duluth, MN) evaluated the protocol, both empirically and conceptually. The ORD laboratory concluded that the tail resorption assay was not sufficiently sensitive to be used as an assay for thyroid disruption activity. The laboratory then suggested that EPA examine the earlier *prometamorphosis* stage, when thyroid hormone synthesis is initiated and the larvae are responsive to lower circulating thyroid hormone concentrations. The ORD laboratory developed a 14-day version of the prometamorphosis protocol with *X. laevis* and recommended NF developmental stage 54 as the stage of initiation. Other protocols were also considered including a full metamorphosis protocol, where embryos are exposed through metamorphic climax; a 21-day version of the prometamorphosis protocol, in which the assay is initiated at an earlier stage of development and exposure occurs over a longer time period; assays using other species; and assays based on contexts outside of an EDSP screening battery (see OECD, 2004 (10) for a full discussion).

The EPA proceeded with validating the 14-day prometamorphosis protocol, and the U.S. presented this protocol to the OECD. However, one of the OECD member states (Germany) recommended a 28-day version beginning at NF stage 51. The U.S. felt the 14-day NF stage 54 version was more practical within the context of an EDSP screening battery. The German and U.S. protocols not only differed in duration and the initial stage of metamorphosis, but also in the preferred exposure method. The Germans used a static renewal method, while the U.S. preferred a flow-through exposure method. An international workshop on amphibian methods preceded an OECD expert group meeting where the two AMA protocols were discussed. It was determined that both should be further evaluated. The Germans agreed to shorten their protocol to 21 days but held that the assay should still be initiated at NF stage 51 of metamorphosis. A Phase 1 validation trial was undertaken using a thyroid agonist (T4) and a thyroid antagonist (6-PTU) as test chemicals (31). One cohort was evaluated beginning at NF stage 51 for 21 days, and another cohort was evaluated at NF stage 54 for 14 days for each chemical. Three labs (one in Germany, one in Japan, and one in the U.S.) tested both protocols and found that, while the results for the two compounds and the two protocols differed slightly, they were generally comparable (further discussion of these results is within chapter 4). Based on the results from these studies, it was determined that the protocol to be validated would begin at stage 51 and continue for 21 days. Section 3.3 describes the

rationale of aspects of this protocol in more depth, and the detailed Method is attached (attachment A).

In addition, two workshops were hosted by U.S. EPA to further the diagnostic capabilities and reproducibility of the histology endpoint. The workshops were held in Washington DC (USEPA) in July 2005 and January 2006. The primary objectives of the initial workshop (EPA Thyroid Histopathology Workshop) were to discuss the various issues associated with interpreting rat and amphibian thyroid gland histopathology and ultimately to develop standardized protocols. The objectives of the second workshop (OECD Expert Consultation on Amphibian Thyroid Histopathology) were similar to the first, but focused solely on thyroid histopathology as it related to the AMA. Guidance documents were developed that describe standardized histopathology reading practices, diagnostic criteria, severity grading, and data reporting. These documents serve to maximize comparability between pathologists and to reduce bias for this endpoint. They also provide a reference atlas of normal microanatomy of *X. laevis* thyroid glands and an atlas of core diagnostic criteria with examples of severity grades (32;33).

3.3 Rationale for the Core Endpoints in the AMA

The primary endpoints used to assess effects of test substances on metamorphic development are: developmental stage, hind limb length, body weight and length, and thyroid gland histology.

Developmental Stage

The developmental stage of *X. laevis* tadpoles is easily determined using the staging criteria of Nieuwkoop and Faber (23). The metamorphosis assay is designed to detect both agonistic and antagonistic effects on the thyroid system. Anti-thyroidal compounds are expected to retard metamorphic development, resulting in delayed development compared to untreated controls. Thyroid hormone agonists are expected to cause accelerated development resulting in more advanced developmental stages compared to untreated controls.

Hind Limb Length

Differentiation and growth of the hind limbs are under control of thyroid hormones and are major developmental landmarks for determining developmental stage. Thyroid hormones exert stimulatory effects on hind limb morphogenesis, consistent with metamorphic development of other tissues. Hind limb development is used qualitatively in the determination of developmental stage, but is considered here as a quantitative endpoint to detect effects on the thyroid axis. Hind limb length may be particularly useful as a diagnostic indicator of agonistic effects when evaluated in the premetamorphic period (NF 51-54), because this developmental period precedes normal onset of T4 synthesis.

Body Length and Wet Weight

Determinations of body length (snout-vent length and whole body length) and wet weight are included in the test protocol to assess possible effects of test substances on the growth rate of tadpoles in comparison to the control group. Although measurements of tadpole wet weight may provide a more precise estimation of tadpole growth, weight determinations during the course of the experiments are not feasible due to the necessary handling of the larvae. Therefore, wet weight measurements are only performed on a sub-sample of larvae (day 7) and on all remaining larvae at test termination (day 21).

There are two different approaches to assess tadpole growth by means of body length measurements: measurement of snout-vent length (SVL) or whole body length (WBL); see (33) for more information.

While growth effects are generally considered as toxic effects, disruption of HPT axis function can also affect tadpole growth. For example, exposure to T4 hastens development which results in reduced organism weights through two mechanisms. First, the normal period of growth that occurs in pre- and pro-metamorphic larvae is truncated upon exposure to T4, so the ultimate weight of the larvae is typically reduced. Second, there is a normal loss of weight throughout metamorphic climax in control animals and this weight loss is accelerated commensurate with the advanced development associated with exogenous T4 exposure. Exposure to TH synthesis inhibitors, on the other hand, prevents the larvae from experiencing the normal weight loss associated with climax and they continue to grow as prometamorphic animals to extraordinary weights. Consequently, any analysis of growth by either length or weight requires some interpretation. For example, reduced growth rates in the absence of developmental and histological effects provide an indication of general toxic effects of a test compound.

Thyroid Gland Histology

While developmental stage and hind limb length are important endpoints to evaluate exposure related changes in metamorphic development, developmental delay cannot, by itself, be considered a diagnostic indicator of anti-thyroidal activity. This is based on the likelihood that chemicals exist that may retard development through other modes of toxicity not related to perturbation of thyroid function. The diagnostic utility of thyroid histology is based on the idea that inhibition of normal thyroid function, which results in reduced circulating TH levels, causes a compensatory increase in TSH release by the pituitary. The thyroid glands respond to TSH stimulation by increasing synthetic activity, which is manifest at the histological level as follicular cell hypertrophy and hyperplasia, and eventually thyroid gland hypertrophy. Therefore, histological analysis of the thyroid glands is a necessary endpoint that improves the diagnostic capability of the assay.

In addition to the compensatory responses of the thyroid glands to TSH, differential histopathological effects can also be observed which may be related to the mechanism of toxicity of the test chemical. Exposure to perchlorate, for example, an inhibitor of thyroidal iodide uptake, results in different histopathological effects than PTU, an

inhibitor of thyroid peroxidase activity. These observations add to the diagnostic contribution of this endpoint.

Histological responses in the thyroid glands to inhibitors of the HPT axis are consistently observed at test chemical concentrations below those that elicit developmental delay. That is, the histological status of the thyroid glands is the most sensitive indicator of thyroid inhibition in the AMA. The simultaneous retention of normal metamorphic development of the organism and manifestation of obvious histopathological effects in the thyroid glands suggests that some level of TH synthesis inhibition can be overcome by the compensatory mechanisms of the HPT axis. While this may not be a major consideration when working with model test chemicals with known modes of thyroid toxicity, it is important when assessing the effects of chemicals during the screening process, where chemicals with lower potency are likely to be encountered.

3.4 Rationale for Exposure Scenarios in the AMA

Number of Animals on Test and Replicates

Tadpoles are equally allocated between treatment groups and control groups, with 20 animals per replicate aquaria. The practical limit of treatment replication of most flow-through aquatic toxicity testing systems is four. Prior to initiation of the Phase 2 inter-laboratory trial, power analyses were performed which demonstrated that the Jonckheere-Terpstra and Williams' tests, each applied in step-down manners, would provide sensitive and viable statistical procedures for tests with 4 replicate treatments (see attachment G). Additionally, power analyses were performed to determine the minimum number of sub-sampled organisms, both on day 7 and day 21, that would allow adequate statistical power for data interpretation of specific endpoints. Assuming equal allocation of four replicates in a five treatment study (control + 4 chemical treatments), the power analyses suggested that a sample size of five animals per replicate tank for the day 7 endpoint measurements does not impair the statistical power of the study at termination. For day 21 measurements, power analyses determined that 15 animals per replicate tank were sufficient to evaluate developmental stage. Accordingly, the minimum number of tadpoles per replicate tank at test initiation is 20 animals. For this assay, it was ultimately determined that three (3) treatment levels with a control would provide adequate statistical resolution for its screening purposes while minimizing the use of animals and spanning a sufficiently wide range of test concentrations.

Exposure System

Although a specifically designed test system is not required to conduct this assay, certain methods should be used. The preferred exposure system uses the flow-through method, meaning that exposure water is added continuously or semi-continuously to the treatment tanks and flows out to waste. The purpose of the flow-through system is to maintain test chemical concentration while eliminating the accumulation of degradates and waste products. Flow-through systems also allow for improved ability to maintain good water

quality including maintenance of appropriate temperature levels, dissolved gasses, pH, and ammonia.

Chemical delivery

It is necessary to obtain reliable information concerning the physicochemical properties of test chemicals to determine likely behavior in the test system and appropriate exposure methods. Factors important in this regard include water solubility, octanol-water partition coefficient (K_{ow}), melting point, density, volatility, stability in water, and biodegradability. Knowledge of these physicochemical properties will determine the appropriate procedure to use in producing a stock solution of the test chemical, as well as the necessary rate of renewal (i.e., derived from aqueous stability). For example, for chemicals that are highly soluble in water (≥1,000 mg/L), stock solutions may be prepared directly in the test water with a slow-stirring method. High-energy stirring, such as with a mechanical blender, generally should be avoided, as it may produce emulsified droplets of pure compound that could enter the exposure system. The stock solution may then be pumped directly into a holding chamber in the test system for subsequent dilution. If the chemical is relatively insoluble in water, and a liquid at the target test temperature (i.e., melting points <25°C), a liquid-liquid saturator is recommended for generating stock solutions; whereas, if the chemical is a solid, a glass wool column saturator is the preferred method of stock solution generation (34).

None of the above methods require solvent or oil carriers, however different test chemicals will possess varied physicochemical properties that will likely require different approaches for preparation of chemical exposure water. It is preferred that effort be made to avoid solvents or carriers because: (1) certain solvents themselves may result in toxicity and/or undesirable or unexpected endocrinological responses, and (2) the use of solvents in longer-term tests can result in a significant degree of "biofilming" associated with microbial activity. For difficult to test substances, a solvent may be employed as a last resort, and the OECD Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (35) should be consulted to determine the best method. If solvent carriers are used, appropriate solvent controls must be evaluated in addition to non-solvent controls. If it is not possible to administer a chemical via the water, either because of physicochemical characteristics (low solubility) or limited chemical availability, it may be necessary to introduce the chemical via the diet. Preliminary work has been conducted on dietary exposures, however, this route of exposure is not commonly used.

3.5 Statistical Analyses

For the quantitative endpoints of the AMA, including hind limb length, body length (both whole body length and snout-vent length), and wet weight, the statistical tests recommended for use are consistent with the OECD Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (36). The recommended statistical protocol for normally distributed (parametric) data sets is to perform an ANOVA test followed by Dunnett or Williams' test if there is a linear dose response.

Pairwise tests are recommended, and the Dunnett test is most powerful and appropriate, provided the data are normally distributed with homogeneous variances. Where these conditions cannot be satisfied, a transformation should be sought that normalize the data and stabilize the variances. If the data are normally distributed but heterogeneous, and no variance stabilizing transformation can be performed, a robust version of Dunnett's test, referred to as the Tamhane-Dunnett test, requiring only normality but not variance homogeneity, can be used.

For non-parametric data, the recommended test is the Mann-Whitney U, or alternatively, the Jonckheere-Terpstra test if there is a linear dose response.

Special Issues for Statistics

Developmental stage is an inherently non-normal, even non-continuous response and parametric analyses such as the Williams' and Dunnett tests are not applicable, nor will any transformation alter that basic fact. Thus Dunnett and Williams' tests cannot be used in the analysis of this endpoint. Rather, Mann-Whitney U and Jonckheere-Terpstra (if there is a linear dose response) tests are recommended. The unit of analysis is the replicate, not individual animals. Since developmental stage is measured on an ordinal scale, not a ratio scale, analysis is based on the replicate *median* rather than the replicate mean.

Like the developmental stage endpoint, histology and histopathology, using severity grading score methods and qualitative assessments, also are not amenable to statistical analyses like those for more quantitative, non-categorical data. Statistical approaches for histopathology are still under development, however because experts in amphibian toxicologic pathology are used to read the studies, expert opinion weighs in on this endpoint.

3.6 Data Interpretation

The AMA is a screening assay designed to identify compounds that have activities within the HPT axis. Because further testing would be necessary to fully characterize the activity of a compound, this screening assay is not meant to determine dose responses, LOECs or NOECs. Rather, interpretation of data from this assay should lead to a determination if further testing is necessary or not. Given this, there are three potential outcomes of the AMA: potentially thyroid active, thyroid inactive, and toxic. Based on the following discussion, compounds that are deemed potentially thyroid active are subject to undergo further testing to characterize the activity. Compounds that are thyroid inactive will not likely undergo further testing to characterize *thyroid* activity. Compounds that are determined to be toxic, where no evidence suggests that the compound has the potential to be both toxic and thyroid active, will not undergo further *thyroid* testing. Compounds with equivocal results, results that indicate that the compound has both thyroid activity and generalized toxicity, or results that indicate that effects are due to non-HPT axis activity (e.g. HPG-axis activity), will be evaluated on a case-by-case basis, and in combination with other assays in the screening battery, to determine the need for further testing.

Potential Endpoint Outcomes:

Advanced development (determined using developmental stage, HHL, BW, SVL, WBL)

Advanced development is only known to occur through effects which are thyroid hormone related. These may be peripheral tissue effects such as direct interaction with the thyroid hormone receptor (such as with T4) or effects which alter circulating thyroid hormone levels. In either case, this is considered sufficient evidence to indicate that the chemical has thyroid activity. Advanced development can be evaluated in one of two ways. First, the general developmental stage can be evaluated using the standardized approach detailed in Nieuwkoop and Faber (23). Second, specific morphological features may be quantified, such as hind limb length, which is positively associated with agonistic effects on the thyroid hormone receptor. If statistically significant advances in development occur, then the test indicates that the chemical is thyroid active.

Delayed development (determined using developmental stage, HHL, BW, SVL, WBL)

Delayed development can occur through anti-thyroidal mechanisms and through indirect toxicity. Mild developmental delays coupled with overt signs of toxicity likely indicate a non-specific toxic effect. Evaluation of non-thyroidal toxicity is an essential element of the test to reduce the probability of false positive outcomes. Excessive mortality is an obvious indication that other toxic mechanisms are occurring. Similarly, mild reductions in growth, as determined by wet weight and/or organism length, also indicate non-thyroidal toxicity. Apparent increases in growth are commonly observed with compounds that negatively affect normal development. Consequently, the presence of larger animals does not indicate non-thyroidal toxicity. However, growth should never be solely relied upon to determine thyroid toxicity. Rather, growth, in conjunction with developmental stage and thyroid histopathology, should be used to determine thyroid activity. Other endpoints should also be considered in determining overt toxicity including edema, hemorrhagic lesions, lethargy, reduced food consumption, erratic/altered swimming behavior, etc. If all test concentrations exhibit signs of overt toxicity, the test compound must be re-evaluated at lower test concentrations before determining whether the compound is potentially thyroid active or thyroid inactive.

Statistically significant developmental delays, in absence of other signs of overt toxicity, indicate that the chemical is thyroid active (antagonistic). In the absence of strong statistical responses, this outcome may be augmented with results from thyroid histopathology.

The inability to clearly establish the developmental stage of an organism using the suite of morphological endpoints considered typical of any given stage indicates that the tissues are developing asynchronously through metamorphosis. Asynchronous development is an indicator of thyroid activity. The only known mode of action of this process is through peripheral actions of the chemical on developing tissues such as is observed with deiodinase inhibitors.

Histopathology

If the chemical does not cause overt toxicity and does not accelerate development, then histopathology of the thyroid glands should be evaluated using the appropriate guidance documents. This analysis needs to be conducted whether or not the chemical retards development. Developmental retardation, in the absence of toxicity, is a strong indicator of anti-thyroid activity, but the developmental stage analysis is less sensitive and less diagnostic than the histopathological analysis of the thyroid gland. Therefore, conducting histopathological analyses of the thyroid glands is required. Effects on thyroid gland histology have been demonstrated in the absence of developmental effects. If changes in thyroid histopathology occur, then the chemical is considered to be thyroid active. If no developmental delays or histological lesions are observed in the thyroid glands, then the chemical is considered to be thyroid inactive. The rationale for this decision is that the thyroid gland is under the influence of TSH and any chemical which alters circulating thyroid hormone sufficiently to alter TSH secretion will result in histopathological changes in the thyroid glands. Various modes and mechanisms of action can alter circulating thyroid hormone. So, while this endpoint is indicative of a thyroid related effect, it is insufficient to determine which mode or mechanism of action is related to the response.

Because this endpoint is not amenable to basic statistical approaches, the determination of an effect associated with exposure to a chemical shall be made through expert opinion by a pathologist.

4 OPTIMIZATION OF THE TEST METHOD

4.1 OECD Phase 1 study

The primary objective of the Phase 1 OECD study was to determine the most practical and sensitive exposure scenario for the AMA for detecting changes in metamorphic development and thyroid system function in response to substances considered to act as potent agonists (L-thyroxine [T4]) and potent antagonists (6-propyl-2-thiouracil [PTU]). The two testing scenarios evaluated were 1) exposure of stage 51 tadpoles for a total of 21 days and 2) exposure of stage 54 tadpoles for a total of 14 days.

Additional goals of the Phase 1 pre-validation work were to: 1) evaluate the intra- and inter-laboratory variability of developmental and growth rates of control animals; 2) evaluate the reproducibility, and ultimately the robustness of the assay when slightly different experimental conditions are used in three different labs, 3) obtain data on the intra- and inter-laboratory variability and reproducibility among the selected core endpoints of the assay, 4) compare the different endpoints with view to their relevance, sensitivity and diagnostic value, 5) identify need for protocol changes to enhance reproducibility, sensitivity, and diagnostic value of the assay, and 6) provide a proposal for a testing protocol to be used in Phase 2 of the validation process – the inter-laboratory evaluation.

The Phase 1 pre-validation studies were conducted in three separate laboratories located in Germany, Japan, and the U.S.

Overview of Methods

General standard operating procedures (SOPs) for the conduct of the in-life portion of the study were developed and approved by the Validation Management Group for Ecotoxicity Testing (VMG-eco) for use in the Phase 1 pre-validation trial. Details of these procedures can be accessed in Annex 5 of the Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 – Optimisation of the Test Protocol (31) (attachment B). In general, the study was conducted according to the proposed method (attachment A), using two different exposure scenarios, with the following differences:

- **Exposure method** – the U.S. and Japanese laboratories used flow-through exposure systems, whereas the German laboratory used a static-renewal system.
- **Dilution water –** the Japanese laboratory used activated carbon filtered, UV-irradiated tap water, the U.S. laboratory used filtered, UV- sterilized Lake Superior water, and the German laboratory used a synthetic test medium.
- **Diet –** Sera Micron was used in the German and Japanese laboratories and the U.S. laboratory used a mixture of trout starter, algae, TetraFin™ flakes, and live brine shrimp.
- **Test animals per replicate –** 20 tadpoles/replicate aquaria were used in the German and Japanese laboratories, whereas the U.S. laboratory study with PTU was conducted using 25 tadpoles/replicate. Test vessel size and tank dimensions were not reported.
- **Developmental stage** – German and Japanese laboratories recorded stage data on days 0, 7, 14 and 21 of exposure, whereas the U.S. lab measured these endpoints only at test termination (day 21).
- **Body length measurement –** the German and Japanese laboratories determined whole body length (WBL) from the snout to the tip of the tail; the U.S. lab determined this endpoint as snout-to-vent length (SVL).
- **Thyroid gland histology –** transverse sections of the lower jaw from dorsal to ventral were prepared and reviewed by the German laboratory, whereas the Japanese laboratory evaluated sagittal sections of the whole body from left to right. The U.S. laboratory prepared transverse sections of the head from caudal to rostral.

Each lab conducted four exposure studies (2 chemicals tested at two different exposure durations) as described in **Table 4-1**.

Table 4-1. Overview of exposure studies conducted by German, Japanese and U.S. laboratories.

Test Material	Concentrations* (µg/L)	Stage at Test Initiation	Duration
PTU	2,500; 5,000; 10,000; 20,000	51	21 days
PTU	2,500; 5,000; 10,000; 20,000	54	14 days
T4	0.25, 0.5, 1.0, 2.0	51	21 days
T4	0.25, 0.5, 1.0, 2.0	54	14 days

* In addition to the concentrations listed above, the U.S. laboratory also tested a PTU concentration of 1,250 mg/L and a T4 concentration of 4.0 µg/L. Test concentrations were selected based previous studies by the participating laboratories.

Data Analysis

Prior to formal statistical analysis, data sets were tested for normality and homogeneity of variance. Data sets for body length, hind limb length and body weight measurements were analyzed using ANOVA and a post-hoc multiple comparison test (Dunnett test). Developmental stage data were analyzed using ANOVA on ranks (Kruskal-Wallis ANOVA) and Dunn's multiple comparison test. Differences in growth and developmental indices between treatments were considered significant at $p<0.05$. Gene expression data (German laboratory only) were log-transformed to satisfy the criteria of normality and homogeneity of variance. A Dunnett test was used to compare control data to all other treatment groups. Differences were considered significant at $p<0.05$.

Overview of Results – Control Organism Performance

A total of 12 exposure experiments with *X. laevis* tadpoles were conducted during the pre-validation study of the AMA. Despite differences in testing conditions between the three laboratories, the overall performance of control organisms, as judged from mortality, developmental, and growth rates was similar (see **Table 4-2**).

Table 4-2. Test termination summary data for control performance for the OECD Phase 1 pre-validation study.

Expsosure scenario	Chemical	Lab	Mortality (number found dead)	Median developmental stage at test termination (range)	Mean body length (mm)	Mean body weight (mg)
Stage 51/21 days	PTU	US	0	59 (7)	19.5 ± 1.9 (SVL)	1047 ± 230
		Germany	0	58 (6)	59.8 ± 2.6 (WBL)	919 ± 162
		Japan	0	58 (5)	64.2 ± 3.9 (WBL)	1065 ± 153
	T4	US	0	59 (6)	19.5 ± 1.7 (SVL)	1042 ± 202
		Germany	0	60 (5)	58.9 ± 4.2 (WBL)	824 ± 169
		Japan	0	58 (6)	56.1 ± 3.4 (WBL)	761 ± 127
Stage 54/14 days	PTU	US	0	58 (4)	19.9 ± 1.2 (SVL)	1069 ± 161
		Germany	0	58 (5)	59.7 ± 2.9 (WBL)	957 ± 139
		Japan	0	57 (3)	60.4 ± 2.9 (WBL)	896 ± 113
	T4	US	0	58 (5)	19.2 ± 0.9 (SVL)	943 ± 157
		Germany	0	58 (4)	59.6 ± 2.4 (WBL)	861 ± 117
		Japan	0	57 (4)	59.6 ± 3.2 (WBL)	882 ± 131

Mortality was not observed in any of the control groups. Slightly lower rates of development based on developmental stage at test termination were observed in the control groups during several experiments conducted in the Japanese laboratory. However, the Japanese laboratory encountered some technical difficulties with a new flow-through diluter system which caused reduced food availability, potentially accounting for some of the differences in developmental rates. This conclusion is based on unpublished data from the U.S. lab where the two different diets were evaluated using the related species, *Silurana* (*X.) tropicalis*. In that study, different concentrations of Sera Micron diet were administered to larvae under static and differing flow-through conditions. It was found that development under static conditions could be greater than flow-through when the same amount of food was provided. This difference was overcome in flow-through conditions with higher Sera Micron feeding rates. Despite different exposure scenarios between the U.S. and German labs, development was similar.

Growth of tadpoles was assessed by means of body length (WBL/SVL) measurements throughout the exposure periods of 21 and 14 days, respectively, and at test termination (represented in **Table 4-2**). Germany and Japan measured WBL whereas the U.S. measured snout vent length (SVL). The overall mean of WBL for the 21-day study was 59.7 ± 2.9 mm with a coefficient of variation (CV) of 4.8%. For the 14-day study, the overall mean of WBL was 59.8 ± 0.3 mm with a 0.6% CV. Overall mean body weights (mg) for the 21-day study were 943 ± 118 with a 12.5% CV. For the 14-day study, overall mean weights were 934 ± 69 with a CV of 7.4%.

Slight differences in performance of control organisms may not represent a concern for the robustness of the assay or preclude a determination of effect in treated groups. If the differences are the result of culture conditions, such as temperature, differential responses to the test chemicals might also be observed. Developmental and growth rates between laboratories using static-renewal systems versus flow-through systems were consistent,

indicating that *X. laevis* development is relatively insensitive to different exposure systems.

Overview of Results – PTU

Summary tables (**Table 4-3** and **Table 4-4**) of results, broken out by exposure scenario, are provided below.

Each of the laboratories was able to identify an inhibitory effect of PTU on metamorphic development in *X. laevis* tadpoles based on comparison of developmental stages from treated organisms to control organisms at test termination, with both exposure scenarios (14 day and 21 day) at 20 mg/L. The German and Japanese laboratories also evaluated developmental stages at 7 days to determine if effects could be detected earlier in the exposure. Neither laboratory, regardless of exposure scenario, detected significant changes in developmental stages at 7 days, however at 14 days, significant effects on developmental stage occurred in the Japanese laboratory at both 20 mg/L and 5 mg/L. The significant difference at 5 mg/L after 14 days of exposure in the Japanese study seems to be an anomalous result and driven by one of the two replicates which does not fit the pattern of the other tests. Furthermore, the apparent significance at 5 mg/L for 14 days by the Japanese lab does not persist at 21 days, suggesting that this observation is not real. Mean hind limb length proved to be a reliable indicator of effect in the 21-day study, whereas mean body length was a useful indicator of effect in only the 14-day study. Significant differences in mean body weight were only detected in the 21-day study.

Based on morphological endpoints, greater sensitivity was achieved using the 21-day exposure protocol (LOECs = 10 mg/L PTU for the 21-day study and 20 mg/L PTU for the 14-day study) in the Japanese laboratory, and a more consistent profile of effect occurred in the 21-day study for the German laboratory. In contrast, the U.S. laboratory found the NF stage 54 exposure protocol to be more sensitive than the NF stage 51 approach (LOECs = 20 mg/L PTU for 21-day study and 10 mg/L PTU for 14-day study). Thus, it was impossible to determine which exposure scenario represented the more sensitive approach. This same experimental question is addressed in the USEPA pre-validation study discussed in the following subsection. Overall, the LOEC values determined for developmental stage did not vary by more than one test concentration between the laboratories and between the different exposure protocols within each laboratory.

Hind limb length, used in the Japanese and German laboratories, confirmed the inhibition of metamorphic development by PTU. Data from both laboratories suggested sensitivity differences of this endpoint between the two exposure protocols. In contrast to the developmental stage endpoint, the 21-day exposure protocol allowed for inhibited limb development at 7 days and at test conclusion in the 20 mg/L PTU treatment. In the German laboratory, 10 and 20 mg/L PTU caused a significant inhibition of hind limb growth at day 14 and day 21, whereas developmental stage was significantly affected only at test termination in the 21-day exposure studies. Thus, developmental delay

caused by 10 and 20 mg/L PTU was more readily detected by hind limb length than developmental stage in the 21-day exposure studies.

The German laboratory found no significant effect on hind limb growth in the 14-day exposure study. Significant delay in hind limb growth was only detected at day 14 at 20 mg/L PTU in the NF stage 54 exposure study in the Japanese laboratory. Overall, these data suggest that hind limb length measurements in a stage 51/21-day exposure protocol could provide a valuable endpoint to more rapidly detect developmental delay caused by anti-thyroidal substances.

Measurements of whole body length at early time points during the exposure phase indicated the presence of weak growth-retarding effects at the highest PTU test concentration (20 mg/L). However, further studies will be necessary to investigate the relationship between tadpole growth and disruption of the thyroid axis in *X. laevis*.

Developmental stage determination and hind limb length measurements represent apical endpoints that can be used to evaluate exposure-related changes in metamorphic development. Both endpoints may or may not be diagnostic of thyroid system-related modes of action. Thyroid gland histopathology was proposed as a core endpoint for the AMA to enhance the assay's specificity for thyroid system-related mechanisms of action. Histopathological analysis of thyroid gland sections revealed a classical goiter; e.g., distension of thyroid follicles, diffuse enlargement of the thyroid gland, depletion of colloid and follicular cell hyperplasia. These findings strongly suggest that PTU acts as an anti-thyroidal agent in *X. laevis* tadpoles.

Table 4-3. Summary data for the 21-day, NF stage 51 exposure scenario using PTU in the OECD Phase 1 pre-validation study.

Expsosure scenario	Lab	Dose (nominal) mg/L	Mean dose (actual) mg/L ± SD	Median developmental stage (range)			Mean hind limb length (mm) ± SD			Mean body length (mm) ± SD GER/JP - WBL / US - SVL				Mean body weight (mg) ± SD
Day				7	14	21	7	14	21	0	7	14	21	21
Stage 51/21 days	US	0.0	0.0	ND	ND	59 (57-63	ND	ND	ND	ND	ND	ND	19.5 ± 1.9	1047.0 ± 230.2
	US	2.5	2.79±0.04	ND	ND	59 (57-63	ND	ND	ND	ND	ND	ND	19.4 ± 1.5	1003.7 ± 164.5
	US	5	5.55±0.05	ND	ND	59 (57-62	ND	ND	ND	ND	ND	ND	20.0 ± 1.9	1081.5 ± 215.8
	US	10	10.81±0.1	ND	ND	59 (56-62	ND	ND	ND	ND	ND	ND	19.7 ± 1.6	1075.5 ± 183.3
	US	20	21.20±0.19	ND	ND	55 (53-59	ND	ND	ND	ND	ND	ND	21.8 ± 1.3	1211.1 ± 204.6
	Germany	0.0	NR	54 (53-55)	56 (55-57)	58 (56-61	2.3 ± 0.3	5.1 ± 0.9	11.2 ± 3.3	28.1 ± 1.0	45.6 ± 2.5	56.3 ± 3.3	59.8 ± 2.6	918.8 ± 161.9
	Germany	2.5	NR	54 (53-54)	56 (55-57)	57 (56-61	2.1 ± 0.2	4.7 ± 0.9	9.7 ± 3.0	28.1 ± 1.0	44.2 ± 2.6	54.7 ± 3.6	58.5 ± 3.0	864.9 ± 170.1
	Germany	5	NR	54(53-55)	56 (55-57)	57 (57-59	2.2 ± 0.3	4.8 ± 0.7	9.8 ± 2.0	28.0 ± 1.0	44.6 ± 2.1	55.1 ± 2.6	58.8 ± 2.4	859.9 ± 125.5
	Germany	10	NR	54 (53-55)	56 (55-56)	57 (56-59	2.1 ± 0.2	4.3 ± 0.7	8.9 ± 2.3	27.6 ± 0.8	43.2 ± 2.2	53.9 ± 3.2	58.3 ± 2.8	844.7 ± 152.2
	Germany	20	NR	54 (53-55)	56 (54-56)	57 (55-59	2.0 ± 0.2	4.2 ± 0.8	8.4 ± 2.6	27.9 ± 1.0	43.9 ± 2.6	54.5 ± 3.1	58.7 ± 3.5	841.6 ± 175.0
	Japan	0.0	0.0	54 (53-55)	58 (57-61)	58 (57-61	2.2 ± 0.4	5.7 ± 1.1	12.8 ± 3.0	26.5 ± 1.7	42.6 ± 3.1	57.3 ± 4.4	64.2 ± 3.9	1065 ± 152.7
	Japan	2.5	2.03±0.78	54 (52-54)	58 (56-62)	58 (56-62	2.3 ± 0.4	5.9 ± 1.3	13.7 ± 3.7	26.2 ± 1.7	44.6 ± 3.3	57.9 ± 3.5	64.3 ± 3.0	1023 ± 139.3
	Japan	5	4.35±2.17	54 (52-55)	57 (55-61)	57 (55-61	2.1 ± 0.4	4.8 ± 1.2	10.8 ± 3.1	26.3 ± 1.9	41.2 ± 3.9	53.3 ± 5.1	62.3 ± 4.4	864 ± 169.6
	Japan	10	11.20±1.91	54 (52-55)	58 (55-60)	58 (55-60	2.3 ± 0.4	5.6 ± 1.2	12.5 ± 3.7	26.6 ± 2.1	45.9 ± 3.9	57.2 ± 4.7	63.4 ± 4.3	967 ± 157.0
	Japan	20	27.62±5.98	54 (51-54)	54 (52-57)	55 (54-58	1.9 ± 0.4	3.1 ± 1.1	5.1 ± 2.7	26.6 ± 1.5	39.9 ± 5.2	53.5 ± 8.7	63.6 ± 6.7	914 ± 227.2

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$. ND = Not done. NR = Not reported.

Table 4-4. Summary data for the 14-day, NF stage 54 exposure scenario using PTU in the OECD Phase 1 pre-validation study.

Expsosure scenario	Lab	Dose (nominal) mg/L	Mean dose (actual) mg/L ± SD	Median developmental stage (range)		Mean hind limb length (mm) ± SD		Mean body length (mm) ± SD GER/JP - WBL / US - SVL			Mean body weight (mg) ± SD
Day				**7**	**14**	**7**	**14**	**0**	**7**	**14**	**14**
Stage 54/14 days	US	0.0	0.0	ND	58 (57-60)	ND	ND	ND	ND	19.9 ± 1.2	1068.8 ± 160.6
		2.5	2.86±0.03	ND	58 (57-59)	ND	ND	ND	ND	19.8 ± 1.3	1043.1 ± 150.1
		5	5.43±0.06	ND	58 (57-59)	ND	ND	ND	ND	20.0 ± 1.0	1058.6 ± 134.9
		10	10.71±0.11	ND	57 (57-59)	ND	ND	ND	ND	19.8 ± 1.4	1017.6 ± 186.7
		20	20.88±0.18	ND	56 (55-59)	ND	ND	ND	ND	21.1 ± 1.1	1139.2 ± 171.0
	Germany	0.0	NR	56 (55-57)	58 (56-60)	5.6 ± 0.9	10.9 ± 2.8	43.2 ± 1.4	57.1 ± 3.0	59.7 ± 2.8	956.8 ± 139.0
		2.5	NR	56 (56-57)	57 (55-60)	5.4 ± 0.8	9.8 ± 2.5	42.6 ± 1.6	56.1 ± 2.9	58.9 ± 2.9	924.2 ± 157.5
		5	NR	56 (55-57)	57 (56-59)	5.3 ± 0.9	9.8 ± 2.5	42.6 ± 1.6	55.5 ± 2.2	58.4 ± 2.2	923.1 ± 129.0
		10	NR	56 (56-57)	57 (57-60)	5.4 ± 0.8	9.9 ± 2.1	42.7 ± 1.5	55.7 ± 2.2	58.5 ± 2.3	916.4 ± 112.1
		20	NR	56 (55-57)	57 (56-59)	5.2 ± 0.8	9.3 ± 2.1	42.9 ± 1.4	55.3 ± 2.5	58.3 ± 2.5	878.6 ± 133.6
	Japan	0.0	0.0	55 (55-56)	57 (56-58)	4.2 ± 0.6	8.9 ± 1.6	39.0 ± 2.1	51.8 ± 2.4	60.4 ± 2.9	896.0 ± 113.4
		2.5	2.93±0.75	55 (54-56)	57 (56-59)	4.0 ± 0.8	8.8 ± 2.1	38.6 ± 3.2	50.4 ± 3.0	60.2 ± 2.7	881.0 ± 116.4
		5	5.15±0.28	55 (54-56)	57 (56-58)	3.9 ± 0.5	8.6 ± 1.6	38.5 ± 2.2	50.7 ± 2.7	59.1 ± 2.5	838.0 ± 117.2
		10	11.17±0.37	55 (54-56)	57 (56-59)	4.2 ± 0.7	8.9 ± 1.9	39.4 ± 2.3	51.9 ± 2.6	60.2 ± 2.2	852.6 ± 99.6
		20	24.19±0.42	55 (54-56)	57 (56-59)	3.9 ± 0.7	7.5 ± 1.7	38.4 ± 2.1	52.2 ± 2.7	60.9 ± 3.5	861.4 ± 129.3

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$.
ND = Not done. NR = Not reported.

Pronounced changes in thyroid gland histology were observed in *X. laevis* tadpoles exposed to PTU from either stage 51 or stage 54. While the light microscopic appearance of thyroid glands from the 2.5 mg/L PTU treatment group did not differ markedly from the control group, exposure-related changes in the thyroid glands of tadpoles treated with higher PTU concentrations included distension of thyroid follicles, diffuse enlargement of the thyroid glands, colloid depletion, follicular cell hypertrophy and hyperplasia. The incidence and severity of these changes increased in a concentration-dependent manner in each of the studies performed. At the highest exposure concentration of PTU (20 mg/L), tadpoles were markedly affected with thyroid follicular cell hypertrophy and hyperplasia accompanied by diffuse thyroid gland enlargement, irrespective of the stage at which exposure was initiated. Overall, the results from the qualitative histological evaluation of thyroid glands were relatively consistent between the laboratories which confirmed the anti-thyroidal activity of PTU in *X. laevis* tadpoles.

The Japanese and German laboratories also quantified changes in selected histological endpoints. By using image analysis techniques, measurements of epithelial cell heights, follicular lumen area and thyroid gland area confirmed the presence of follicular cell hypertrophy, diffuse enlargement of follicles and diffuse thyroid gland enlargement at high PTU concentrations. High variability in measured values for the selected endpoints prevented the detection of statistically significant differences at low PTU concentrations. In these studies, it was found that changes in epithelial cell height provided a less sensitive endpoint than follicular cell and thyroid gland size.

Using the data interpretation criteria from section 3.3, a summary of results for the PTU studies is presented in **Table 4-5**. All laboratories detected significant inhibition of development in both exposure scenarios. Inconsistencies occur in the measurement of wet weight and SVL/WBL. The U.S. measured SVL whereas Germany and Japan measured WBL. It is possible that rates of tail resorption confound these findings, accounting for the discrepancies. In the Japanese laboratory, significant changes in body weight were equivocal (increased and decreased depending on dose) and not dose dependent. The most consistent endpoint was thyroid histopathology. All laboratories, regardless of exposure scenario, detected significant changes in the thyroid glands, and therefore the assays were determined to be “positive”.

Table 4-5. Summary results for PTU from the OECD Phase 1 study.

Criterion		Stage 51/21 days US	Stage 51/21 days GR	Stage 51/21 days JP	Stage 54/14 days US	Stage 54/14 days GR	Stage 54/14 days JP
Overt Toxicity	Mortality	-	-	-	-	-	-
	Hemorrhagic lesions	-	-	-	-	-	-
	Edema	-	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-	-
	Lethargy	-	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-	-
Advanced Development		-	-	-	-	-	-
Delayed Development	Developmental stage (d7 or 14/21)	↓	↓	↓	↓	↓	↓
	Wet weight (d7 or 14/21)	↑	-	↓	-	-	-
	SVL/WBL (d7 or 14/21)	↑	↓	E	↑	↓*	-
	HHL (d7 or 14/21)	ND	↓	↓	ND	-	↓
	Thyroid histopathology	+	+	+	+	+	+
	Overall assay result:	+	+	+	+	+	+

ND = not done
E = Equivocal
*No concentration response relationship

Overview of Results – T4

Summary tables (**Table 4-6**, **Table 4-7**) of results, broken out by exposure scenario, are provided below.

In the T4 study, all laboratories detected the expected accelerating effects on metamorphic development in *X. laevis* tadpoles. Each of the laboratories was able to detect significant acceleration of tadpole development at the highest test concentration of T4 (2.0 µg/L), regardless of exposure scenario, but not at lower concentrations. Data sets from the German and Japanese laboratories showed that exposure of NF stage 51 and 54 tadpoles to 2.0 µg/L T4 for 7 days was sufficient to detect statistically significant developmental acceleration compared to the control group. Thus, exogenous addition of low concentrations of T4 may produce a biologically relevant increase in circulating T4 levels during this early developmental phase. The highest dose used in this study was 2.4 nM. The peak whole body content, or plasma concentration of T4, is ca. 7-10 nM in *X. laevis* (25;37). Since *X. laevis* tadpoles are capable of taking up and concentrating thyroid hormones 4-6 times above the environmental concentration (37), the low dose of T4 may actually result in plasma concentrations at or above those achieved at metamorphic climax.

Different results were obtained in the three laboratories with regard to the sensitivity of the two exposure protocols for detection of stimulatory effects of T4 on development. No sensitivity differences were observed in the U.S. lab (LOEC = 2.0 µg/L T4 for both exposure protocols). However, data sets from the German and Japanese laboratories showed a higher sensitivity of the NF stage 51 exposure protocol (LOECs = 1.0 µg/L T4 for NF stage 51 and 2.0 µg/L T4 for NF stage 54). Thus, based on developmental stage data from at least two laboratories, exposure initiated at NF stage 51 for 21 days was

more sensitive than exposure initiated at NF stage 54 for 14 days. Overall, the LOEC values determined for developmental stage did not vary by more than one test concentration between the laboratories, and between the different exposure protocols within each laboratory.

As with the PTU studies, hind limb length was used in the Japanese and German laboratories as an additional endpoint to detect exposure-related alterations in development. Data from both laboratories indicated a sensitivity difference of this endpoint between the two exposure protocols. In the 21-day experiments, hind limb length measurements at day 7 provided the most sensitive endpoint detecting the stimulatory effects of T4 (LOEC = 0.5 µg/L T4 in the German laboratory and 1.0 µg/L in the Japanese laboratory). In the 14-day experiments, hind limb length measurements were less straight forward due to a heterogenous effect in the Japanese laboratory. The German investigators noted that the sensitivity of hind limb length measurements to detect T4 effects diminished in the latter stage of the 21-day exposure study. With the exception of the Japanese laboratory, these data indicate that hind limb length measurements at day 7 in a 21-day/stage 51 exposure protocol can be a sensitive endpoint to detect agonist activities.

At exposure termination, reduced mean whole body length, snout-to-vent length, and body weight were consistently observed at the greater T4 concentrations in all three laboratories. Overall, the results suggested that the T4 concentrations tested accelerated development leading to decreased mean values of whole body length, snout-to-vent length and body weight at test termination. As with the PTU studies, evaluation of body size-related endpoints at early time points during the exposure period could potentially provide relevant information in this assay system.

Table 4-6. Summary data for the 21-day, NF stage 51 exposure scenario using T4 in the OECD Phase 1 pre-validation study.

Expsosure scenario	Lab	Dose (nominal) µg/L	Mean dose (actual) µg/L ± SD	Median developmental stage (range)			Mean hind limb length (mm) ± SD			Mean body length (mm) ± SD GER/JP - WBL / US - SVL				Mean body weight (mg) ± SD
Day				**7**	**14**	**21**	**7**	**14**	**21**	**0**	**7**	**14**	**21**	**21**
Stage 51/21 days	US	0.0	0.0	ND	ND	59 (57-62)	ND	ND	ND	ND	ND	ND	19.5 ± 1.7	1041.8 ± 201.9
		0.25	0.27±0.03	ND	ND	59 (57-62)	ND	ND	ND	ND	ND	ND	19.4 ± 1.5	1051.7 ± 196.0
		0.5	0.52±0.07	ND	ND	59 (57-62)	ND	ND	ND	ND	ND	ND	18.9 ± 1.4	997.6 ± 152.9
		1.0	1.04±0.11	ND	ND	60 (59-62)	ND	ND	ND	ND	ND	ND	17.8 ± 1.8	844.8 ± 173.5
		2.0	2.13±0.51	ND	ND	61(59-63)	ND	ND	ND	ND	ND	ND	16.4 ± 1.9	639.7 ± 217.3
	Germany	0.0	NR	55 (54-55)	57 (56-57)	60 (58-62)	2.4±0.3	6.9±0.8	16.7±2.6	27.4±0.5	46.3±2.3	58.6±2.3	58.9 ± 4.2	823.5 ± 168.7
		0.25	NR	55 (54-55)	57 (56-57)	60 (57-62)	2.7±0.3	7.0±0.9	17.0±2.9	27.5±0.5	46.5±2.6	58.9±3.5	59.3 ± 3.4	761.9 ± 131.9
		0.5	NR	55 (54-55)	57 (56-57)	60 (58-62)	2.9±0.4	7.1±0.8	17.1±2.7	27.5±0.5	47.1±2.8	58.6±3.3	59.1 ± 2.9	755.9 ± 129.1
		1.0	NR	55 (55-55)	57 (57-58)	60 (59-62)	3.1±0.3	7.6±1.0	18.2±2.2	27.4±0.5	46.9±3.0	58.6±3.6	57.1 ± 4.1	691.5 ± 128.2
		2.0	NR	55 (55-56)	58 (57-58)	62 (61-65)	3.7±0.3	9.1±1.0	19.4±1.4	27.4±0.5	45.5±2.4	57.0±2.9	50.2 ± 8.6	549.6 ± 103.4
	Japan	0.0	0.0	54 (53-54)	56 (54-56)	58 (56-61)	2.0±0.3	4.8±1.0	11.3±2.6	23.9±1.4	39.2±3.1	51.2±3.8	56.1 ± 3.4	761.2 ± 127.4
		0.25	0.32±0.11	54 (53-54)	56 (55-57)	58 (57-62)	2.2±0.3	4.9±1.0	11.7±2.7	23.5±1.3	39.8±4.0	51.8±5.0	57.8 ± 3.4	845.0 ± 117.0
		0.5	0.48±0.01	54 (54-55)	56 (55-57)	59 (57-62)	2.2±0.3	5.2±1.1	12.0±2.8	23.8±1.6	39.3±3.9	51.8±4.1	55.8 ± 3.6	752.1 ± 138.8
		1.0	1.22±0.11	55 (54-55)	56 (56-57)	59 (58-62)	2.6±0.6	5.9±0.8	14.1±1.9	23.6±1.7	39.0±2.8	52.0±3.2	56.2 ± 2.7	735.2 ± 109.7
		2.0	2.74±0.42	55 (54-55)	58 (58-58)	60 (58-62)	2.9±0.4	7.4±0.9	14.6±1.6	23.6±1.5	38.4±3.0	48.7±3.4	51.1 ± 4.4	556.5 ± 106.1

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$. ND = Not done. NR = Not reported.

Table 4-7. Summary data for the 14-day, NF stage 54 exposure scenario using T4 in the OECD Phase 1 pre-validation study.

Expsosure scenario	Lab	Dose (nominal) µg/L	Mean dose (actual) µg/L ± SD	Median developmental stage (range)		Mean hind limb length (mm) ± SD		Mean body length (mm) ± SD GER/JP - WBL / US - SVL			Mean body weight (mg) ± SD
Day				7	14	7	14	0	7	14	14
Stage 54/14 days	US	0.0	0.0	ND	58 (56-60)	ND	ND	ND	ND	19.2 ± 0.9	942.7 ± 156.6
		0.25	0.21±0.01	ND	58 (57-60)	ND	ND	ND	ND	19.3 ± 1.1	984.1 ± 170.7
		0.5	0.47±0.05	ND	58 (56-61)	ND	ND	ND	ND	19.0 ± 1.0	931.5 ± 131.7
		1.0	0.83±0.14	ND	58 (58-59)	ND	ND	ND	ND	19.1 ± 0.9	924.2 ± 138.0
		2.0	1.71±0.13	ND	59 (58-62)	ND	ND	ND	ND	18.1 ± 1.0	813.6 ± 134.6
	Germany	0.0	NR	56 (55-57)	58 (57-60)	5.0 ± 0.7	12.2 ± 2.6	38.8 ± 0.9	53.7 ± 2.3	59.6 ± 2.4	860.7 ± 116.6
		0.25	NR	56 (55-57)	59 (57-62)	5.5 ± 0.7	13.1 ± 2.6	39.2 ± 1.2	54.4 ± 2.3	59.3 ± 2.6	870.6 ± 110.1
		0.5	NR	56 (55-57)	58 (57-60)	5.2 ± 0.6	12.1 ± 2.1	38.8 ± 0.8	54.1 ± 2.1	59.6 ± 3.2	875.9 ± 139.1
		1.0	NR	56 (55-57)	59 (57-62)	5.5 ± 0.6	13.1 ± 2.2	38.8 ± 0.9	53.7 ± 2.1	59.2 ± 3.0	858.4 ± 153.0
		2.0	NR	57 (56-58)	59 (58-61)	6.5 ± 0.5	14.9 ± 1.6	38.8 ± 1.0	52.8 ± 2.2	56.5 ± 3.4	735.1 ± 124.9
	Japan	0.0	0.0	55 (55-57)	57 (57-60)	4.2 ± 0.9	10.1 ± 2.6	37.5 ± 2.7	51.5 ± 3.4	59.6 ± 3.2	882.0 ± 130.7
		0.25	0.35±0.08	55 (55-57)	57 (57-60)	4.1 ± 0.9	8.9 ± 2.4	37.6 ± 2.5	47.9 ± 3.3	54.5 ± 2.7	689.1 ± 97.4
		0.5	0.65±0.05	55 (55-57)	57 (57-59)	4.2 ± 0.6	9.3 ± 1.7	37.3 ± 2.0	49.8 ± 3.1	57.2 ± 4.1	810.6 ± 163.3
		1.0	1.06±0.23	55 (55-56)	57 (57-59)	4.0 ± 0.4	8.4 ± 1.4	37.4 ± 2.2	45.0 ± 3.4	52.3 ± 5.0	603.7 ± 162.6
		2.0	2.00±0.84	56 (56-57)	59 (58-60)	5.0 ± 0.5	10.8 ± 1.4	37.3 ± 2.3	46.1 ± 3.8	52.2 ± 4.1	603.6 ± 126.2

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$.
ND = Not done. NR = Not reported.

Results from histopathological analyses of thyroid glands of T4-treated tadpoles were less consistent between the laboratories and were, thus, more difficult to interpret compared to the studies with PTU. The U.S. laboratory primarily observed changes in colloid content and colloid density. At the two greatest T4 concentrations examined, prevalent histological changes included: 1) collapsed follicles, and 2) reduced or absent follicular colloid. Similar effects on follicular colloid content were only rarely observed at the 2.0 µg/L T4 in the German laboratory. Quantitative analysis performed by the Japanese laboratory revealed a significant reduction in follicular lumen area and thyroid gland area at the two highest T4 concentrations (1.0 and 2.0 µg/L T4) in the 14-day exposure studies, but not in the 21-day experiment. The most dramatic change determined by the German laboratory was an increase in epithelial cell height at the highest T4 test concentration. Depletion of colloid stores and increases in epithelial cell height are known to occur at climax stages during normal development (38) when TSH synthesis and release by the pituitary (39;40) and T4 synthesis and secretion by the thyroid gland (25;37;38) reach maximum levels. Therefore, interpretation of the histological findings of the present study was confounded by the ability to determine whether the selected changes occurred in response to T4-induced alterations of the functional state of the HPT axis, or simply reflected the advanced stage of the tadpoles in the corresponding T4 treatment groups.

Summary results using data interpretation criteria are provided in **Table 4-8**. Because developmental acceleration is only known to occur with thyroid agonists, this trial indicates that all the laboratories detected T4 regardless of exposure scenario. Thyroid histopathology was less consistent in this trial, with both atrophy and hypertrophy of the thyroid glands reported, depending on the laboratory. It is possible that differences in the developmental stages that were evaluated contributed to the inconsistencies, however this theory has not been thoroughly investigated. Regardless of this inconsistency, the strong effects in developmental and growth markers deemed this assay “positive” in all laboratories.

Table 4-8. Summary data for T4 from the Phase 1 trial.

Criterion		Stage 51/21 days US	Stage 51/21 days GR	Stage 51/21 days JP	Stage 54/14 days US	Stage 54/14 days GR	Stage 54/14 days JP
Overt Toxicity	Mortality	-	-	-	-	-	-
	Hemorrhagic lesions	-	-	-	-	-	-
	Edema	-	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-	-
	Lethargy	-	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-	-
Delayed Development		-	-	-	-	-	-
Advanced Development	Developmental stage (d7 or 14/21)	↑	↑	↑	↑	↑	↑
	Wet weight (d7 or 14/21)	↓	↓	↓	↓	↓	↓
	SVL/WBL (d7 or 14/21)	↓	↓	↓	↓	↓	↓
	HHL (d7 or 14/21)	ND	↑	↑	ND	↑	E
	Thyroid histopathology	+	E	-	+	E	+
	Overall assay result:	+	+	+	+	+	+

ND = not done
E = Equivocal

Given the differences in the protocols used by the different laboratories, the Phase 1 exercise does not strictly allow for lab to lab comparisons. On the other hand, given the relatively consistent growth and development of the test organisms, and the reproducible effects of the test chemicals among the participating labs over three continents, these studies strongly indicate that the model system is relatively robust and not subject to variation as a function of the test protocols employed. These results, along with additional consultations, led to a more uniform approach for the Phase 2 validation studies, whose parameters were more tightly defined and controlled in order to permit inter-laboratory comparisons.

4.2 Multi-Chemical Study and Method Demonstration – USEPA

Objectives

The U.S. EPA conducted a multi-chemical study to demonstrate and evaluate the AMA, as developed by the U.S. EPA Office of Research and Development/National Health and Environmental Effects Laboratory (ORD/NHEERL) and OECD, using the following test materials: PTU, methimazole, T4, dexamethasone, phenobarbital, and pregnenolone-16-α-carbonitrile (PCN). The recommended protocols under consideration were: 1) 21-day exposure initiated at NF stage 51, and 2) 14-day exposure initiated at NF stage 54, which were generally consistent with the protocols used in the OECD Phase 1 study described in the preceding subsection. A full report of this study is included in attachment H.
The specific experimental objective was to determine the differential sensitivity of beginning exposure with NF stage 51 and stage 54 larvae, and to establish the appropriate duration of exposure. To achieve this objective, two concurrent studies with organisms from the same cohort, exposed from stage 51 for 21 days and from stage 54 for 14 days, were conducted for each chemical.

Overview of Methods

Prior to initiating the in-life studies, methods of analysis, including determination of method detection limits (MDLs); verification of chemical purity; and evaluation of stability (degradability), were defined. A summary of the experimental conditions used in the study is provided in **Table 4-9** with modifications identified below. Additional background work and range-finding studies were needed to assist in the ultimate selection of test concentrations for phenobarbital and PCN. Range-finding studies consisted of 4-day static-renewal exposures of NF stage 51 larvae to establish lethal and malformation-inducing concentrations. Nominal test concentrations are provided in **Table 4-9**. On exposure days 21 and 14 for the NF stage 51 and NF stage 54 experiments, respectively, all organisms in the study were anesthetized in MS-222, evaluated for developmental stage, weighed, and fixed in Bouin's solution for histological analysis of the thyroid glands. Within 96 hours, preserved specimens were rinsed 3 times in 70% (v/v) reagent alcohol and stored in 10% neutral buffered formalin (NBF).

The primary differences in the methodologies used from the OECD Phase 1 protocol included:

- **Exposure method** – PTU, methimazole, T4, and dexamethasone were tested using a flow-through diluter system, whereas phenobarbital and PCN were tested using a static-renewal approach due to the high test concentrations required and limits of aqueous solubility for these two materials.
- **Diet –** A mixture of trout starter, algae, TetraFin™ flakes, and live brine shrimp were used (USEPA diet).
- **Developmental stage** – Determined only at test termination (day 21).
- **Body length measurement –** Not performed.
- **Thyroid gland histology –** Preparation of histological samples and histological interpretation were performed by a board certified pathologist.

Table 4-9. Overview of nominal exposure concentrations used in the USEPA protocol demonstration and optimization studies.

Test Material	Concentrations (mg/L)	Exposure	Stage at Test Initiation	Duration
PTU	2.5, 5.0, 10.0, 20.0	FT	51/54	21/14 days
Methimazole	6.25, 12.5, 25.0, 50.0	FT	51/54	21/14 days
T4	0.0005, 0.001, 0.002, 0.004	FT	51/54	21/14 days
Dexamethasone	0.001, 0.0039, 0.0156, 0.0625, 0.250	FT	51/54	21/14 days
Phenobarbital	125.0, 250.0, 500.0, 1,000.0	SR	51	21 days
Phenobarbital	250.0, 500.0, 1,000.0, 1,500.0	SR	54	14 days
PCN	0.0074, 0.0222, 0.0667, 0.200	SR	51/54	21/14 days

The primary endpoints evaluated in the study were mortality, developmental stage, growth (wet weight), and thyroid histology. These endpoints, including the measurement format and the frequency of measurement during the proposed studies, are described in **Table 4-10**. Gross morphology was evaluated as a secondary endpoint.

Table 4-10. Primary and secondary endpoints for the USEPA protocol demonstration and optimization studies.

Primary Endpoint	Frequency of Measurement	Measurement Format
Mortality	Daily	% at Test Conclusion
Developmental Stage	At Test Conclusion	Individual Stage Distribution
Growth	At Test Conclusion	Weight
Thyroid Histology	At Test Conclusion	Pathology Summary
Secondary Endpoint		
Gross Morphology	At Test Conclusion	% External Abnormality

For each chemical, five randomly selected larvae from two replicates of each test concentration were allocated for thyroid gland histology. Developmental stage was analyzed based on the distribution of stages after 14 and 21 days of exposure, respectively. Non-parametric Kruskall Wallis-ANOVA was conducted on all developmental stage data sets. When warranted, pair-wise comparisons between treatments and control were conducted using Dunn's test ($p \leq 0.05$). Wet weight was analyzed using one-way ANOVA or Kruskall Wallis-ANOVA followed by pair-wise comparisons between treatments and control using Dunn's test or Bonferroni t-test ($p \leq 0.05$ for both).

Quality Assurance

Quality assurance measures used in the study were described in a Quality Assurance Project Plan (QAPP) specifically prepared for this study. The Quality Assurance Unit (QAU) reviewed each of the guidance documents, the data generated from the study, and the report. Study deviations and their significance on the outcome of the study are described in the latter portion of this subsection.

Overview of Results – Controls

Control mortality frequencies were ≤2.5% in every study. The mean stage of development obtained in controls (dechlorinated tap water) ranged from 56.0 ± 0.21 (dexamethasone experiment) to 57.7 ± 0.17 (phenobarbital experiment) for NF stage 51 specimens cultured for 21 days. The mean stage of development obtained in controls ranged from 57.1 ± 0.17 (dexamethasone) to 58.4 ± 0.26 (methimazole) for NF stage 54 specimens cultured for 14 days. In the tables below, statistical analyses of median stages are reported to provide consistency with the Phase 1 experiments. Mean wet weight of controls at the conclusion of the NF stage 51/21-day method ranged from 0.48 ± 0.02 g (PCN) to 0.85 ± 0.03 g (phenobarbital). Mean wet weight of controls at the conclusion of the NF stage 54/14-day method ranged from 0.59 ± 0.01 g (T4) to 0.90 ± 0.03 g (phenobarbital). Overall, the intra-study variability in control weight was low; however, the inter-study variability for wet weight of controls was somewhat greater. In addition, control body weight was more variable than developmental rate from both an intra- and inter-study standpoint. In each case, thyroid glands from control specimens appeared normal based on histological examination. As a point of reference, the rate of normal *X. laevis* development from NF stage 51 for 21 days and NF stage 54 for 14 days should yield NF stage 56 to 57 and 57 to 58, respectively, in accordance with Nieuwkoop and Faber (23).

Overview of Results – PTU

A summary of results for the PTU study can be found in **Table 4-11**.

Table 4-11. Summary data for PTU for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) mg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	57 (56-60)	0.736 ± 0.022
	2.5	57 (56-60)	0.728 ± 0.027
	5.0	57 (55-59)	0.726 ± 0.022
	10.0	56 (54-58)	0.729 ± 0.024
	20.0	54 (53-59)	0.757 ± 0.037
Stage 54/14 days	0.0	58 (56-60)	0.825 ± 0.030
	2.5	58 (57-60)	0.846 ± 0.020
	5.0	58 (56-60)	0.851 ± 0.025
	10.0	57 (56-59)	0.859 ± 0.028
	20.0	57 (56-59)	0.879 ± 0.022

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$.

PTU exposure slowed the rate of development in a concentration-dependent manner, but had no apparent effect on body weight. In NF stage 51 specimens (21-day exposure), increased frequency of enlarged, diffuse follicles was noted at 5.0 mg/L PTU. An increase in the frequency of follicular hyperplasia was noted at 10.0 mg/L PTU. A decrease in follicular colloid was noted in two specimens exposed to 20 mg/L PTU. In NF stage 54 specimens (14-day exposure), increased frequency of enlarged, diffuse follicles; follicular distension (increase in luminal area); and follicular hyperplasia were noted at 5.0 mg/L PTU. Changes in follicular colloid were not noted. Overall, histological effects were slightly more dramatic in NF stage 54 specimens exposed to PTU for 14-days compared to NF stage 51 specimens exposed to PTU for 21-days. No significant gross morphological abnormalities were detected. These results are consistent with the findings in the Phase 1 trial using PTU. Combined summary results are provided in **Table 4-12**. Given the combined developmental delay with significant histopathological changes in the thyroid glands, the assay was deemed “positive” for both exposure scenarios.

Table 4-12. Summary data for PTU from Contractor lab coupled with results from Phase 1.

Criterion		Stage 51/21 days US	GR	JP	C	Stage 54/14 days US	GR	JP	C
Overt Toxicity	Mortality	-	-	-	-	-	-	-	-
	Hemorrhagic lesions	-	-	-	-	-	-	-	-
	Edema	-	-	-	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-	-	-	-
	Lethargy	-	-	-	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-	-	-	-
Advanced Development		-	-	-	-	-	-	-	-
Delayed Development	Developmental stage (d7 or 14/21)	↓	↓	↓	↓	↓	↓	↓	↓
	Wet weight (d7 or 14/21)	↑	-	↓	-	-	-	-	-
	SVL/WBL (d7 or 14/21)	↑	↓	E	ND	↑	↓*	-	ND
	HHL (d7 or 14/21)	ND	↓	↓	ND	ND	-	↓	ND
Thyroid histopathology		+	+	+	+	+	+	+	+
Overall assay result:		+	+	+	+	+	+	+	+

ND = not done
E = equivocal
C = contractor laboratory
*No concentration response relationship

Overview of Results – Methimazole

A summary of results for the methimazole study can be found in **Table 4-13**.

Table 4-13. Summary data for methimazole for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) mg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	57 (55-59)	0.539 ± 0.020
	6.25	56 (54-59)	0.592 ± 0.027
	12.5	55 (52-57)	0.503 ± 0.025
	25.0	54 (53-55)	0.554 ± 0.042
	50.0	54 (53-56)	0.606 ± 0.040
Stage 54/14 days	0.0	59 (56-62)	0.636 ± 0.026
	6.25	59 (56-61)	0.708 ± 0.026
	12.5	57 (56-59)	0.864 ± 0.041
	25.0	57 (55-58)	0.755 ± 0.032
	50.0	56 (55-59)	0.752 ± 0.037

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$. Statistics reported for the stage 51/21 days weight data in the report were performed on medians instead of means. Re-analysis of these data using the mean body weight is reflected above.

Methimazole exposure inhibited the rate of development in a concentration-dependent manner for organisms exposed for 21 days (NF stage 51) and 14 days (NF stage 54). There was no apparent effect on the body weight for organisms in the 21-day study, but a slight increase in growth (weight) in larvae in the 14-day study was detected. In NF stage 51 specimens (21-day exposure), an increase in follicular distension; enlarged, diffuse

follicles; and follicular hyperplasia were noted at 6.25 mg/L methimazole. A decrease in follicular colloid was noted in several specimens exposed to 12.5 mg/L methimazole. In NF stage 54 specimens (14-day exposure), increased frequency of enlarged, diffuse follicles; follicular distension; follicular hyperplasia; and a reduction in follicular colloid were noted at 12.5 mg/L methimazole. The concentration dependent delay in development coupled with altered thyroid gland histology deemed this assay “positive”.

Overview of Results – T4

A summary of results for the T4 study can be found in **Table 4-14**.

Table 4-14. Summary data for T4 for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) µg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	57 (55-59)	0.531 ± 0.018
	0.5	56 (56-59)	0.549 ± 0.017
	1.0	59 (59-62)	0.464 ± 0.016
	2.0	60 (59-63)	0.288 ± 0.016
	4.0	63 (63-64)	0.102 ± 0.008
Stage 54/14 days	0.0	59 (57-61)	0.587 ± 0.014
	0.5	58 (57-61)	0.614 ± 0.019
	1.0	59 (57-61)	0.556 ± 0.016
	2.0	59 (59-61)	0.490 ± 0.018
	4.0	62 (61-64)	0.239 ± 0.010

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$.

T4 exposure accelerated the rate of development and reduced larval body weight in a concentration-dependent manner in both the 21-day and 14-day assays. High larval mortality was observed in the 4.0 µg/L group in the 21-day study. Due to this mortality, histology was not performed on this set. For the other specimens in the 21-day study, decreased follicular size and colloid content were noted at 0.5 µg/L T4. An increase in the foamy nature of colloid was noted with increasing concentration. In NF stage 54 specimens (14-day exposure), thyroid gland tissue was not found in three of the 10 specimens exposed to 2.0 µg/L T4 and no thyroid gland tissue could be found in NF stage 54 larvae exposed to 4.0 µg/L T4. Decreased colloid was observed at 0.5 µg/L T4. In general, the histological effects were more dramatic in the NF stage 51 larvae exposed to T4 for 21 days. **Table 4-15** presents summary results of the assay coupled with results from the Phase 1 studies. Like the U.S. lab in the Phase 1 studies, thyroid histology revealed evidence of glandular atrophy, and in some larvae, the glands could not be found. Because advanced development is only known to occur through thyroid-mediated pathways, this assay was deemed “positive”.

Table 4-15. Summary assay results for T4 coupled with Phase 1 results.

Criterion		Stage 51/21 days US	GR	JP	C	Stage 54/14 days US	GR	JP	C
Overt Toxicity	Mortality	-	-	-	+	-	-	-	-
	Hemorrhagic lesions	-	-	-	-	-	-	-	-
	Edema	-	-	-	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-	-	-	-
	Lethargy	-	-	-	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-	-	-	-
Delayed Development		-	-	-	-	-	-	-	-
Advanced Development	Developmental stage (d7 or 14/21)	↑	↑	↑	↑	↑	↑	↑	↑
	Wet weight (d7 or 14/21)	↓	↓	↓	↓	↓	↓	↓	↓
	SVL/WBL (d7 or 14/21)	↓	↓	↓	ND	↓	↓	↓	ND
	HHL (d7 or 14/21)	ND	↑	↑	ND	ND	↑	E	ND
Thyroid histopathology		+	E	-	+	+	E	+	+
Overall assay result:		+	+	+	+	+	+	+	+

ND = not done
E = equivocal
C = contractor laboratory

Overview of Results – Dexamethasone

A summary of results for the dexamethasone study can be found in **Table 4-16**.

Table 4-16. Summary data for dexamethasone for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) µg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	56 (54-59)	0.695 ± 0.041
	1.0	56 (55-59)	0.721 ± 0.034
	3.9	56 (55-59)	0.609 ± 0.028
	15.6	56 (53-59)	0.489 ± 0.024
	62.5	55 (53-57)	0.375 ± 0.017
	250.0	55 (53-57)	0.355 ± 0.018
Stage 54/14 days	0.0	57 (55-59)	0.790 ± 0.026
	1.0	57 (55-59)	0.782 ± 0.025
	3.9	57 (55-60)	0.696 ± 0.020
	15.6	57 (55-59)	0.623 ± 0.018
	62.5	56 (55-57)	0.564 ± 0.013
	250.0	56 (55-57)	0.513 ± 0.013

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$. The attached report indicates that the median was used for statistical analysis of the body weights. Re-analysis using means was performed and is reflected above.

Dexamethasone exposure inhibited the rate of development and reduced growth (weight) in a concentration-dependent manner. In the 21-day exposure specimens, minimal decreases in follicular size and colloid were noted at 15.6 µg/L and minimal to mild decreases were noted at 62.5 and 250 µg/L dexamethasone. In the 14-day study, no histological effects were noted. There were no significant findings suggesting overt

toxicity such as excessive mortality or abnormal behaviors. Strong developmental effects suggest that this chemical affects the HPT axis, however without strong evidence of histological changes and further endpoint analysis (HHL, WBL), it is difficult to determine the mode of action of this chemical. Therefore, using the data interpretation criteria, this chemical could potentially be considered toxic as opposed to positive.

Overview of Results – Phenobarbital

A summary of results for the Phenobarbital study can be found in **Table 4-17**.

Table 4-17. Summary data for phenobarbital for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) mg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	57 (56-59)	0.852 ± 0.033
	125.0	59 (57-59)	0.912 ± 0.026
	250.0	59 (57-59)	0.958 ± 0.024
	500.0	59 (57-60)	0.932 ± 0.028
	1000.0	59 (57-59)	1.016 ± 0.030
Stage 54/14 days	0.0	59 (56-60)	0.897 ± 0.031
	250.0	59 (57-59)	1.007 ± 0.037
	500.0	58 (57-59)	1.095 ± 0.044
	1000.0	59 (57-59)	0.993 ± 0.027
	1500.0	58 (57-59)	0.966 ± 0.027

Summary data are pooled across replicates. Shaded cells indicate statistically significant differences from control at $p<0.05$.

Phenobarbital exposure nominally increased the rate of development and weight in NF stage 51 larvae exposed for 21 days. Phenobarbital exposure had no appreciable effect on the rate of development and weight in NF stage 54 organisms exposed to phenobarbital for 14 days. In the 21-day study specimens, an increase in enlarged, diffuse follicles, and follicular cell hyperplasia were noted at 1,000 mg/L phenobarbital. In the 14-day study specimens, increased frequency of follicular cell hyperplasia was noted at 1,500 mg/L phenobarbital. Developmental acceleration is thought to only occur through thyroid-mediated mechanisms, and the current data interpretation criteria would therefore deem this a positive for the stage 51/21 day exposure and positive for the stage 54/14 day exposure based on histological effects. Accelerated development, however, was not expected, and interpretation of these data is difficult owing to the sedative effects of phenobarbitol. Thyroid histopathology did not support the findings of accelerated development given that follicular cell hyperplasia and enlarged follicular lumen were observed, findings typically associated with antagonistic chemicals. The Contract laboratory did not report excessive mortality, however other signs of overt toxicity were not evaluated. The assay result is considered “positive”.

Overview of Results – PCN

A summary of results for the PCN study can be found in **Table 4-18**.

Table 4-18. Summary data for PCN for the USEPA protocol demonstration and optimization studies.

Expsosure scenario	Dose (nominal) µg/L	Median developmental stage (range)	Mean body weight (g) ± SEM
Stage 51/21 days	0.0	56 (55-59)	0.481± 0.023
	7.4	57 (56-59)	0.758 ± 0.025
	22.2	57 (55-59)	0.680 ± 0.021
	66.7	57 (57-59)	0.731 ± 0.015
	200.0	57 (56-59)	0.650 ± 0.016
Stage 54/14 days	0.0	58 (57-59)	0.827 ± 0.025
	7.4	58 (56-59)	0.783 ± 0.023
	22.2	59 (57-60)	0.830 ± 0.019
	66.7	59 (57-60)	0.845 ± 0.020
	200.0	57 (57-59)	0.853 ± 0.027

Summary data are pooled across replicates.

PCN exposure had no appreciable effect on the rate of development or weight in either the 21-day study or the 14-day study. It should be noted, however, that the potential effect of PCN on larval growth (weight) is not completely clear since the control weights for the 21-day study were low. In NF stage 51/21-day specimens, an increase in follicular cell hyperplasia was noted at 22.2 µg/L PCN. Increased incidence of follicular karyomegaly was noted in several specimens exposed to 66.7 µg/L PCN. In the 14-day study, increased frequency of enlarged, diffuse follicles and follicular hyperplasia were noted at 22.2 µg/L PCN. Follicular karyomegaly was found in only one specimen (22.2 µg/L treatment) in the NF stage 54 larvae exposed to PCN for 14 days. These effects were not reported at lower test concentrations. Lack of developmental delay or acceleration and clear histological results, data interpretation would result in a negative assay outcome.

Quality Assurance

Study deviations and data errors regarding the in-life phase of study were found to be relatively minor in nature, generally inconsequential, and corrected in each case by the Study Director. Several deviations in test chemical analyses were noted. These deviations included sample holding time ascendances during evaluation of T4 and phenobarbital. In addition, the measured concentration of PCN was substantially less (ca. 10-fold less) than originally intended. This may have been the result of difficulties in solubilizing a relatively insoluble chemical in large quantities of test material (50 L). In turn, the less than anticipated stock and test concentrations may have been the result of sample degradation. Differences in analytical methodologies used, liquid chromatography-mass spectrometry (LC-MS)/MS and high performance liquid chromatography (HPLC)/photodiode array detector may have also been a factor in the success of analysis. The final set of PCN samples analyzed from the in-life test, with the exception of the stock solution, which measured a concentration consistent with previous

stock solution samples, was found to be not detectable. Since the high concentration and stock solution were theoretically the same nominal concentration, this discrepancy cannot be explained by the study team. Tissue residues were measured to evaluate exposure and uptake. Tissue residues were found to be generally consistent with exposure and followed a concentration-exposure pattern.

Endpoint Sensitivity

Of the primary endpoints measured in this study, changes in thyroid gland histology were noted with each of the six test materials. Alteration of developmental stage was also found in studies with five of the six test materials. Both T4, and to a lesser degree, phenobarbital accelerated development, whereas PTU, methimazole, and dexamethasone, were found to inhibit developmental rate. Alteration in wet weight was found in specimens exposed to five of the six test materials. T4 and dexamethasone both inhibited growth (weight), whereas phenobarbital (NF stage 51 at 21-day exposure only), and methimazole (NF stage 54 at 14-day exposure only) increased weight. Histological changes were found at test concentrations below those inducing effects on developmental rate or weight in organisms exposed to PTU, methimazole, T4, phenobarbital, and PCN. Based on these findings, thyroid gland histology appeared to be the most consistent and sensitive indicator of thyroid axis impairment in this study. Alteration of developmental rate was somewhat less sensitive than histology, but was more sensitive than weight. This relationship was not necessarily surprising since thyroid histology is the most direct link to the thyroid axis of the endpoints measured. Effects on weight, and to a lesser extent developmental stage, may or may not be directly indicative of effects induced at the thyroid axis level. Further, effects on growth alone may be the result of indiscriminant toxicological insult. Without clear histological data, understanding the effects of phenobarbital and PCN on the thyroid axis is difficult.

Sensitivity of Exposure Protocol

Results from this study suggest that the sensitivity of the stages used for exposure and the importance of the duration of exposure were test material-dependent. Measurement endpoint sensitivity was also reflected in the sensitivity of the two different exposure periods. For example, in terms of developmental rate and weight, PTU was nearly equipotent in both exposure profiles. However, somewhat more dramatic histological effects were observed in organisms exposed at NF stage 54 for 14 days than NF stage 51 organisms exposed for 21 days. In this case, both the 21-day and 14-day exposures were similarly sensitive to the development of diffuse enlargement of the thyroid gland. However, the occurrence of follicular cell hyperplasia and pale, foamy colloid was greater among the 14-day exposure specimens than the 21-day exposure specimens.

A similar relationship between the two exposure profiles was observed with methimazole. Similar to PTU, both exposure scenarios detected development of diffuse enlargement of the thyroid gland, follicular hyperplasia, and the presence of pale and foamy colloid. However, the 14-day exposure resulted in an increased sensitivity to decreased colloid compared to the 21-day exposure. T4 was equipotent in terms of

effects on developmental rates and weight in both exposure profiles. Like T4, dexamethasone was equipotent in terms of effects on developmental rates and weight in both exposure profiles. Marked stage sensitivity differences were not observed with either phenobarbital or PCN. A stimulatory effect on weight in NF stage 54 organisms exposed to methimazole for 14 days was noted, but not in NF stage 51 organisms exposed to methimazole for 21 days.

In terms of the total number of significant responses for effects on developmental stage and weight for both exposure scenarios evaluated in the study, a greater number of positive or negative responses were observed with the NF stage 51 organisms exposed to test material for 21 days. This response was not necessarily surprising because of the longer and more inclusive exposure period. Without consideration of thyroid histology, the 21-day exposure profile would appear to be more sensitive. However, when thyroid histology was considered in the endpoint battery, the 14-day exposure using NF stage 54 organisms appeared to be somewhat more sensitive for PTU, which is a potent goitrogen. The slightly decreased sensitivity of the NF stage 51 organisms exposed to PTU, and to a lesser extent methimazole, may be the result of increased compensatory response stemming from prometamorphic exposure. However, more work will be needed to delineate this hypothesis. On the contrary, the 21-day exposure profile appeared to be somewhat more sensitive for T4, dexamethasone, phenobarbital, and PCN. Overall, the study demonstrated the importance of histology endpoints to more clearly evaluate thyroid axis impairment. In the case of dexamethasone, which induced only slight histological changes in thyroid tissue in the 21-day study, but not in specimens subjected to dexamethasone for 14 days, alteration in development and weight was most likely the result of a non-thyroidal mechanism of action. It is possible, however, that dexamethasone affected TH metabolism.

Conclusions – Multi-Chemical Study and Method Development

Results from this demonstration study indicated that PTU, methimazole, and dexamethasone were capable of inhibiting the rate of development as marked by developmental stage, whereas T4 was capable of accelerating development. Results from phenobarbital were equivocal given accelerated developmental stages in two dose groups (in a non-dose dependent manner) – a finding contrary to expected effects of this chemical. PCN had no effect on the rate of development. T4 and dexamethasone inhibited growth (weight), while methimazole slightly increased growth (weight). Phenobarbital exposure did not markedly alter weight. The effect of PCN on larvae weight is not completely clear since the control weight for the study with NF stage 51 larvae was low. PTU, methimazole, T4, phenobarbital, and PCN each induced histological change in thyroid gland tissue characteristic of thyroid axis disturbance. Although minimal histological effects on the thyroid glands were detected in NF stage 51 organisms exposed to dexamethasone for 21 days, no substantive histological changes were noted in the 14-day exposure studies, indicating that the developmental and growth effects may not be the result of effects on the thyroid gland, or the thyroid axis. Both exposure scenarios were effective in detecting developmental and growth changes with the appropriate test substances. However, organisms exposed to PTU, and to a lesser

extent methimazole, at NF stage 54 for 14 days demonstrated a slightly increased sensitivity to histological changes of the thyroid gland, whereas specimens exposed to T4 and phenobarbital at NF stage 51 for 21 days demonstrated a slightly greater sensitivity to histological changes.

5 INTER-LABORATORY EVALUATION OF THE AMA (PHASE 2)

The work that was performed during the OECD Phase 1 studies, in combination with work conducted independently of the formal OECD validation effort (9;41-43) demonstrated that test protocol optimization and standardization, including husbandry practices, choice of developmental period, and duration of the assay dramatically increased the reproducibility and overall quality of method performance. Primary decisions that resulted from these studies include:

1. The stage of initiation of the method for the AMA is NF stage 51 tadpoles;
2. The duration of exposure is 21 days;
3. Flow-through exposure scenario is preferred;
4. Method endpoints include: developmental stage, hind limb length, wet weight, body length, and thyroid histology.

In addition, prior to the performance of the OECD Phase 2 study, improvements in test endpoint measurement and evaluation, and techniques for minimizing the variability of individual growth and developmental rates within the population of test animals were evaluated. To harmonize and standardize the methodologies for morphometric length measurements in different labs, a technical guidance document, ***Amphibian Metamorphosis Assay Part 1: Technical guidance for morphologic sampling and histological preparation*** was prepared (attachment A – appendix 1). This document details the methods for measuring body and limb lengths as well as the acquisition, fixation, and embedment of tissues for subsequent thyroid histology. Because thyroid histology is a key diagnostic endpoint of the study protocol, efforts were undertaken to standardize the methods used to obtain, process, and analyze thyroid tissue sections in order to improve the comparability of the information. In addition to the technical guidance document, through several consultations with pathologists experienced in reading studies using amphibian tissues, a second document, ***Amphibian Metamorphosis Assay Histopathology Part 2: Approach to reading studies, diagnostic criteria, severity grading, and atlas*** (attachment A – appendix 2), was prepared. This document presents general guidelines for study reading practices, detailed diagnostic criteria and severity grading schemes for various observations, a reference atlas of normal microanatomy of *X. laevis* thyroid glands, an atlas of core diagnostic criteria with examples of tissues exhibiting effects with different severity grades, and a recommendation for uniform data recording and compilation. These documents were distributed to all participating laboratories.

Given these technical improvements, an OECD Phase 2 study was initiated to assess inter-laboratory transferability and reliability of the test method using three chemicals with different modes of activity on the HPT axis: T4, perchlorate (PER) and iopanoic

acid (IOP). T4 is the native prohormone and is used as a reference compound for agonist activity on the thyroid system. Perchlorate is a well-known inhibitor of iodine uptake by the thyroid gland and has been shown to retard metamorphosis in *X. laevis* tadpoles (44). IOP is an inhibitor of all monodeiodinases, enzymes which catalyze the metabolism of TH in peripheral tissues (45). Therefore, IOP was used as a reference compound for modulation of peripheral TH action. IOP was expected to enhance TH action in tissues that are normally protected from TH by expression of type III monodeiodinase (e.g., tail), while this compound may block TH action in tissues that require efficient conversion of T4 to T3 (e.g., hind limbs).

Concerning validation, the OECD Phase 2 study evaluated transferability of the test method, measured by the ability of laboratories to perform the assay with general guidance from experienced laboratories, and reliability, measured by assessing the diagnostic value of each endpoint and the consistency in which it was determined in each of the participating laboratories. Overall, the study sought to determine the relevance of the biological model to detect thyroid activity.

5.1 Study Design

For more details concerning the study design, please refer to the OECD Phase 2 report (attachment C). Generally, the Phase 2 studies involved 6 international laboratories which performed a total of 14 exposure experiments including 5 experiments with PER, 4 experiments with T4, and 5 experiments with IOP. Randomly selected NF stage 51 larvae were exposed to four test concentrations and a dilution water control for 21 days. Each test concentration and control was run in quadruplicate, with 20 organisms per replicate. Once in the test system, mortality observations were made daily and any dead larvae were immediately removed. On day 7, developmental stage, hind limb and body lengths, and wet weight were determined on larvae randomly selected (5/replicate), euthanized, and preserved for possible histology. The tests were terminated on day 21, at which time all test animals were staged (NF), measured (cm), weighed (g), and visually observed for morphological abnormalities. Five larvae per replicate were randomly selected, euthanized, and preserved for histology. Critical test parameters and experimental conditions for the in-life study are presented in **Table 5-1.**

Statistical Analyses

The statistical approach used in Phase 2 was the step-down Jonckheere-Terpstra test, unless there was compelling evidence of a non-monotone dose-response. No such evidence was found in the Phase 2 study. An alternative to the Jonckheere-Terpstra test that was considered was the Williams' test, also applied in step-down fashion. Unlike the Jonckheere-Terpstra test, Williams' test requires the data be normally distributed with homogeneous variances. Fortunately, with only two exceptions, these conditions were satisfied. Where pairwise tests were desired, the Dunnett test was most powerful and appropriate, provided the data were normally distributed with homogeneous variances. When these conditions were not satisfied, a transformation was sought that normalized the data and stabilized the variances. If the data were normally distributed but heterogeneous, and no variance stabilizing transformation could be found, a robust

version of a Dunnett test, referred to as the Tamhane-Dunnett test, requiring only normality but not variance homogeneity, was used. When the data could not be normalized, Dunn's test was the choice among those considered in the accompanying power analysis. It was also an alternative to consider in the normal, heterogeneous case, but it has lower power than the Tamhane-Dunnett test.

Developmental stage was analyzed using the Jonckheere-Terpstra, Dunn and Mann-Whitney tests. The issue of individual frogs versus tanks as the unit of analysis was the same for this response as for the other four. However, since developmental stage is measured on an ordinal scale, not a ratio scale, analysis was based on the replicate median rather than the replicate mean. Regarding the statistical analysis of data from the severity grade scoring of thyroid tissue, a specific statistical approach is currently under development and therefore is not reported here.

5.2 Performance of the Controls

The consistency and performance of control animals, under specified husbandry conditions, was evaluated so as to minimize uncertainty when making comparisons to dosed groups. In doing so, it is also imperative to refine husbandry methods and other test factors to ensure optimal and consistent performance of controls. Because changes in body size, body weight and morphology are the basis for assessing chemical-induced effects on TH-dependent metamorphosis, consistency in control performance is paramount. The performance of control animals was evaluated using growth and development rates among the three laboratories. Control performance was characterized using mean population rates, and within-population variance in growth and development.

Developmental Stage

To evaluate the consistency and reproducibility of control animal growth and development, a set of criteria was established. The criteria included: 1) the time since hatching required to achieve test initiation NF stage 51 (pre-test phase), 2) the developmental stage obtained 7 days after the initiation of the exposure protocol (beginning at NF stage 51), and 3) the developmental stage obtained at the conclusion of the exposure phase (21 days). For the first criterion, according to Nieuwkoop and Faber (23) and experience from several of the participating investigators, *X. laevis* tadpoles should reach stage 51 within approximately 14-17 days post-hatch at a temperature of 22-24°C. In general, the age of larvae at test initiation (NF stage 51) fell into a range of 11 to 17 days, with a majority falling between 11 and 14 days of age.

The second criterion is the median developmental stage reached by the control animal population within the initial 7 days of the exposure protocol. Under optimal rearing conditions, control tadpoles should reach early prometamorphic stages (NF stages 54-55) by study day 7. Based on historical records within the participating laboratories, tadpoles should have a whole body length of > 35 mm and a wet weight of >300 mg at NF stages 54-55. Nieuwkoop and Faber (23) suggested that the body length of a stage 54 tadpole was 58-65 mm. Slow or sporadic growth in tadpoles from within a given experiment was thought to be indicative of sub-optimal culture. Analysis of the stage distribution on day 7 indicated that less than optimal control performance occurred in two experiments during the study. Control animals from the IOP study by laboratory

3 showed development to only stage 53 within 7 days suggesting slower development than anticipated, yet consistent with that reported by Nieuwkoop and Faber (23). An increased range of stages (4 stages) determined in the controls of the T4 experiment in laboratory 3 indicated increased heterogeneity in developmental trajectory at an early time point during this experiment, although lower variance in developmental stages was achieved by the conclusion of the test.

The final criterion evaluated was the developmental stage obtained at the conclusion of the experiment. Increased variability relative to the other laboratories in developmental stages in the control group was found in all three experiments performed in laboratory 3. A primary difference between laboratory 3 and the other laboratories was the feeding rate of the tadpoles. In the study protocol, a recommendation for feeding rates was to begin feeding with a ration of ~600 mg Sera Micron/tank/day which equates to ~30 mg food per animal. This ration increased proportionally as tadpoles grew. Approximately 30-50% of the food amount applied in other labs was available to the animals in laboratory 3. Overall, the control data from experiments with daily feeding rates of 30 mg/animal and higher indicate that such feeding rates control the variability of developmental rates. However, this conclusion has not been substantiated statistically, and alternative explanations may exist.

Table 5-1. Critical test parameters and conditions for the OECD Phase 2 inter-laboratory studies.

Test Animal		*Xenopus laevis* larvae	
Initial Larval Stage		NF Stage 51	
Exposure Period		21 days	
Larvae Selection Criteria		Developmental Stage and Optional Total Length	
Stock Concentration / Test Chemical Concentrations	T4	4.04 mg/L and 0.25, 0.5, 1.0, 2.0 µg/L	
	PER	25.0 mg/L and 8.0, 32.0, 128.0, 512.0 µg/L	
	IOP	1000 mg/L and 750, 1500, 3000, 6000 µg/L	
Exposure Regime		Flow-Through	
Pre-Exposure Flow Rate / Test System Flow-Rate		50 mL/min / 25 mL/min	
Endpoints / Determination Days		Mortality	Daily
		Developmental Stage	Days 0, 7, and 21
		Hind Limb Length	Days 7 and 21
		Whole Body Length	Days 7 and 21
		Snout-Vent Length	Days 7 and 21
		Wet Body Weight	Days 7 and 21
		Thyroid Histology	Days 7 and 21
Dilution Water / Laboratory Control		Dechlorinated Tap Water (charcoal-filtered)	
Larval Density		20 Larvae / Test Vessel (5 / L)	
Test Solution / Test Vessel		4 L (10-15 cm minimum water)	
Replication		4 Replicates / Test Concentration and Control	
Acceptable Mortality Rate in Controls		± 5%	
Thyroid Fixation	Number Fixed	5 / Replicate (randomly selected)	
	Region	Head	
	Fixation Fluid	Davidson's Fixative	
Feeding	Food	Sera Micron	
	Frequency / Amount	Twice daily / Quantity adjusted with Larval Age	
Lighting	Photoperiod	12 h Light : 12 h dark	
	Intensity	600 to 2000 lux at Water Surface	
Water Temperature		22º ± 1ºC	
pH		6.5 to 8.5	
Dissolved Oxygen (DO) Concentration		>3.5 mg/L (>40% Air Saturation)	
Analytical Chemistry Sample Schedule		Once / Week (4 Sample Events / Test)	

Growth Metrics

Growth endpoints, including WBL, SVL and wet weight, were also determined as a measure of control animal performance. Mean values for WBL, SVL and wet weight determined for individual replicate tanks in the control groups were compared among all 14 tests performed during the study. In general, tadpole growth was highly reproducible among replicate tanks of the control group in experiments performed within laboratories 1-2 and laboratories 4-6. Variability was evaluated by calculating the percent coefficient of variation (%CV) for each aquarium using individual animals as replicates. The %CV values for replicate tanks was <15% for WBL and SVL, and <30% for wet weight in laboratories 1-2 and 4-6. By contrast, variability of growth rates within and among replicates was markedly higher in all three experiments performed in laboratory 3. The higher %CV values for all three growth-related parameters on day 21 in laboratory 3 may be the result of high proportion of tadpoles developing to climax stages at the time of the measurement. Climax stages are associated with marked decreases in body size and body weight. Thus, mean values for WBL, SVL and wet weight will be naturally more variable if a larger number of animals reach climax stages 61 or greater. Inter-laboratory comparisons showed highly reproducible tadpole growth rates; data from laboratory 3 were excluded from the analysis owing to high variability in development and growth rates and the large proportion of animals reaching climax stages.

5.3 Summary of Results – Sodium Perchlorate

Exposure studies using sodium perchlorate were conducted in each of five laboratories. The nominal concentrations for perchlorate were 0, 62.5, 125, 250 and 500 µg/l. Summary data for gross morphological endpoints are presented in **Table 5-2**. Based on measurements of WBL, SVL, wet weight, and mortality, it was concluded that exposure to perchlorate did not induce systemic toxicity in the developing larvae at any of the doses tested. Analyses of developmental stage and HLL suggested that exposure to perchlorate induced developmental delay in two of the experiments conducted. Histological assessment of thyroid tissue was performed in four of the five participating laboratories. Exposure to perchlorate induced marked effects on thyroid gland histology in all four laboratories. Furthermore, the occurrence and severity of histological changes was concentration-dependent. Thus, these data provide evidence for disruption of the tadpole thyroid axis by perchlorate, which is consistent with findings reported in the scientific literature (43;44;46;47).

Test performance for the five experiments with perchlorate was generally acceptable and is discussed in the following paragraphs. Increased variability in control stage development and lower growth rates compared to other laboratories was noted in results reported from laboratory 3. Lower feeding rates used in this laboratory appeared to be responsible for these results. Measured test concentrations were close to the expected nominal concentrations. For the 62.5 µg/l dose group, labs maintained actual concentrations between 92-143% of nominal; for the 125 µg/l dose group, labs maintained concentrations between 91-123% of nominal; for the 250 µg/l dose group, labs maintained concentrations between 88-115% of nominal; and for the 500 µg/l dose group, labs maintained concentrations between 89-118% of nominal.

Mortality

No mortality was observed in any treatment group in laboratories 3 and 5. The frequency of mortality was very low (< 5%) in tests performed in laboratories 1, 2 and 4.

Growth

Growth measurements were performed for a subset of animals (n=5 animals per replicate tank) on study day 7 and for the remaining animals at test conclusion (day 21). Results from four of five tests conducted noted significant increases in body length and weight – lab 4 found significant differences at day 7 at the 250 and 500 µg/L perchlorate concentrations, whereas the other 3 labs detected significant differences at day 21. Interestingly, laboratory 4 did not detect significant differences in growth at test conclusion. Concentration-related increases in all three growth-related parameters (WBL, SVL and wet weight) were observed at 125, 250 and 500 µg/L perchlorate in laboratory 1, and at 250 and 500 µg/L perchlorate in laboratory 3. In laboratory 1, a large number of control animals showed development to climax stages which corresponded to weight loss and a reduction in body size, thus distorting assessment of growth effects at higher perchlorate concentrations. Significant increases in tadpole length (250 and 500 µg/L perchlorate), but not wet weight, were observed in results from laboratory 5.

The sensitivity of WBL and SVL measurements to detect changes in tadpole size were similar in laboratories 1, 3, and 4; whereas SVL was slightly less sensitive than WBL in laboratory 5. The relationship between length and weight was not consistent. For example, in laboratory 1, significantly greater wet weights corresponded to greater WBL and SVL at each of the perchlorate exposure concentrations tested. In laboratory 3, wet weight increased at 125 µg/L perchlorate, but WBL and SVL were unaffected at this dose. In contrast, significant changes in tadpole WBL or SVL did not correspond with significant changes in wet weight in studies conducted by laboratory 5.

Table 5-2. Summary data for gross morphological endpoints from the OECD Phase 2 perchlorate study.

	Mean Whole Body Length										Mean Snout Vent Length										Mean Wet Weight									
	Day 7					Day 21					Day 7					Day 21					Day 7					Day 21				
Dose (µg/L) / Lab	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
62.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
125	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	+	-	-
250	-	-	-	+	-	+	-	+	-	+	-	-	-	+	-	+	-	+	-	-	-	-	-	+	-	+	-	+	-	-
500	-	-	-	+	-	+	-	+	-	+	+	-	-	+	-	+	-	+	-	+	+	-	-	+	-	+	-	+	-	-

	Mean Hind Limb Length										Median Developmental Stage									
	Day 7					Day 21					Day 7					Day 21				
Dose (µg/L) / Lab	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
62.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
125	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
250	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	+	-	-
500	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	+	+	-

+ indicates statistically significant differences from controls (p<0.05).

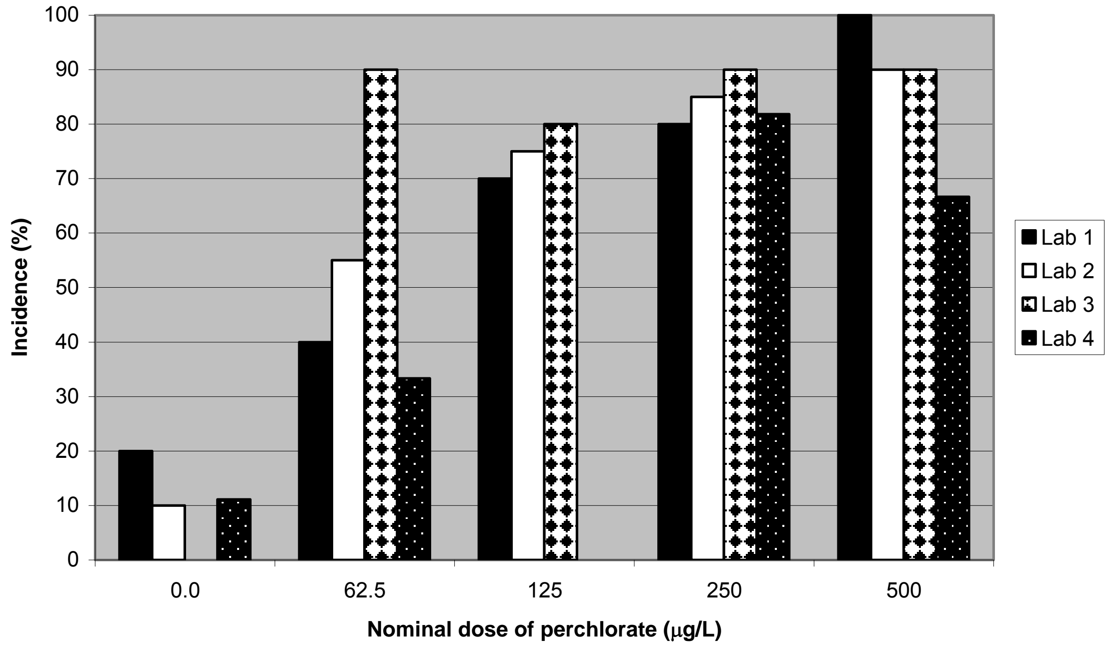
Development

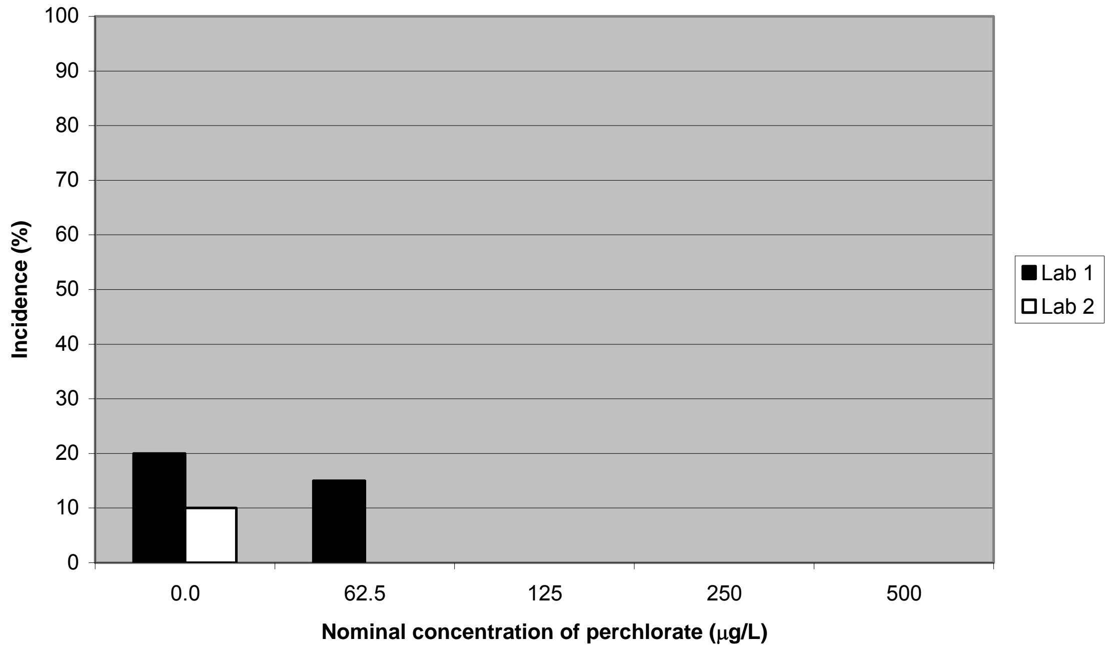
Perchlorate exposure induced developmental delay in two of the five laboratories (3 and 4). Effects on developmental stage and HLL were greatest at test termination (study day 21) compared to study day 7. On day 7, a small but statistically significant effect on median developmental stage was detected in larvae exposed to 500 µg/l perchlorate in laboratory 1. Significant trends toward delayed development, as measured by the median developmental stage obtained on day 21 of exposure, were detected at the highest perchlorate dose (500 µg/l) in laboratory 4. By contrast, significant effects of a range of perchlorate doses (125, 250 and 500 µg/l) were detected in laboratory 3. Suppression of hind limb growth by perchlorate at test termination was observed in specimens from laboratories 3 and 4. Mean HLL at test conclusion was significantly reduced relative to controls at 250 and 500 µg/l perchlorate in laboratories 3 and 4. In the other three laboratories, tadpole development (developmental stage or HLL) was not significantly affected by perchlorate exposure.

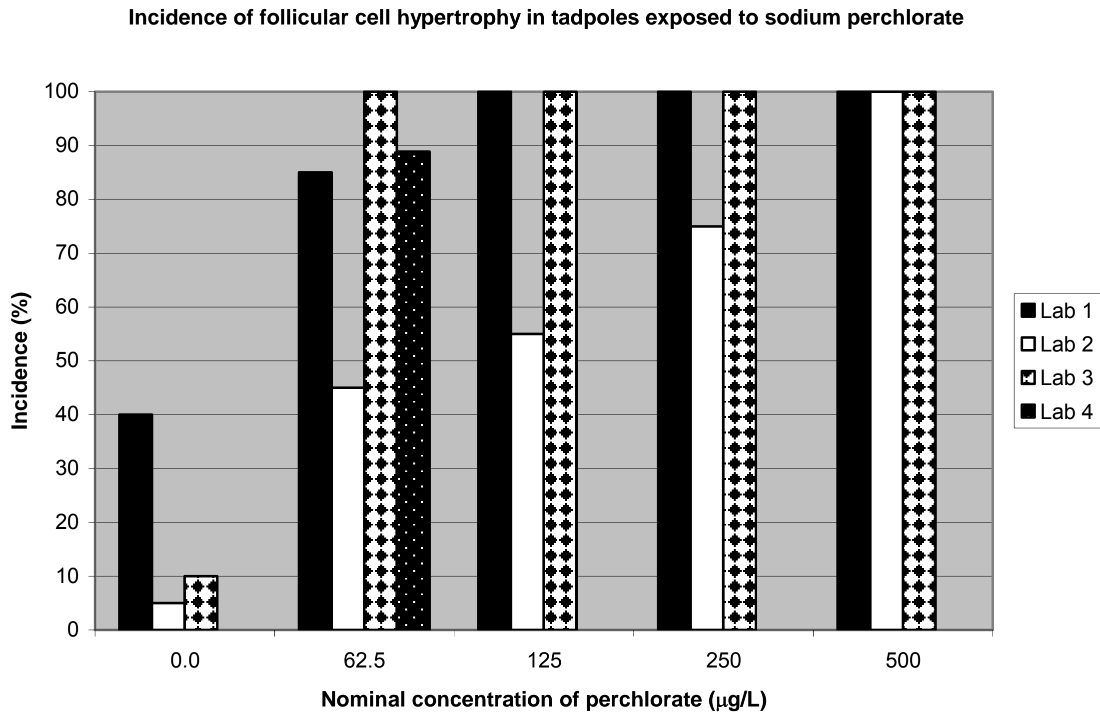
Thyroid Gland Histology

In four of the five laboratories, histological analyses on thyroid glands were conducted. In each laboratory, perchlorate caused alterations in thyroid histology. A semi-quantitative histological analysis was performed using a 3-scale severity grading approach (5-scale scheme was used by laboratory 3) to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. In addition, quantitative techniques were used by laboratories 1 and 2 to evaluate the impact of perchlorate exposure on epithelial cell height, thyroid volume, and thyroid cross section area. The natures of the alterations caused by perchlorate exposure at test termination were consistent among the laboratories and included concentration-dependent decreases in colloid content, increases in overall thyroid gland size, and hypertrophic and hyperplastic changes in the follicular epithelium. Graphs depicting incidence of the core criteria are presented in **Figure 5-1** (note – laboratory 4 did not evaluate thyroid glands from the 125 µg/L dose group). Partial colloid depletion, glandular hypertrophy, and follicular cell hypertrophy were noted with each perchlorate concentration tested from each of the four laboratories. Colloid depletion, thyroid gland hypertrophy, and changes in follicular epithelium were generally mild to moderate at the lowest concentration (62.5 µg/L), but moderate to severe at the highest concentration (500 µg/L). Overall, the morphometric data were consistent with the results from the qualitative characterization and the severity grading approach demonstrating significant increases in epithelial cell height (follicular cell hypertrophy), total glandular volume and maximum cross section area (thyroid gland hypertrophy).

Of the measurements conducted, thyroid gland histology was the most concentration-sensitive endpoint analyzed, and the most reproducible among the participating laboratories.


Incidence of glandular hypertrophy in tadpoles exposed to sodium perchlorate
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0.0
62.5
125
250
500
Nominal dose of perchlorate (µg/L)
Lab 1
Lab 2
Lab 3
Lab 4


Incidence of glandular atrophy in tadpoles exposed to sodium perchlorate
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0.0
62.5
125
250
500
Nominal concentration of perchlorate (µg/L)
Lab 1
Lab 2



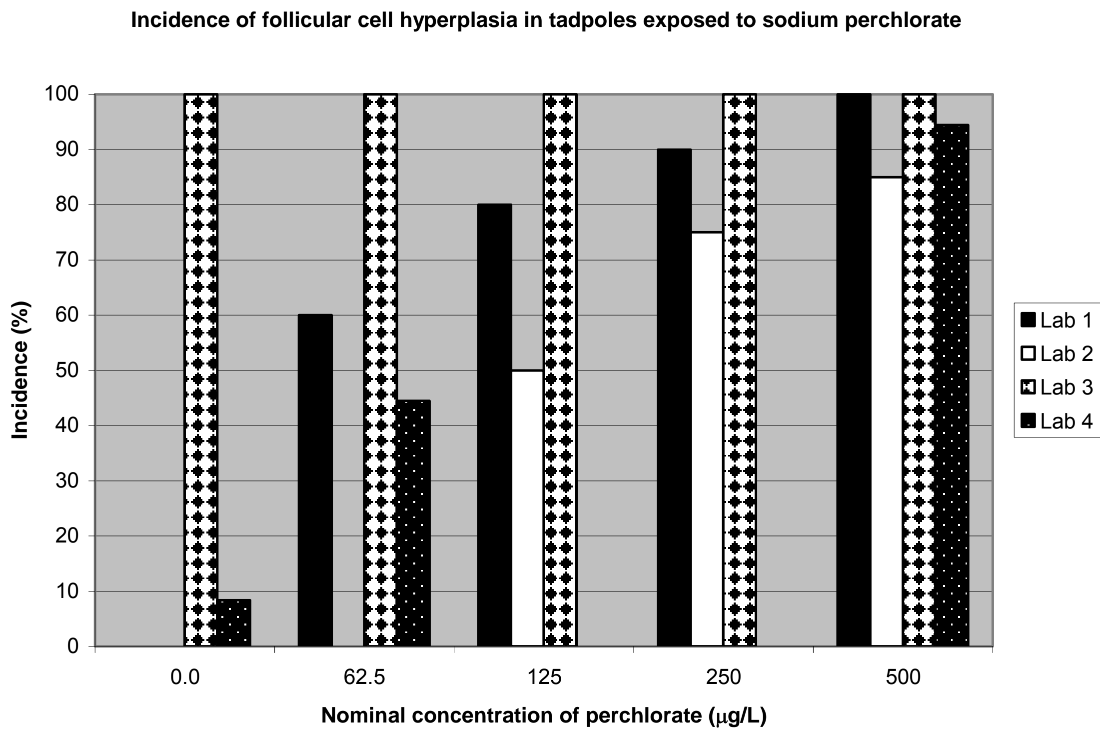


Figure 5-1. Incidence of core histopathological lesions associated with perchlorate exposure from the OECD Phase 2 study. Please refer to the OECD Phase 2 report for details pertaining to the analyses, including further information regarding numbers of animals evaluated, severity grades and on secondary diagnostic criteria.

Overall Laboratory Comparisons – Perchlorate

Using the data interpretation criteria discussed in section 3.6, the overall outcome of the assay when challenged with perchlorate was evaluated across laboratories and is presented in **Table 5-3**.

Table 5-3. Overall assay results - perchlorate.

Criterion		Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Overt Toxicity	Mortality	-	-	-	-	-
	Hemorrhagic lesions	-	-	-	-	-
	Edema	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-
	Lethargy	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-
Advanced development		-	-	-	-	-
Delayed development	Developmental stage (d7 or 21)	-	-	↓	↓	-
	Wet Weight (d7 or 21)	↑	-	↑	↑	-
	SVL/WBL (d7 or 21)	↑	-	↑	↑	↑
	HHL (d7 or 21)	-	-	↓	↓	-
Thyroid histopathology		+	+	+	+	ND
Overall assay result:		+	+	+	+	-

ND = Not done

Results from four of the five laboratories (80%) that tested perchlorate indicate that the assay successfully detected this thyroid antagonist. None of the labs reported evidence of overt toxicity, hence the developmental delays observed were determined to be associated with thyroid activity. Developmental delay was not detected in laboratory 2 using morphological endpoints, however thyroid histopathology from this laboratory revealed a strong response consistent with thyroid antagonism, and therefore was deemed to have a positive assay outcome. In laboratory 5, only SVL was statistically impacted by perchlorate exposure, and no histopathology was performed. Due to the lack of a strong developmental delay response and no histological data, the overall outcome in this laboratory was considered negative.

5.4 Summary of Results – T4

Four of the participating laboratories conducted exposure studies using thyroxine (laboratories 1, 2, 3, and 5). The nominal concentrations of T4 used in each laboratory were 0, 0.25, 0.5, 1.0, and 2.0 µg/l. Summary data for the gross morphological endpoints are reported in **Table 5-4**. Each of the laboratories found that exposure to T4 caused a dose-dependent acceleration of metamorphosis, consistent with the well known effects of thyroid hormone on amphibian metamorphosis. Significant differences in developmental stage caused by T4 were evident only at test termination, while significant increases in mean HLL were observed on study day 7 and at test termination. At test termination T4-treated animals exhibited significantly reduced WBL, SVL, and wet weight, which is consistent with accelerated metamorphosis. No signs of overt toxicity were recorded. Effects of T4 treatment on thyroid gland histology were variable, and histological

endpoints did not present an obvious concentration-response pattern. T4 exposure resulted in atrophy of the thyroid gland at lower T4 concentrations but hypertrophy at greater T4 concentrations.

Test performance for the five experiments with T4 was generally acceptable as discussed in the following paragraphs. Variability in control stage development and lower growth rates compared to other labs was noted in results from laboratory 3. As discussed previously, lower food availability may have been the underlying cause for these results. Measured test concentrations of T4 were reasonably close to nominal concentrations in laboratory 2. However, technical problems associated with measurements of T4 were encountered in the other laboratories preventing verification of the nominal test concentrations.

Mortality

No mortality was observed in any treatment group in the test performed in laboratory 2 and treatment-related mortality was low (< 5%) in tests performed in laboratories 1 and 3. Laboratory 5 reported that all cases of mortality were due to handling errors.

Growth

At test termination, all four laboratories reported a reduction in size and weight at the higher T4 test concentrations. Laboratory 3 reported significantly reduced mean values for WBL, SVL and wet weight at all but the lowest T4 concentration. Overall, WBL, SVL and wet weight were similarly sensitive in detecting dose-dependent effects of T4 in all tests.

Table 5-4. Summary data for gross morphological endpoints from the OECD Phase 2 T4 studies.

	Mean Whole Body Length								Mean Snout Vent Length								Mean Wet Weight							
	Day 7				Day 21				Day 7				Day 21				Day 7				Day 21			
Lab	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Dose (µg/L) 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-
1.0	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	+	-	-	+	-	-	-	+	+
2.0	-	-	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	-	+	+	+	+	+	+

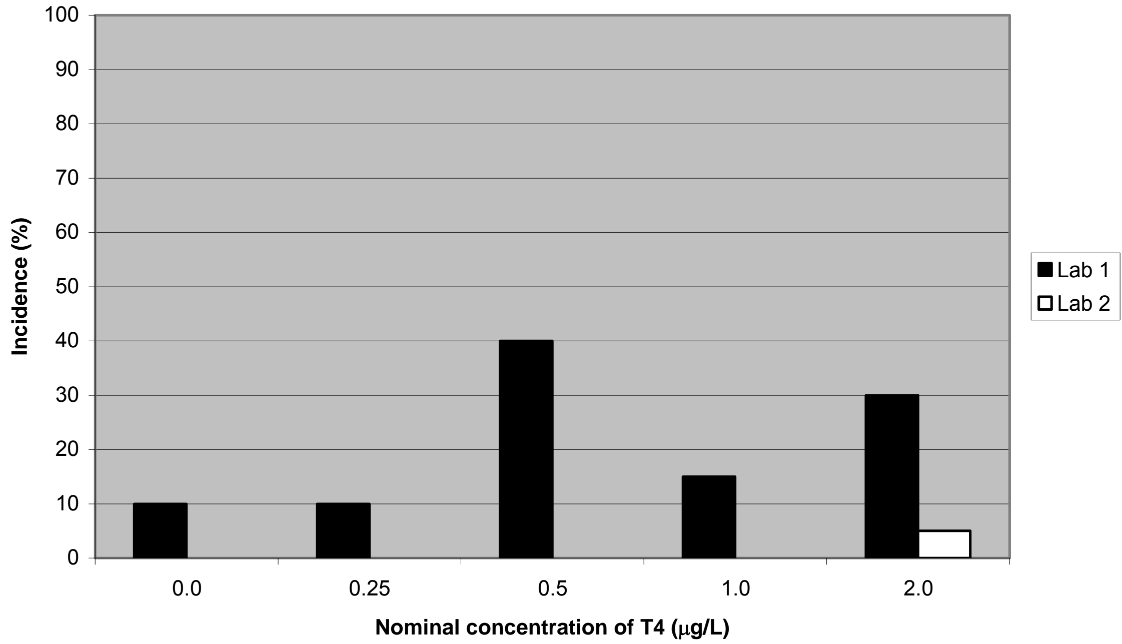
	Mean Hind Limb Length								Median Developmental Stage							
	Day 7				Day 21				Day 7				Day 21			
Lab	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Dose (µg/L) 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.25	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.5	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.0	+	+	-	+	+	-	+	+	-	-	-	+	+	-	+	+
2.0	+	+	-	+	+	+	+	+	-	-	-	+	+	+	+	+

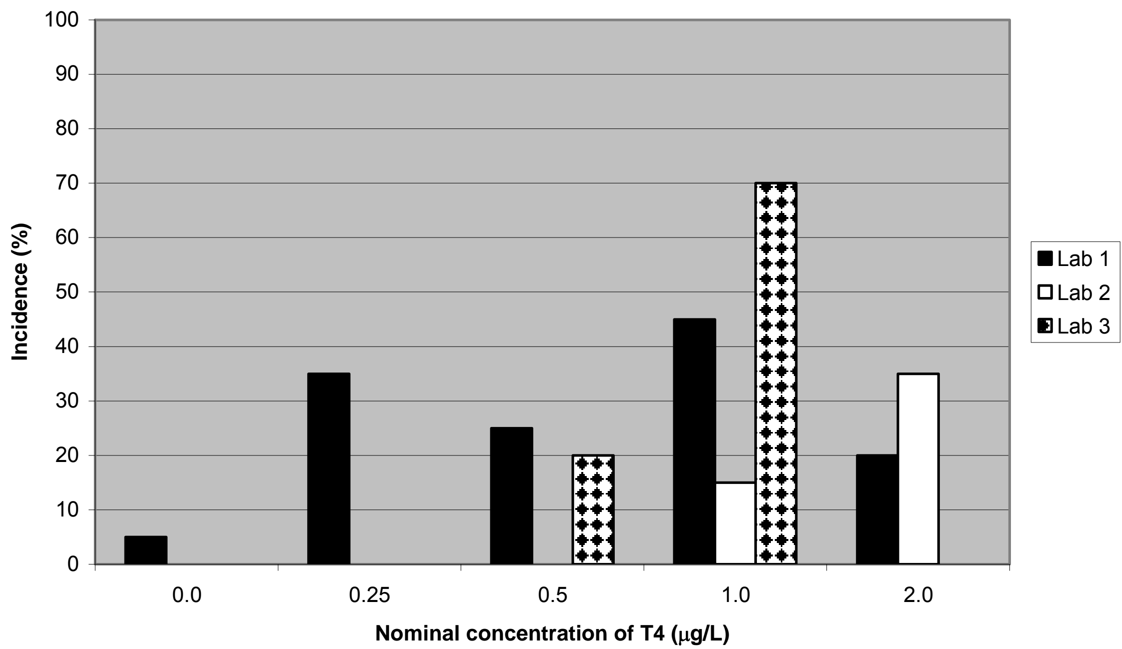
+ indicates statistically significant differences from controls ($p<0.05$).

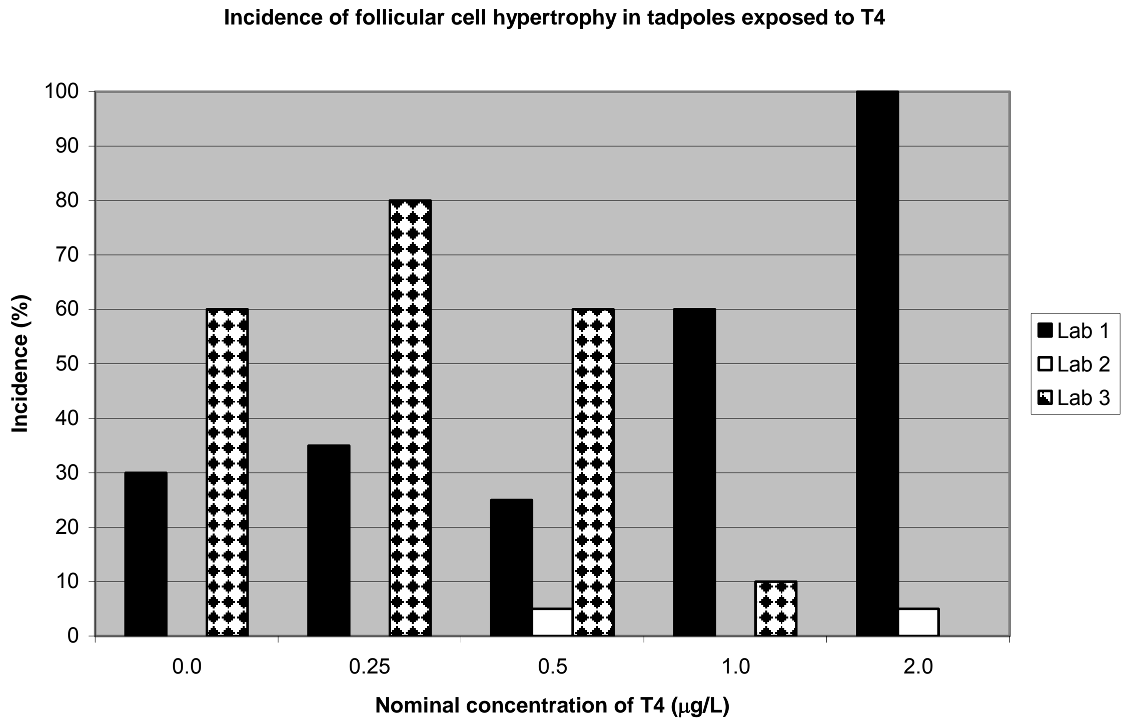
Development

T4 treatment caused a consistent acceleration of tadpole development in all experiments. Developmental stage and HLL showed differential sensitivity to detect developmental acceleration depending on the observation time point. Significant increases in HLL were detected for all four T4 concentrations (0.25 – 2.0 µg/L) on study day 7 in laboratories 1 and 2. However, on day 7 there was no significant effect of T4 exposure on median developmental stage in either laboratory. The effects pattern observed on day 7 in laboratory 5 was somewhat different as the two highest T4 concentrations (1.0 and 2.0 µg/L) caused significant increases in HLL and significant increases in median stages.

HLL measurements were less sensitive in detecting developmental acceleration by T4 at day 21. Significantly increased mean HLL values were detected after treatment of tadpoles with 1.0 and 2.0 µg/L T4, but lower T4 test concentrations did not cause significant effects on HLL at day 21. Measurement of developmental stage at day 21 showed marked acceleration of development at 2.0 µg/L T4 in each laboratory, and at 1.0 µg/l in laboratories 1 and 3. Overall, HLL measurements and stage determination on day 21 were equally sensitive in detecting T4-related acceleration of morphological development.


Incidence of glandular hypertrophy in tadpoles exposed to T4
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0.0
0.25
0.5
1.0
2.0
Nominal concentration of T4 (µg/L)
Lab 1
Lab 2


Incidence of glandular atrophy in tadpoles exposed to T4
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0.0
0.25
0.5
1.0
2.0
Nominal concentration of T4 (µg/L)
Lab 1
Lab 2
Lab 3



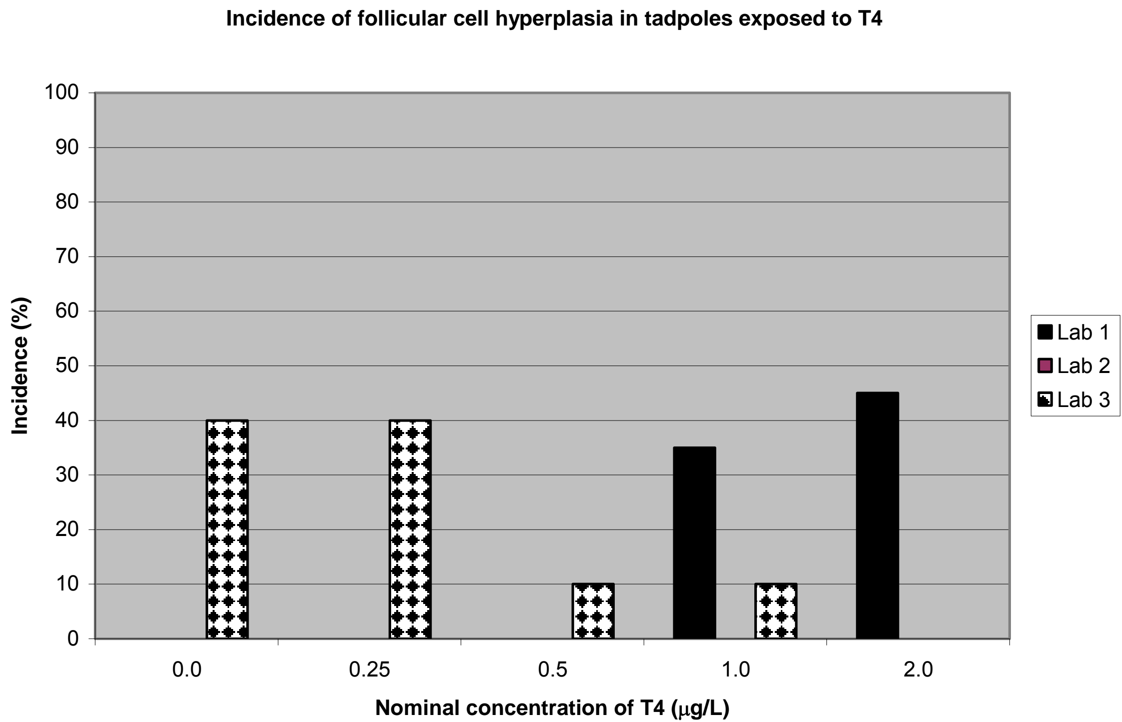


Figure 5-2. Incidence of core histopathological lesions associated with T4 exposure from the OECD Phase 2 study. Please refer to the OECD Phase 2 report for details pertaining to the analyses, including further information regarding numbers of animals evaluated, severity grades and on secondary diagnostic criteria.

Thyroid Gland Histology

Exposure to T4 caused only moderate changes on thyroid gland histology. See **Figure 5-2** for incidence information for the core diagnostic criteria. Variable effects on thyroid gland size were observed in laboratory 1. Mild to moderate glandular hypertrophy was noted in a number of tadpoles exposed to 0.5 µg/L and 2.0 µg/L. In contrast, mild glandular atrophy was observed in tadpoles exposed to 0.25 and 2.0 µg/L T4. Neither of these findings was deemed to be concentration-related. Mild glandular atrophy was also observed in 7 of 10 tadpoles from the 1.0 µg/L T4 treatment in laboratory 3, but not in the other treatments. No changes in glandular size were observed by laboratory 2. This finding was further confirmed by morphometric analyses.

The effects of T4 on the morphology of follicular cells were also different amongst the studies conducted. In laboratory 1, histological alterations were mainly detected at the two highest T4 concentrations of 1.0 and 2.0 µg/L and included mild to moderate follicular cell hypertrophy, mild colloid depletion and a slight increase in the incidence of mild follicular cell hyperplasia. Measurement of epithelial cell height and thyroid volume by laboratory 1 confirmed the results from the severity grading approach, showing significant increases in epithelial cell height (follicular cell hypertrophy) at 1.0 and 2.0 µg/L T4. However, no statistically significant differences in total glandular volume (thyroid gland hypertrophy) were observed between treatments. It should be noted that the thyroid hypertrophy observed at 1.0 and 2.0 µg/L T4 in laboratory 1 was detected in tadpoles at more advanced stages of development compared to the control animals which provided the reference tissues. In laboratory 2, effects on the epithelial cell layer were not detected. Laboratory 3 reported a decreased incidence of mildly hypertrophic follicular cells at the higher T4 concentrations. It should be noted that these results were not compared to the changes in thyroid gland histology associated with the normal course of development, only to concomitant control animals.

Overall Laboratory Comparisons – T4

Using the data interpretation criteria discussed in section 3.6, the overall outcome of the assay when challenged with T4 was evaluated across laboratories and is presented in **Table 5-5**. Results from all four laboratories (100%) indicate that the AMA consistently detects a strong thyroid agonist. No labs reported evidence of overt toxicities. All labs reported significantly advanced development. As indicated in the data interpretation section (section 3.6), accelerated development can only be induced through thyroid-mediated mechanisms, hence all assay outcomes were deemed to be positive. Thyroid histopathology was inconsistent between the three labs that collected information on this endpoint, however the strong developmental response was deemed to be sufficient to conclude that the assays successfully detected T4.

Table 5-5. Overall assay results in Phase 2 - T4.

Criterion		Lab 1	Lab 2	Lab 3	Lab 5
Overt Toxicity	Mortality	-	-	-	-
	Hemorrhagic lesions	-	-	-	-
	Edema	-	-	-	-
	Abnormal behavior	-	-	-	-
	Lethargy	-	-	-	-
	Reduced food consumption	-	-	-	-
Advanced Development	Developmental stage (d7 or 21)	↑	↑	↑	↑
	Wet weight (d7 or 21)	↓	↓	↓	↓
	SVL/WBL (d7 or 21)	↓	↓	↓	↓
	HHL (d7 or 21)	↑	↑	↑	↑
Thyroid histopathology		+	-	E	ND
Overall assay result:		+	+	+	+

ND = not done
E = equivocal

5.5 Summary of Results – IOP

During the validation Phase 2 study, a total of five experiments using iopanoic acid (IOP) were performed in five different laboratories (lab 1, lab 2, lab 3, lab 5, and lab 6). The nominal aqueous concentrations for IOP were 0, 0.75, 1.5, 3.0, and 6.0 mg/L. The experiment in lab 6 included an additional dose of 0.375 mg/L IOP. Summary data for the gross morphological endpoints are reported in **Table 5-6**. Analysis of WBL, SVL, wet weight and mortality rates indicated slight effects of IOP treatment on tadpole growth and survival in several laboratories. Marked growth retardation was determined to be the result of overt toxicity by IOP treatment in one laboratory. Overall, IOP treatment caused a decrease in HLL and asynchronous morphological development. Asynchronous development refers to observations that individual tadpoles exhibited morphological landmarks that corresponded to different developmental stages according to the criteria of Nieuwkoop and Faber (23). Thus, significant difficulty was encountered in assigning specific developmental stages to many of the IOP-exposed animals. Histological assessment of thyroid tissue after IOP exposure was performed in four labs. In general, IOP treatment induced mild alterations in thyroid gland histology, including mild to moderate glandular hypertrophy which was associated with mild to moderate follicular cell hypertrophy.

An initial assessment of the test performance among the five laboratories indicated that results from laboratory 3 were marked by higher variability in development and growth in control animals compared with other laboratories. The lower food availability in this lab may have been the underlying cause for this deviation from the other laboratories. Overall, the measured test concentrations were close to nominal target concentrations. For the 0.75 mg/L dose group, labs maintained actual concentrations between 78-113% of nominal; for the 1.5 mg/L dose group, labs maintained concentrations between 92-111% of nominal; for the 3.0 mg/L dose group, labs maintained concentrations between 91-108% of nominal; and for the 6.0 mg/L dose group, labs maintained concentrations between 91-107% of nominal.

Mortality

No mortality was observed in any treatment group in the test performed in laboratory 5. Mortality was low (< 5%) in the test performed in laboratory 1. Slightly greater mortality (≥ 5%) at the highest IOP test concentration was observed in laboratory 2 (6.25%), laboratory 3 (5%), and laboratory 6 (15%). Laboratory 6 experienced mortality in all treatment groups, and reported higher mortality (10%) in the 1.5 mg/L IOP treatment group.

Growth

IOP treatment caused no significant changes in body size measurements at day 7 of treatment in any of the laboratories except lab 3. In this test, all IOP concentrations caused a marked decrease in tadpole body size with significant effects observed for all of the three endpoints WBL, SVL, and wet weight at test termination. When growth-related endpoints were analyzed at test termination, every laboratory reported a significant reduction in tadpole body size owing to the IOP treatment. In three IOP tests (laboratories 1, 2, and 6), significantly reduced mean values were detected at all tested IOP concentrations for all of the growth-related parameters. Less marked effects of IOP were noted in results collected by laboratory 1 where WBL and wet weight were significantly reduced at the two highest IOP concentrations (1.0 and 2.0 µg/L), and SVL was significantly reduced at the highest IOP concentration (2.0 µg/L). Overall, WBL, SVL and wet weight displayed a similar sensitivity to detect treatment-related changes in tadpole growth in response to IOP treatment. An exception to this effects profile for day 21 growth-related parameters occurred in laboratory 3. In this test, no effects of IOP treatment were detectable for WBL and SVL, and only the highest IOP concentration (6.0 mg/l) caused a significant reduction in tadpole wet weight.

Table 5-6. Summary data for gross morphological endpoints from the OECD Phase 2 IOP studies.

	Mean Whole Body Length										Mean Snout Vent Length									
	Day 7					Day 21					Day 7					Day 21				
Dose (µg/L) / Lab	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.75	-	-	+	-	-	+	+	-	-	+	-	-	+	-	-	+	+	-	-	+
1.5	-	-	+	-	-	+	+	-	-	+	-	-	+	-	-	+	+	-	-	+
3.0	-	-	+	-	-	+	+	-	+	+	-	-	+	-	-	+	+	-	-	+
6.0	-	-	+	-	-	+	+	-	+	+	-	-	+	-	-	+	+	-	+	+

	Mean Wet Weight										Mean Hind Limb Length									
	Day 7					Day 21					Day 7					Day 21				
Dose (µg/L) / Lab	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.75	-	-	+	-	-	+	+	-	-	+	-	-	+	-	-	+	+	+	-	+
1.5	-	-	+	-	-	+	+	-	-	+	-	-	+	-	-	+	+	+	-	+
3.0	-	-	+	-	-	+	+	-	+	+	-	-	+	-	-	+	+	+	-	+
6.0	-	-	+	-	-	+	+	+	+	+	-	-	+	-	-	+	+	+	-	+

	Median Developmental Stage									
	Day 7					Day 21				
Dose (µg/L) / Lab	1	2	3	5	6	1	2	3	5	6
0	-	-	-	-	-	-	-	-	-	-
0.75	-	-	-	-	-	-	-	-	-	-
1.5	-	-	-	-	-	-	-	-	-	-
3.0	-	-	-	-	-	-	-	-	-	-
6.0	-	-	-	-	-	-	-	-	-	-

+ indicates statistically significant differences from controls ($p<0.05$). Note – median developmental stage data could not be quantified due to asynchronous development.

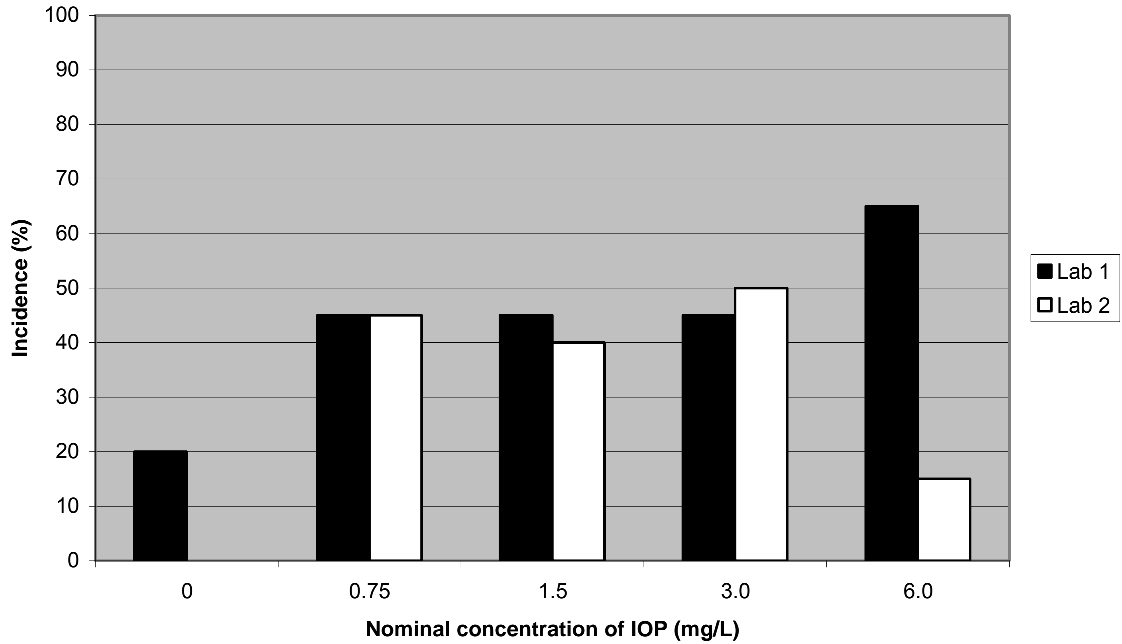
Development

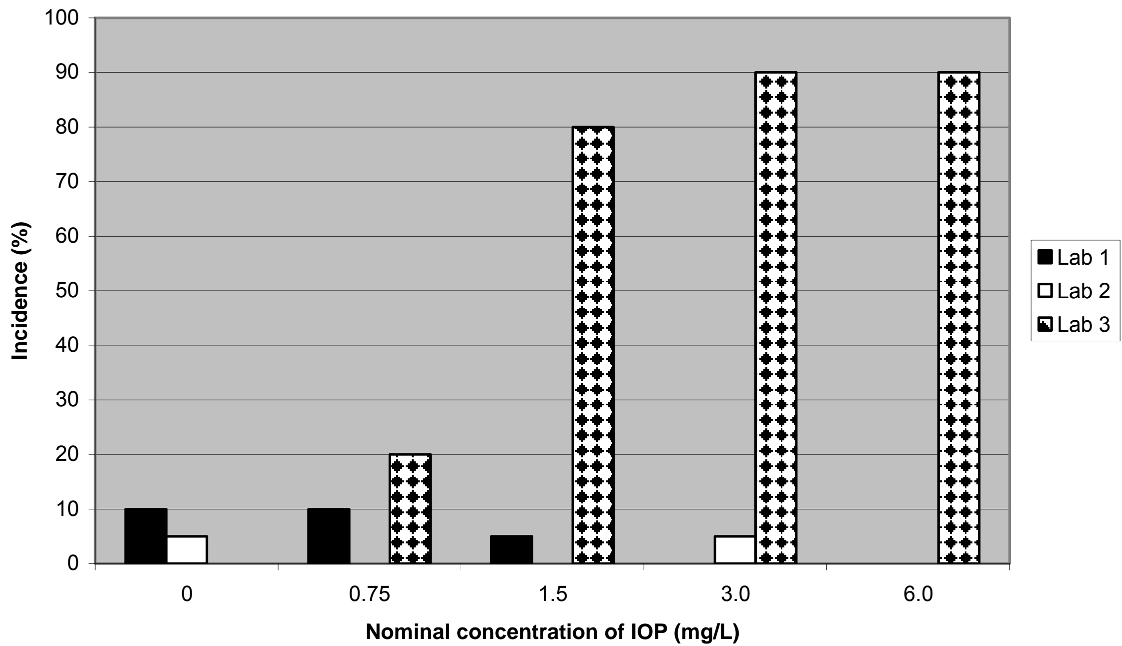
As previously mentioned, asynchronous morphological development was observed in tadpoles exposed to IOP. As a consequence, it was not possible to determine the stage of many test animals. Three laboratories provided a record of the incidence of remarkable cases of asynchronous morphological development. The prevailing type of morphological abnormalities in IOP-treated animals was characterized as retarded hind limb morphogenesis relative to advanced metamorphic changes of head structures. Another observation was that developmental changes in tail tissue appeared more advanced relative to hind limb development. The occurrence of such morphological abnormalities represented a confounding factor for a statistical assessment of median stages and individual laboratories processed the information differently. Thus, statistical analysis of the stage data was not performed. With the exception of results from laboratory 3, the effects of IOP exposure on tadpole development were only detected on study day 21. In addition, morphometric analyses of HLL at test termination revealed reductions in mean HLL in response to all IOP concentrations in three of the four laboratories. No significant effect on HLL was detectable in studies conducted in laboratory 5. The reason for this difference has not been determined. Thus, the effects of

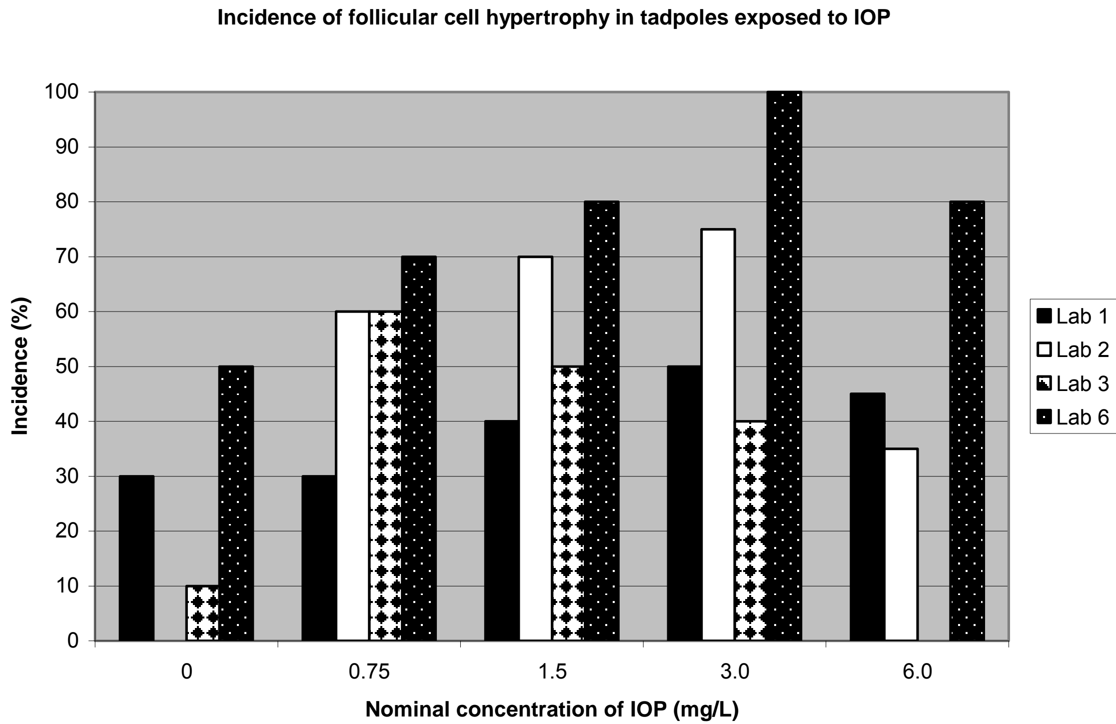
IOP on hind limb growth must be interpreted with caution for the following reasons: 1) for the selected concentration range of IOP, data from HLL measurements on day 21 was not concentration-related in two laboratories (laboratories 1 and 2), and 2) perhaps more important, the interpretation of reduced HLL in IOP treatments was complicated by the finding that concurrent WBL and SVL measurements indicated an overall reduction in tadpole size in the IOP treatments.

Thyroid Gland Histology

Four of the five laboratories reported data on thyroid histology and summary incidence data are reported in **Figure 5-3**. Qualitatively similar changes in thyroid tissue were observed in three experiments with IOP. Two labs reported mild to moderate increases in thyroid gland size, whereas in laboratory 3, glandular atrophy was reported. Mild to moderate follicular cell hypertrophy was reported in all four labs. Follicular cells were considered to be hypertrophic if they were tall columnar cells as opposed to simple cuboidal cells. Tall columnar cells occurred in thyroid glands of controls, but the incidence and follicular cell heights were increased by IOP. Mild hyperplasia of the follicular epithelium was observed in several cases. These findings were not associated with distinct changes in colloid content. Thus, the glands from IOP-treated animals did not show the entire spectrum of histological alterations that is typical for a moderate increase of glandular activity in *X. laevis* tadpoles. A clear, linear concentration-response relationship for the histological criteria could not be established for the range of concentrations tested, and a reverse U-shaped curve was demonstrated for follicular cell hypertrophy in labs 1, 2 and 4, and less apparently in lab 3. The highest mean severity values for follicular cell hypertrophy were observed in the 3.0 mg/L rather than the 6.0 mg/L IOP treatment. The incidence of response differed markedly between the various IOP treatments. The qualitative severity grading approach suggested a mild response, but quantitatively, significant increases in epithelial cell height (follicular cell hypertrophy), total glandular volume and maximum cross sectional area (thyroid gland hypertrophy) were found. These quantitative data also confirmed the absence of a linear dose-response relationship between IOP treatment levels and histological alterations.


Incidence of glandular hypertrophy in tadpoles exposed to IOP
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0
0.75
1.5
3.0
6.0
Nominal concentration of IOP (mg/L)
Lab 1
Lab 2


Incidence of glandular atrophy in tadpoles exposed to IOP
Incidence (%)
100
90
80
70
60
50
40
30
20
10
0
0
0.75
1.5
3.0
6.0
Nominal concentration of IOP (mg/L)
Lab 1
Lab 2
Lab 3



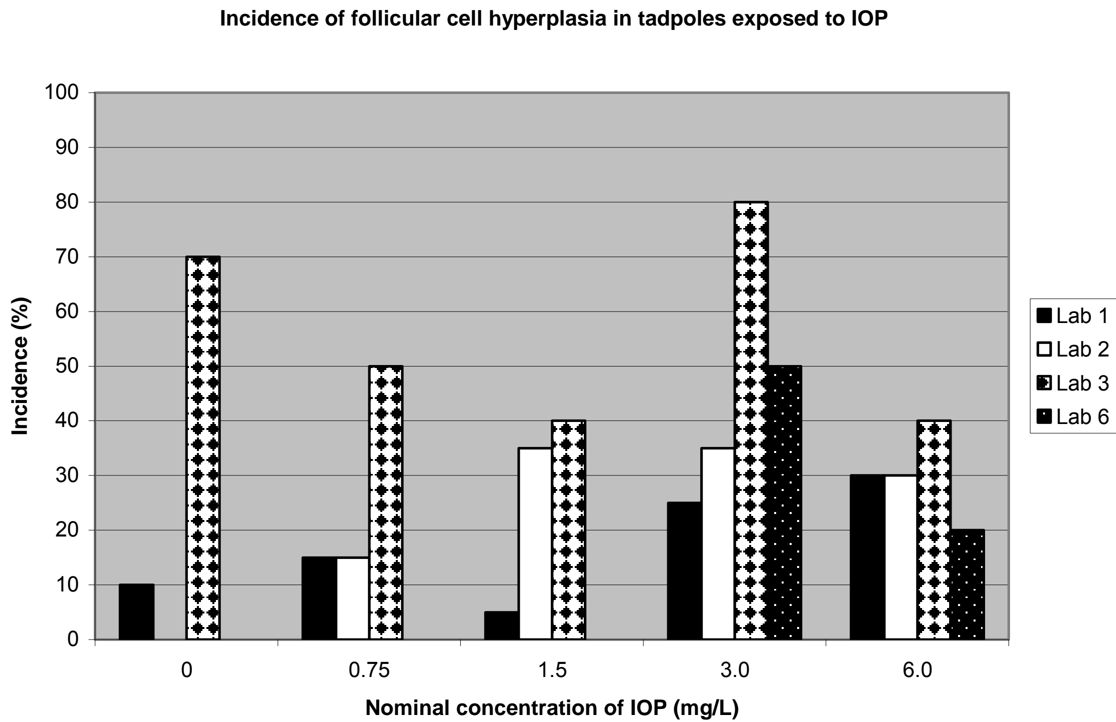


Figure 5-3. Incidence of core histopathological lesions associated with IOP exposure from the OECD Phase 2 study. Please refer to the OECD Phase 2 report for details pertaining to the analyses, including further information on numbers of animals evaluated, severity grades and on secondary diagnostic criteria.

Table 5-7. Overall assay results - IOP.

Criterion		Lab 1	Lab 2	Lab 3	Lab 5	Lab 6
Overt Toxicity	Mortality	-	-	-	-	+
	Hemorrhagic lesions	-	-	-	-	-
	Edema	-	-	-	-	-
	Abnormal behavior	-	-	-	-	-
	Lethargy	-	-	-	-	-
	Reduced food consumption	-	-	-	-	-
Altered Development	Developmental stage (d7 or 21)	+*	+*	+*	+*	-*
	Wet Weight (d7 or 21)	↓	↓	↓	↓	↓
	SVL/WBL (d7 or 21)	↓	↓	↓	↓	↓
	HHL (d7 or 21)	↓	↓	↓	-	↓
Thyroid histopathology		+	+	+	ND	+
Overall assay result:		+	+	+	+	E

ND = not done
*See discussion above on developmental staging for tadpoles exposed to IOP
E = equivocal

Overall Laboratory Comparisons – IOP

Using the data interpretation criteria discussed in section 3.6, the overall outcome of the assay when challenged with IOP was evaluated across laboratories and is presented in **Table 5-7**. Four of the five labs successfully detected developmental effects from IOP exposure. Results from laboratory 6 were considered equivocal because of signs of overt toxicity, regardless of statistically significant changes in wet weight, body length and hind limb length. Labs 1, 2, 3 and 5 reported asynchronous development in tadpoles. This finding, coupled with significant decreases in weight, length and/or hind limb length, was deemed sufficient to justify a positive assay outcome.

5.6 Overall Endpoint Performance

The suite of endpoints used in the OECD Phase 2 study protocol served to obtain information about exposure-related alterations in tadpole growth, metamorphic development, and HPT axis function. Considering the complexity of the biological model utilized, interpretation of the assay results focus on a weight-of-evidence analysis integrating the response profile observed for all endpoints.

Body Length and Weight

Apical morphological endpoints including WBL, SVL, and wet weight were used to assess effects of the test substances on larval growth – a key indicator of metamorphic development. Growth may be affected through thyroid-mediated activity in addition to systemic toxicity, particularly before day 7 of exposure prior to significant decreases in body size during metamorphic climax. Within the first seven days of the Phase 2 study, tadpoles showed substantial growth, doubling in body size and markedly increasing in body weight. The absence of significant effects from the three test compounds on

growth-related parameters at day 7 indicates that systemic toxicity was generally not achieved at the doses tested (with the exception of laboratory 6/IOP test).

Growth assessments at test conclusion (day 21) are meaningful measures of thyroid-mediated activity. For compounds that act through thyroid antagonism, there is potential that these data are confounded by significant reductions in growth parameters during metamorphic climax, and therefore should be interpreted judiciously. During the Phase 2 study, only a small proportion of the test animals, with the exception of animals exposed to T4, developed beyond NF stage 61. Because of this, these data were significant. These data were particularly useful for detecting HPT axis antagonism with IOP, where the determination of developmental stages was confounded by asynchronous development and where thyroid histopathology detected non-linear dose responses to the chemical.

Hind Limb Length

As previously discussed, hind limb differentiation is regulated by TH. Results from the OECD Phase 1 studies indicated that HLL measurements, particularly when performed on day 7, were capable of detecting the activity of TH agonists. This finding was further supported by the results of the T4 experiments in the OECD Phase 2 study where increases in HLL on day 7 provided the most sensitive endpoint. The capacity of HLL measurements on day 7 to detect the antagonistic effects of IOP and perchlorate was generally low for the concentrations used in Phase 2 studies. However, OECD Phase 1 studies indicated that higher doses of PTU reduced HLL following 7 days of exposure.

Determination of HLL on day 21 was useful to detect acceleration of development by T4 in all four studies, and to detect inhibition of metamorphosis by perchlorate in two of five studies. Among the compounds studied in Phase 2, IOP had the most marked effect on hind limb growth. Mean HLL on day 21 was reduced at all IOP concentrations in four of five tests. This finding is consistent with the plausible inhibition of D2 activity in hind limbs by IOP and the reduced generation of local T3 (48-50). However, IOP treatment also caused a slight reduction in whole body length at all tested concentrations, so it is possible that the reduced HLL might merely reflect the reduction in overall body size.

Developmental Stage

Determination of the developmental stage of test animals provided an integrative measure of test substance effects on metamorphosis. Developmental stage data is essential for sound interpretation of all other endpoint results. Determination of developmental stage on day 21 was a sensitive endpoint to detect acceleration of metamorphic development in all T4 experiments. Combined with morphometric analyses of HLL, stage determination was a more sensitive and diagnostic indicator of the agonistic activity of T4 than histopathological examination of the thyroid gland. Developmental stage was less sensitive in detecting antithyroidal activity of perchlorate in the concentration range tested. Only two of five experiments revealed significant delay in metamorphic development due to treatment with the highest perchlorate concentration. Considering

the marked histological changes of thyroid tissue at all tested perchlorate concentrations, these results support the hypothesis that alterations in development in response to antithyroidal agents might only occur when the compensatory capacity of the HPT axis has become exhausted. Stage determination in the IOP studies detected disruption of the sequential remodeling of tadpole tissues eventually leading to asynchronous morphological development. These results warranted the identification of criteria and examination approaches to efficiently identify and describe such morphological changes. Because asynchronous development is a direct result of thyroid system interaction, this finding, in itself, can be considered a “positive” result.

Thyroid Histology

Histopathology of thyroid tissue was relevant and reliable for the detection of anti-thyroidal test compounds. Thyroid histopathology was the most sensitive endpoint to detect the effects of perchlorate on the thyroid gland. The demonstration of mild increases in thyroid gland size and mild to moderate follicular cell hypertrophy following exposure to IOP supported the value of thyroid histopathology to detect substances with extra-thyroidal modes of action. Considering the role of thyroid histopathology as a diagnostic endpoint to detect exposure-related alterations in thyroid system function, possible confounding factors that affect the interpretation of this endpoint must be considered. The results of the OECD Phase 2 studies indicated that differences in the developmental stage of sampled test organisms need to be carefully integrated in the interpretation of histopathological findings. Changes in the histological appearance of thyroid glands occur during normal development, and therefore should be taken into consideration when identifying a lesion or severity grade of a lesion. Culture water characteristics, such as iodide content, as well as background concentrations of goitrogenic substances must be considered. A review of results from experiments during OECD Phase 1 and Phase 2 suggested that the histological characteristics of thyroid glands in *X. laevis* tadpoles indicated that distinct changes in thyroid histology during normal development can differ markedly across experiments. Therefore, a study is currently underway to evaluate the tissues from the Phase 2 studies in relation to developmental stages to determine if stage-matching is necessary. This is, however, confounded by the IOP study where staging of some organisms was impossible. There is potential that two data interpretation approaches are necessary to fully appreciate histological alterations of the thyroid gland in amphibians going through metamorphosis – one that is stage-dependent and one that is stage-independent.

5.7 Summary of the OECD Phase 1 and Phase 2 study results

Combining the results from the Phase 1 and Phase 2 OECD studies, the study leads derived the following conclusions regarding the capacity of the amphibian metamorphosis assay to detect thyroid system disruption.

- *Inhibitors of thyroid synthesis are detected most sensitively by thyroid histopathology of day 21 tissue samples. This was true for inhibitors of iodide uptake (e.g., PER) and inhibitors of iodide organification (e.g., PTU). By using*

improved guidance protocols for tissue processing and histological assessment of tissue sections, concentration-response relationships could be clearly demonstrated for core histological parameters.

- *TH agonists are most sensitively detected by demonstration of accelerated morphological development. Morphometric analyses of hind limb length on day 7 appear to provide the most sensitive endpoint to detect agonist action. Stage determination on day 21 provided a sensitive and robust endpoint to confirm developmental acceleration. Notably, weak agonistic activities were not reliably detected by thyroid histopathology of day 21 tissue samples.*
- *Modulators of iodothyronine deiodinase activity (e.g., IOP) and peripheral action of TH are detected by unique effects on morphological development (developmental stage). Consistent with a tissue-specific role of different iodothyronine deiodinases, IOP caused asynchronous morphological development characterized by retarded hind limb development in the presence of advanced developmental remodeling of craniofacial and tail structures. Thyroid histopathology of day 21 tissue samples also served as a sensitive endpoint in the IOP studies.*

The study leads also concluded that:

- *These conclusions are in line with the current concept of thyroid system control of amphibian metamorphosis supporting the validity of the use of this biological model for the purpose of developing a testing tool for thyroid system disruption.*

6 OECD PHASE 3 STUDY – NEGATIVE AND WEAK POSITIVE

The AMA is intended to yield data that can be interpreted as either positive or negative to determine the necessity and manner in which to conduct Tier 2 tests. To assess the sensitivity of the AMA, and to evaluate its ability to distinguish chemicals with HPT-axis-mediated effects from those that are active in other endocrine systems, a third phase of validation was undertaken through the OECD. The chemicals used for this exercise included 17β-estradiol (E2 CAS: 50-28-2), a potent endocrine active compound that acts via signalling pathways not directly related to the thyroid system, and benzophenone-2 (BP-2 CAS: 131-55-5), a UV filter thought to have weak effects on the thyroid system based on previous studies performed on rats (51-53). The full Phase-3 report can be found in attachment D, however it should be noted that this report is still in draft format and its contents and conclusions are subject to change.

6.1 Study Design

The testing protocol used during Phase-3 was identical to the test protocol used in the Phase-2 studies, with the exception of a more standardized feeding scheme which is outlined in the current method. Two laboratories independently conducted tests on BP-2, and one laboratory evaluated E2. Four doses and a control were evaluated for each

chemical as follows: BP-2: 0.75 mg/L, 1.5 mg/L, 3.0 mg/L and 6.0 mg/L; E2: 0.08 µ/L, 0.4 µ /L, 2.0 µ /L and 10.0 µ /L.

6.2 Overview of results – E2

A summary of results for the e2 study can be found in **Table 6-1**.

Table 6-1. Overall assay results - E2.

Criterion		Lab 1
Overt Toxicity	Mortality	-
	Hemorrhagic lesions	-
	Edema	-
	Abnormal behavior	-
	Lethargy	-
	Reduced food consumption	-
Altered Development	Developmental stage (d7 or 21)	-
	Wet Weight (d7 or 21)	↑
	SVL/WBL (d7 or 21)	↑
	HHL (d7 or 21)	↓
	Thyroid histopathology	-
	Overall assay result:	-

No mortality or gross morphological or behavioral abnormalities indicating systemic toxicity were observed in this experiment. On study day 7, analysis of growth-related parameters (e.g., WBL, SVL, wet weight) and developmental parameters (e.g., HLL and developmental stage) did not show any difference between control and E2 treatment groups. On study day 21, size and weight of E2-treated tadpoles was increased at the two highest E2 concentrations. The interpretation of these results is, however, confounded by the fact that a large number of tadpoles in all treatment groups already shown development to climax stages (which is associated with weight loss and a reduction in body size). In *X. laevis* tadpoles, body size and weight are highest in tadpoles at developmental stages 59/60 and body size and weight decrease at more advanced stages. In the E2 experiment, the percentage of tadpoles at stages >60 were 50%, 53%, 40%, 30% and 18% for the control and the 0.08, 0.4, 2.0 and 10 µg/l E2 treatments, respectively. The higher mean values of WBL, SVL and wet weight at high E2 concentrations are, therefore, likely the result of the reduced number of climax stage tadpoles observed in high E2 treatments on day 21. Overall, the high number of climax stage tadpoles (stage >60) observed throughout all treatment groups on day 21 of the E2 experiment prevents a sound analysis of possible E2 effects on growth using data from day 21 growth-related measurements.

Results from stage determination on day 21 did not reveal significant effects of E2 on developmental stage, but HLL measurements indicated a slight reduction in hind limb growth at the two highest E2 concentrations. A trend toward slight developmental delay, as suggested by the reduced number of tadpoles reaching metamorphic climax in the highest concentration, is consistent with a general toxic response.

Analysis of thyroid tissue from the various E2 treatments did not reveal remarkable effects of E2 on thyroid size. The follicular architecture of the thyroid glands from E2-treated tadpoles also did not deviate from the control group. Minor differences between control and E2-treated animals were only noted with regard to the incidence and severity of follicular cell hypertrophy. At the two highest E2 concentrations, stimulation of follicular epithelium appeared less prominent compared to the control and the lower E2 concentrations. At 10 µg/l E2, most thyroid follicles were composed of cuboidal epithelial cells and the mild to moderate follicular cell hypertrophy observed in the majority of control samples was not as prevalent in this treatment group. A semi-quantitative analysis of the histological alterations was performed using a severity grading approach to assess the incidence and severity of three core diagnostic parameters including thyroid gland size, follicular cell hypertrophy and follicular cell hyperplasia. No significant differences for gland size or follicular cell hyperplasia were detected. Severity scores for follicular cell hypertrophy tended to decrease at high E2 concentrations but no statistically significant differences were detected for mean severity scores (n= 4 replicate tanks) between E2 treatments and the control group. There is potential that the differences in follicular cell size are due to differences in developmental stage of tadpoles evaluated.

Gross morphological assessment of gonads showed a concentration-dependent feminization of tadpoles. Abnormal testicular development was observed at 0.4 µg/L E2 while higher E2 concentrations caused an almost complete male-to-female sex reversal as indicated by the lack of animals with testicular tissue.

In this study, the chemical evaluated had a relatively well-characterized mode of action, and therefore increases in growth parameters were expected. While increased growth, as was apparent here from exposure to E2, can be associated with thyroid toxicity, this endpoint, when taking into consideration developmental stage and thyroid histopathology, was not considered sufficiently conclusive to deem this assay “positive”.

6.3 Overview of results – BP-2

During the validation Phase-3 study of the AMA, two exposure experiments with benzophenone-2 (BP-2) were performed. In both exposure studies, the nominal aqueous concentrations for BP-2 were 0, 0.75, 1.5, 3.0, and 6.0 mg/L. Tadpoles treated with the highest BP-2 concentration showed signs of weak systemic toxicity including growth retardation (lab 1) and slightly reduced survival (lab 2). Results from stage determination and HLL measurements showed BP-2 treatment to cause developmental delay in both experiments, however responses between the labs were inconsistent, with lab 1 experiencing delayed development with decreased weight and length, and lab 2 experiencing delayed development with increased weight and length. Histopathological assessment of thyroid tissue revealed remarkable effects of the highest BP-2 concentration on the thyroid gland in lab 2 providing strong evidence for disruption of the thyroid system by BP-2. The histological effects on the thyroid glands in lab 1 were less remarkable, however a significant increase in the severity of follicular cell hypertrophy

was noted in the highest BP-2 concentration. Gross morphological assessment of gonads showed abnormal testicular development indicative of estrogenic activity of BP-2. A summary of results for the BP-2 study can be found in **Table 6-2**.

Table 6-2. Overall assay results - BP-2.

Criterion		Lab 1	Lab 2
Overt Toxicity	Mortality	-	*
	Hemorrhagic lesions	-	-
	Edema	-	-
	Abnormal behavior	-	-
	Lethargy	-	-
	Reduced food consumption	-	-
Altered Development	Developmental stage (d7 or 21)	↓	↓
	Wet Weight (d7 or 21)	↓	↑
	SVL/WBL (d7 or 21)	↓	↑
	HHL (d7 or 21)	E	↓
	Thyroid histopathology	+**	+
	Overall assay result:	+/-	+

ND = not done
*less than 5% mortality
**the histological response was very mild in lab 1

A comparison of endpoint results between the two labs indicates that BP-2 effects were relatively inconsistent, compared to the consistency found with strong antagonists, however this may be expected given that this chemical is thought to act as a weak positive. The results from lab 2 clearly identify this chemical as thyroid-active given the strong histological responses in the thyroid glands. The results from lab 1 are suggestive of thyroid activity, however these results would likely not provide conclusive evidence alone. Rather, in a circumstance such as this, these results would have to be coupled with results from other assays to determine whether further, Tier 2 testing would be warranted. The degree to which the iodide concentration in the dilution water of lab 1 may have attenuated the response is uncertain and could be evidence that BP-2 imposes a reversible effect on iodine uptake.

7 ADDITIONAL STUDIES SUPPORTING THE VALIDATION OF THE AMA

The original DRP covered the relevant literature up to about 2003. Since its publication, several laboratory studies have been published which provide additional support to the validity of the proposed test guideline. All of these studies were *ad hoc*, in that they were not part of a coordinated research and validation effort, like those that emanated from the OECD validation studies previously presented. Consequently, the collection of recent open literature studies do not necessarily conform to the parameters outlined in the AMA protocol. The literature considered in this chapter is limited to laboratory studies that utilize chemicals with known thyroidal activity and include developmental analyses and/or thyroid gland histology. Field studies and studies that focused on molecular endpoints were specifically excluded.

7.1 Studies using *X. laevis*

Several studies have been published which include methods that closely approximate the AMA method, but differ significantly in one or more of the AMA parameters (**Table 7-1**). These studies utilize stage-specific aqueous exposure approaches which include some or all of the pre-metamorphic and/or pro-metamorphic periods. They also include thyroid gland histology as a diagnostic endpoint, as well as morphological assessments of development. The test chemicals represented in these studies are all well known inhibitors of thyroid hormone synthesis and include: ETU, 6-PTU, perchlorate, and methimazole. Test duration ranged from 12-90 days. All of the studies were able to detect differences in metamorphic development and thyroid gland histology, despite experimental differences.

Chemical	Test Concentration	Initial Stage	Exposure Duration	Development	Thyroid Histology	Reference
ETU	50 mg/L	NF52	12 d	*	*	(54)
Sodium perchlorate	20 mg/L	NF52	12 d	*	*	(54)
ETU	1.0 – 50 mg/L	NF51	<90d	*	*	(55)
6-PTU	1.25 – 20 mg/L	NF51 NF54	14 d 14 d	* *	* *	(41)
Methimazole	6.25 – 100 mg/L	NF51 NF54	14 d 14 d	* *	* *	(41)
Perchlorate	16 -4,000 ug/L 16 – 125 ug/L	NF51 NF54 NF51	14 d 14 d 44 d	* * *	* * not available	(43)

Table 7-1. Summary of tests conducted with *X. laevis* that include analyses of developmental effects and thyroid gland histology. Asterisks indicate that significant effects were observed. (Note: Additional 6-PTU and T4 studies in Degitz, et al, 2005 are already represented in section 4 and were excluded from this table.)

An alternative test method for thyroid disruption using *X. laevis*, which preceded the development of the current AMA, was originally proposed by Opitz *et al* (56). This method utilized NF 48-50 in a 28 day exposure and relied primarily on metamorphic development as an effects endpoint, but did not include thyroid histology. This test was evaluated by six different laboratories using ETU as the test chemical. ETU is a known thyroid hormone synthesis inhibitor in mammalian systems. 6-PTU and T4 were used as positive controls for antagonism and agonism, respectively. A total of ten trials were performed. PTU retarded development and T4 accelerated development in all studies. Furthermore, ETU retarded development in all studies with LOECs ranging from 10 to 50 mg/L among the six laboratories. The responses of other morphometric measurements (whole body length and tail length) to ETU exposure were more variable. This was the first inter-laboratory study to demonstrate that the effects of a thyroid synthesis inhibitor on development were repeatedly observed among several laboratories when following a standardized testing protocol.

7.2 Studies using *Silurana (Xenopus) tropicalis*

Silurana (Xenopus) tropicalis, the West-African clawed frog, is similar to *X. laevis* in that it is a predominantly aquatic species that can be reared and bred in the laboratory using similar methods and conditions. Its utility as a model organism for research is increasing due to its rapid life cycle, smaller size, and diploid genome. Most of the methods that are suitable for *X. laevis* studies can also be used with *S.* (*X.) tropicalis* with minimal modifications.

Two studies on thyroid disruption in *S.* (*X.) tropicalis* have recently been published. These studies examined the effects of 6-PTU using protocols similar to the AMA (57;58). The Carlsson and Norrgren study most closely approximated the AMA by using stage NF51 organisms, a 14 day exposure duration, and thyroid gland histology. Using this approach, 6-PTU retarded metamorphic development at 75 mg/L but resulted in thyroid gland histological changes at the lower exposure concentrations (2 – 5 mg/L), reinforcing the previous observations in *X. laevis* that thyroid gland histology is a much more sensitive endpoint than development. The extent and severity of the histological changes also exhibited a clear dose response relationship.

The Mitsui *et al* (58) study followed the protocol described by Opitz *et al* (56) by using stage NF48 – 50 organisms for test initiation, conducting the exposure for 28 days, and using 6-PTU and T4 as described above. Thyroid gland histology was not included in this study. Consistent with results of the inter-laboratory study described previously with *X. laevis*, 6-PTU inhibited development and T4 accelerated development.

7.3 Discussion

A total of sixteen additional individual studies were identified in which *X laevis* was used as the test species in protocols which approximate the AMA (41;43;56;59). All of these studies used chemicals that were known thyroid antagonists. Despite methodological differences, geographic location, water source, etc., these studies demonstrate internally consistent responses of the assay system. Furthermore, these responses are consistent with the results of the Phase 1 and Phase 2 validation studies conducted through the OECD. As such, these studies provide additional support for the AMA protocol. In addition, three studies were identified in which S. (*X.) tropicalis* was the test species. These, too, were consistent with the *X. laevis* studies, further supporting the concept that the AMA protocol is representative of other amphibian species.

One study not considered above was that of Gutleb *et al.* (60), who proposed an assay protocol whereby developmental synchronization is achieved by pre-exposure to thiourea. They used this protocol to examine the developmental effects of Clophen, PCB77, and sediment extracts. Their results demonstrated developmental delay by Clophen and some of the sediment extracts. However, their protocol did not include any

endpoint diagnostic of thyroid specific effects, so it is unclear as to whether the developmental delay actually was due to thyroid disruption. Furthermore, the use of thiourea as a chemical means of developmental synchronization is problematic because it alters the status of the HPT axis to the extent that the test is conducted on a highly modulated system. Presumably, the developmental arrest achieved by thiourea exposure results in highly upregulated compensatory mechanisms, such as increases in circulating TSH and thyroid follicular cell hypertrophy and hyperplasia. Therefore, the validity of this protocol as a screening method is questionable.

8 DISCUSSION/SUMMARY OF THE VALIDATION OF THE AMA

8.1 Reproducibility of the assay

The reproducibility of the AMA, for screening purposes, has been well demonstrated using several representative thyroid-active chemicals across geographically diverse laboratories. As previously discussed, all three laboratories in the OECD Phase 1 trials detected the accelerating effects on metamorphic development in tadpoles exposed to the thyroid agonist and endogenous hormone T4. Significant acceleration was observed in each laboratory at the highest test concentration but not at the lowest concentration, regardless of methodological differences, which demonstrates the robustness of the assay. Likewise, each of the same laboratories was able to consistently detect the inhibitory effect of the thyroid antagonist PTU using thyroid histopathology. Although histopathology findings were generally consistent in the PTU trials, observations for T4 were less consistent in specific diagnoses across the Phase 1 participating laboratories, but still consistent in detecting glandular pathology, albeit less sensitive than morphological development.

In the OECD Phase 2 trials, good inter-laboratory concordance was demonstrated across the three chemicals tested. Variability in control stage development and lower growth rates observed in one laboratory were attributed to lower food availability, but this did not seem to appreciably affect the overall assay conclusions for the compounds tested. Specifically, thyroid histopathology was the most sensitive endpoint for inhibitors of iodine uptake (perchlorate). TH agonists (T4) are more sensitively detected by morphological development and morphometric analysis, and modulators of iodothyronine deodinase activity (IOP) are detected by asynchronous morphological development.

A comparison of results from the inter-laboratory trial for BP-2 revealed less consistent results than with the strong agonists and antagonists, however both labs detected changes in endpoints that were either suggestive of thyroid activity, or conclusively demonstrated thyroid activity. As with the strong antagonists, thyroid histopathology provided the strongest evidence of thyroid activity, and both labs detected developmental delay. Morphological endpoints such as weight and length, while affected by BP-2, were less conclusive.

8.2 Recommended refinements to the method

As reflected in the test method (attachment A), there are several minor refinements to the protocol that have been recommended by the amphibian expert group based on the analyses of both the Phase 1 and Phase 2 studies. The intention of these enhancements is to further reduce variability of the assay. The following discussion outlines those changes and provides the rationale for the suggestions.

Dietary regime

Following the analyses of the Phase 2 study, it was recognized tadpole/froglet size at the end of the study is influenced by the feeding regime. In one lab, final size of test organisms was significantly smaller than in the other labs. This finding did not significantly impact the ability of the laboratory to detect the effects of T4, perchlorate or IOP. However, it is thought that with less potent compounds, this effect could mask more subtle changes associated with size. Therefore, it is recommended that the feeding regime be changed to account for changing energy needs of the developing tadpole. The recommended feeding regime can be found in the test method.

Water iodine level

There was international participation in the Phase 2 trials conducted for the inter-laboratory component of the validation exercise. Therefore, water sources differed greatly, and the salts and minerals present in the sources also differed. In order for the thyroid gland to synthesize TH, sufficient iodide needs to be available to the larvae through a combination of aqueous and dietary sources. Currently, there are no empirically derived guidelines for minimal iodide concentrations. However, iodide availability may affect the responsiveness of the thyroid system to thyroid active agents and is known to modulate the basal activity of the thyroid gland, an aspect that deserves attention when interpreting the results from thyroid histopathology. Therefore, determination of aqueous iodide concentrations must be considered as an important measurement in any bioassay used for the detection of thyroid active substances. Given this, it is recommended that the minimum iodine concentration in the test water be 0.5 μ/L. If the test water is reconstituted from deionized water, iodine must be added at a minimum concentration of 0.5 μ/L. Any additional supplementation of the test water with iodine or other salts must be noted in the report.

Dose levels

During the Phase 1 and Phase 2 trials, the minimum number of test concentrations used by the participating laboratories was four, with a clean water control, each with four replicates per group. This approach was useful for the validation exercises because it allowed for optimum ability to detect effects between dose groups and with controls. Given that this assay is meant to be a screen, rather than a definitive test, it has been recommended that the minimum number of test concentrations be reduced to three, rather than four, with a clean water control and vehicle control (if necessary). This approach

will reduce animal use and cost of the assay while still allowing for sufficient power to detect effects.

Stage matching for histopathology

It is recognized that the morphology of the thyroid gland, particularly follicular cell height, changes over the course of metamorphosis and is stage-dependent. Examples of the progression of the thyroid gland, over time, can be found in appendix B of the test method. However, during the Phase 1 and Phase 2 trials, tadpoles were randomly selected for histopathology at day 21, not taking into account differences in stages. This has the potential to introduce uncertainty into the histological analysis. For this reason, it is recommended that, when possible, stage matching be performed prior to selection for histopathology as follows. A total of five tadpoles are selected from each of the 16 treatment tanks (80 animals) for histopathological analysis of the thyroid gland. In order to select stage-matched individuals, all larvae are first staged prior to selection and subsequent processing for data collection and preservation. This is necessary because normal divergence in development will result in differential stage distributions within each replicate tank. After the larvae are segregated by stage, the distribution of stages across the test is evaluated. Larvae should then be sampled from the most advanced stage with sufficient number of larvae (n=5) in all replicates. If there are more than five larvae from each replicate, then five are randomly selected. If there are fewer than five larvae, then randomly selected individuals from the next lower developmental stage bin are used to reach a total sample size of five. If stages cannot be determined, as with the experiments performed with IOP, larvae are randomly selected for histological analysis.

Performance criteria

Based on results from the Phase 1 and 2 studies and standards in aquatic toxicology, performance criteria were developed for determining the acceptability of test conditions for data reporting and submission. These criteria were developed to reduce variability in the system, to provide optimum conditions for tadpole metamorphosis, to ensure adequate data sets are available for analysis, and to accommodate the use of static-renewal systems when necessary. **Table 8-1** describes these criteria.

Table 8-1. Performance criteria for the AMA.

Criterion	Acceptable limits
Test concentrations	Maintained at ≤ 20% CV(variability of measured test concentration) over the 21 day test
Mortality in controls	≤10%
Minimum median developmental stage of controls at end of test	57
Dissolved Oxygen	≥ 40% air saturation
pH	pH should be maintained between 6.5-8.5. The inter-replicate/inter-treatment differentials should not exceed 0.5.
Water temperature	22º ± 1ºC The inter-replicate/inter-treatment differentials should not exceed 0.2 ºC
Test concentrations without overt toxicity	≥ 1
Replicate performance	< 2 replicates across the test can be compromised
Special conditions for use of a solvent	If a carrier solvent is used, both a solvent control and water control must be used and results reported
	No statistically significant differences between solvent control and water control groups can be detected with any endpoint
Special conditions for static renewal system	Representative chemical analyses before and after renewal must be reported
	Ammonia levels must be measured immediately prior to renewal
	All water quality parameters listed in Table 1 of the test method must be measured immediately prior to renewal
	Renewal period may not exceed 72 hours

Improved data interpretation criteria

In recognition that some of the endpoints in the AMA, such as growth, could potentially be affected through non-thyroid mediated mechanisms, either through general toxicity or through other endocrine pathways (for example – estradiol could potentially cause a change in weight of tadpoles, as has been demonstrated in rat assays, through non-thyroid mediated pathways), an improved method for data interpretation was devised. Based on the results obtained during the Phase-2 validation experimental work, decision logic (data interpretation criteria) was developed for the AMA to provide logical assistance in the conduct and interpretation of the results of the bioassay (see flow chart in **Figure 8-1**). The decision logic, in essence, weights the endpoints in that advanced development, asynchronous development, and thyroid histopathology are weighted heavily, and developmental stage, growth parameters and hind limb length, parameters that can potentially be affected by general toxicity, are weighted less heavily. While this logic is very similar to the logic used in determining the overall outcomes of the Phase 2 data, it

simplifies distinguishing between generalized toxic effects causing developmental delay or altered growth, and effects due to direct thyroid inhibition.

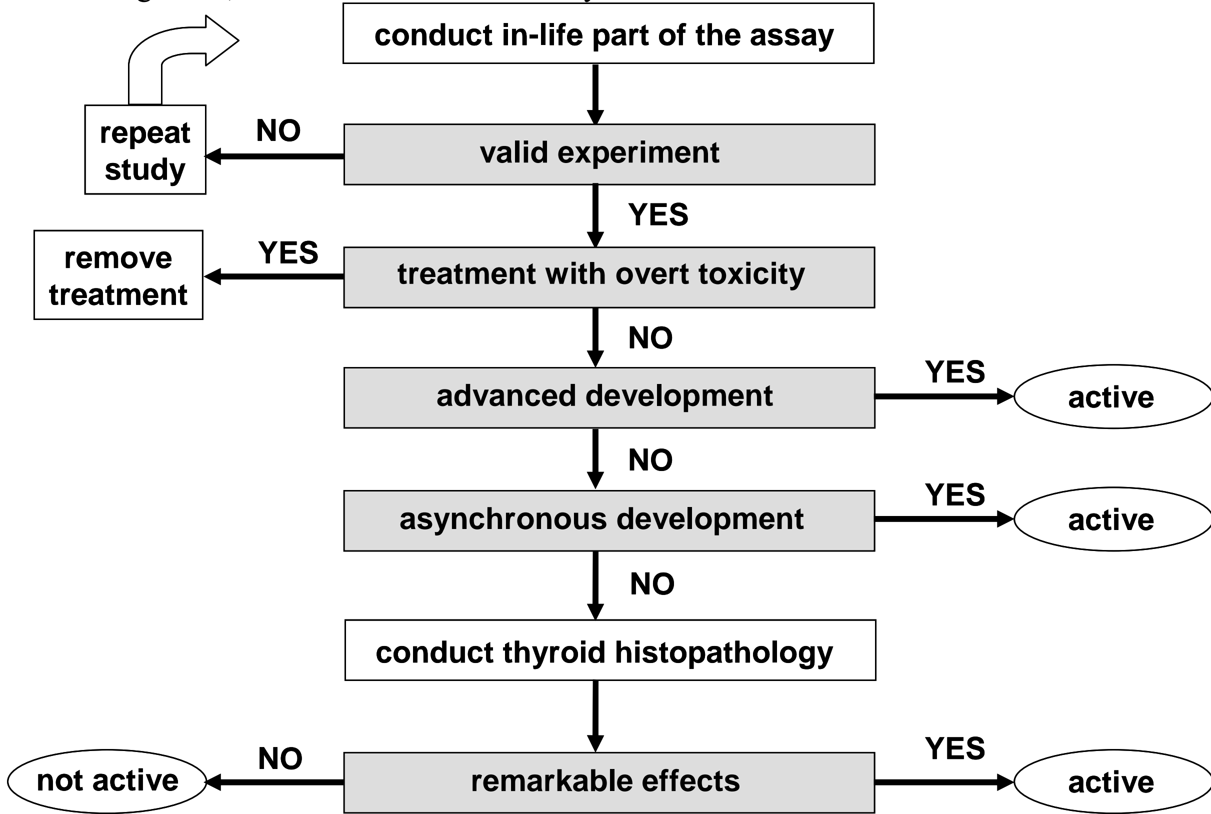


Figure 8-1. Flow chart describing decision logic for the interpretation of results obtained with a test substance in the AMA.

Applying the proposed decision logic for the studies performed in the Phase 1, multi-chemical studies, and the Phase 2 experiments yield the following results (see **Table 8-2**, **Table 8-3**, **Table 8-4**, **Table 8-5**, **Table 8-6**, and **Table 8-7**):

Table 8-2. Overall assay results for PTU from the Phase 1 studies and Contractor studies using the refined data interpretation criteria.

	Stage 51/21 days				Stage 54/14 days			
Criterion	**US**	**GR**	**JP**	**C**	**US**	**GR**	**JP**	**C**
Overt Toxicity	No	No	No	No	No	No	No	No
Advanced Development	No	No	No	No	No	No	No	No
Asynchronous Development	No	No	No	No	No	No	No	No
Remarkable histological effects on thyroid glands	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Overall assay result:	+	+	+	+	+	+	+	+

C = Contractor lab

Table 8-3. Overall assay results for T4 from Phase 1 studies and Contractor lab using the refined data interpretation criteria.

	Stage 51/21 days				Stage 54/14 days			
Criterion	**US**	**GR**	**JP**	**C**	**US**	**GR**	**JP**	**C**
Overt Toxicity	No	No	No	Yes	No	No	No	No
Advanced Development	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Asynchronous Development	No	No	No	No	No	No	No	No
Remarkable histological effects on thryoid glands	Yes	E	No	Yes	Yes	E	Yes	Yes
Overall assay result:	+	+	+	+	+	+	+	+

C = Contractor lab
E = equivocal

Table 8-4. Overall assay results for the multi-chemical study performed by a Contractor for four chemicals using the refined data interpretation criteria.

	Stage 51/21 days				Stage 54/14 days			
Criterion	**Dexamethasone**	**Phenobarbital**	**Methimazole**	**PCN**	**Dexamethasone**	**Phenobarbital**	**Methimazole**	**PCN**
Overt Toxicity	No	No	No	No	No	No	No	No
Advanced Development	No	Yes	No	No	No	No	No	No
Asynchronous Development	No	No	No	No	No	No	No	No
Remarkable histological effects on thryoid glands	Yes	Yes	Yes	No	No	Yes	Yes	No
Overall assay result:	+	+	+	-	-	+	+	-

Table 8-5. Overall assay results for perchlorate from Phase 2 using the refined data interpretation criteria.

Criterion	**Lab 1**	**Lab 2**	**Lab 3**	**Lab 4**	**Lab 5**
Overt Toxicity	No	No	No	No	No
Advanced development	No	No	No	No	No
Asynchronous development	No	No	No	No	No
Remarkable histological effects on the thyroid glands	Yes	Yes	Yes	Yes	ND
Overall assay result:	+	+	+	+	-

ND = not done

Table 8-6. Overall assay results for T4 from the Phase 2 studies using the refined data interpretation criteria.

Criterion	Lab 1	Lab 2	Lab 3	Lab 5
Overt Toxicity	No	No	No	No
Advanced Development	Yes	Yes	Yes	Yes
Asynchronous Development	No	No	No	No
Remarkable histological effects on thyroid glands	Yes	No	E	ND
Overall assay result:	+	+	+	+

ND = not done
E = equivocal

Table 8-7. Overall assay results for IOP from Phase 2 using the refined data interpretation criteria.

Criterion	Lab 1	Lab 2	Lab 3	Lab 5	Lab 6
Overt Toxicity	No	No	No	No	Yes
Advanced Development	No	No	No	No	No
Asynchronous Development	Yes	Yes	Yes	Yes	No
Remarkable histological effects on thyroid glands	Yes	Yes	Yes	ND	Yes
Overall assay result:	+	+	+	+	+

ND = not done

OECD Phase 3 - 17β-estradiol

An assessment of the general test performance indicated that the E2 experiment in Phase 3 should be regarded a valid experiment. The absence of overt toxicity implies that all test concentrations of E2 should be included in the effects assessment. Treatment of tadpoles with E2 did not cause advanced development nor were any signs of asynchronous development observed. Thyroid histopathology did not reveal remarkable effects of E2 treatment on thyroid tissue. Consequently, E2 was classified as being not active on the thyroid system given the analysis using the refined data interpretation criteria.

OECD Phase 3 - BP-2

The general performance of the two tests indicates that the BP-2 studies in Phase 3 were valid. Some mortality occurred in the highest concentration of the BP-2 study in lab 2, however it did not exceed what is considered an acceptable amount of mortality for validity (5%), and therefore was included in the analysis. No other evidence of overt toxicity was reported. Treatment with BP-2 did not result in advanced or asynchronous development. Histological analysis of the thyroid glands revealed strong evidence of thyroid antagonism in lab 2, but only mild changes in lab 1. Consequently, BP-2 was classified as being thyroid active in lab 2. Results from lab 1 are suggestive of thyroid

activity, however this result would likely need other supportive evidence to deem this chemical thyroid active.

No significant changes in assay outcomes occurred when using the refined data interpretation criteria as compared to the initial data interpretation criteria discussed in section 3.6. One potential disadvantage of using this scheme is that it dismisses the possibility that an unknown chemical may cause developmental delay through thyroid-mediated mechanisms, but not cause significant histological changes in the thyroid glands. Therefore, it is recommended that, in combination with the proposed scheme, a common sense weight-of-evidence approach be applied to every unknown chemical.

8.3 Strengths and limitations of the assay

The Amphibian Metamorphosis Assay provides for the ability to screen for chemicals that have the potential to interact with the HPT axis, and therefore fills a gap in current testing paradigms for endocrine activity. As with any toxicological assay, there are strengths and limitations of the systems which must be recognized and considered when evaluating data and drawing conclusions. Significant advantages of the AMA are:

1. Amphibian metamorphosis is a well-studied developmental process that is dependent on thyroid hormone, thus effects on metamorphic development are relatively specific indicators of HPT axis perturbation;
2. It is the only thyroid-specific assay in the EDSP/OECD which is performed on an animal undergoing morphological development;
3. It is an *in vivo* assay allowing for evaluation of effects from parent compounds and metabolic degradates;
4. The combination of morphological endpoints and histological endpoints allow for evaluation of substances that can potentially interfere with the HPT axis through both central homeostatic mechanisms (i.e., hypothalamus, pituitary, thyroid gland) *and* through other regulatory mechanisms operating in peripheral (e.g. deiodinases) tissues;
5. It incorporates a standard, easily acquired laboratory model species, *Xenopus* and utilizes common aquatic toxicology methods;
6. Peer-reviewed protocols for standardized histological assessments of *Xenopus* thyroid glands are available;
7. It is reasonably rapid, straightforward, and cost effective.
8. This model is relevant to other taxa when conserved elements of the HPT axis are considered. For example, several studies have demonstrated that TPO and NIS inhibitors that effect TH homeostasis in mammals also effect similar endocrinological changes in *Xenopus*.
9. It provides toxicological data in a taxon (amphibians) underrepresented in available Agency protocols.

While there are many strengths of the AMA, there are also several limitations of the assay, both technical and those that are due to general lack of knowledge, which should be considered.

1. Relevance of this model to other taxa, including mammals, has not been fully established with regard to all possible mechanisms involved in HPT disruption. This is particularly true with regard to compounds that disrupt TH function and regulation in peripheral mechanisms and tissues, including hepatic metabolism, local TH regulation, and TH transport in blood. Amphibian transthyretins (TTRs), for example, are T3 binding proteins, whereas mammalian TTRs are T4 binding proteins. Thus, effects of compounds that disrupt binding of TH to TTR may differ between amphibians and mammals. This could also be true for other vertebrate taxa, and for other thyroid hormone binding proteins. A document comparing the approaches for detecting thyroid-active compounds is currently being composed and will be available in early 2008;
2. While the amphibian thyroid system is regulated, and responds in a similar manner to perturbation as that of other tetrapods, including mammals, it should be noted that the thyroid system of a developing animal may be regulated differently from that of the juvenile or adult. As such, the regulatory relationships among the different organs and the actions of hormones on tissues (morphological and gene expression) vary among stages of the life cycle. Thus, the morphological or molecular signatures of perturbation of the axis in a tadpole can differ markedly from those of the adult animal;
3. Sensitivity of this assay has not been fully characterized, especially as it relates to detection of weakly active substances;
4. Like most aquatic toxicology methods that rely on testing aqueous concentrations of the test chemical, there are inherent technical difficulties testing substances that are poorly soluble in water in aquatic systems, and methods for delivering such substances to the test system have not been fully explored nor compared. This issue has been generally addressed in the OECD Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (35), however not specifically in terms of validating a method. The AMA requires an aqueous exposure protocol whereby test chemical is introduced into the test chambers via a flow-through system or through static renewal, and these approaches have not been directly compared. Additionally, prior to using this method, baseline information about the chemical must be obtained to determine testability and the best exposure method. Common characteristics which indicate that the chemical may not be testable include: high octanol water partitioning coefficients (log K_{ow}), high volatility, susceptibility to hydrolysis, and susceptibility to photolysis under ambient laboratory lighting conditions;
5. Non-thyroidal toxicities have the potential to affect some of the morphological endpoints of the assay. For example, reductions in growth, as determined by

wet weight and/or organism length, can be affected by non-thyroidal toxicity. Apparent increases in growth may also be observed with compounds that negatively effect normal development. Clearly, the use of a multiple endpoint, weight-of–evidence approach is necessary to rule out non-thyroidal toxicities while taking into account overt evidence of such toxicities such as acute mortality, morbidity, hemorrhagic lesions, lethargy, and neurological abnormalities. However, further characterization of effects of non-thyroidal toxicities on morphological endpoints may be necessary;

6. A potential problem with the histological analysis used in this assay is that observations made after a fixed exposure period are conducted on organisms from different treatment groups that may have progressed to different developmental stages. Analysis of the thyroid glands of individuals in different developmental stages should be done in the context of what is normal for different developmental stages.

Importantly, it should be recognized that the limitations listed above represent gaps in our knowledge that have been identified during the course of developing and refining the Amphibian Metamorphosis Assay, and that similar gaps exist in the development and validation of any assay. These gaps in knowledge represent an understanding that additional advantages or limitations of the assay likely exist, but are not yet defined.

9 REFERENCES

1. EDSTAC. Final Report from the Endocrine Disruptor Screening and Testing Advisory Committee, Volums I-II. 1998.
2. FRN. Federal Register Notice. 71542. 1998.
3. SAB, SAP. Review of the Endocrine Disruptor Screening Program by a Joint Subcommittee of the Science Advisory Board and Scientific Advisory Panel. EPA-SAB-EC-99-013. 1999.
4. FRN. Federal Register Notice. 71542. 1998.
5. NIEHS. 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. 1997. Research Triangle Park, NC.
6. OECD. Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. 1996.
7. OECD. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. 34. 2005.
8. EDSP. Validation of Assays in the EDSP. 2005.
9. Battelle. Final Detailed Review Paper for Amphibian Metamorphosis Assay. 2004.
10. OECD. Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances. 46. 2004.
11. U.S.EPA. U.S. EPA Peer Review Handbook. EPA/100/B-06/002. 2006.
12. OECD. Detailed Review Paper on Thyroid Hormone Disruption Assays. 2006.
13. Zoeller RT, Tan SW. Implications of research on assays to characterize thyroid toxicants. Crit Rev Toxicol 2007;195-210.

14. Shi YB. Amphibian Metamorphosis: From Morphology to Molecular Biology. New York, NY: Wiley & Sons, 2000.
15. Denver RJ, Glennemeier KA, Boorse GC. Endocrinology of complex lifecycles: Amphibians. In Hormones, Brain and Behavior. Elsevier Science, 2002;469-513.
16. Dent JN. Survey of Amphibian Metamorphosis. In: Etkin W, Gilbert LI, eds. Metamorphosis: A Problem in Developmental Biology. New York: Appleton-Crofts, 1968;271-311.
17. Just JJ, Kraust-Just J, Check DA. Survey of Chordate Metamorphosis. In: Gilbert LI, Frieden E, eds. Metamorphosis: A Problem in Developmental Biology. New York: Plenum Press, 1981.
18. Tata JR. Gene expression during metamorphosis: an ideal model for post-embryonic development. Bioessays 1993;239-48.
19. Tata JR. Amphibian metamorphosis as a model for studying the developmental actions of thyroid hormone. Ann Endocrinol (Paris) 1998;433-42.
20. Dodd MHI, Dodd JM. The Biology of Metamorphosis. In: Lofts B, ed. Physiology of Amphibia. New York: Academic Press, 1976;467-599.
21. Etkin W. Metamorphosis. A Problem in Developmental Biology. New York: Appleton, 1964.
22. Etkin W. Hormonal Control of Amphibian Metamorphosis. In: Etkin W, Gilbert LI, eds. Metamorphosis, a Problem in Developmental Biology. New York: Appleton, 1968.
23. Nieuwkoop PD, Faber J. Normal Table of *Xenopus Laevis*. New York: Garland Publishing, 1994.
24. Tata JR. Autoinduction of nuclear hormone receptors during metamorphosis and its significance. Insect Biochem Mol Biol 2000;645-51.
25. Leloup J, Buscaglia M. La triiodothyronine: Hormone de la metamorphose des amphibiens. C.R.Acad Sci 1977;2261-3.
26. Shi Y-B, Wong J, Puzianowska-Kuznicka M, Stolow MA. Tadpole competence and tissue-specific temporal regulation of amphibian metamorphosis: Roles of thyroid hormone and its receptors. Bioessays 1996;391-9.
27. Denver RJ. Proximate mechanisms of phenotypic plasticity in amphibian metamorphosis. Am Zool 1997;172-84.
28. Denver RJ. Hormonal correlates of environmentally induced metamorphosis in the Western spadefoot toad, *Scaphiopus hammondii*. Gen Comp Endocrinol 1998;326-36.
29. Hayes T, Chan R, Licht P. Interactions of temperature and steroids on larval growth, development, and metamorphosis in a toad (*Bufo boreas*). J Exp Zool 1993;206-15.
30. Tata JR. Hormonal regulation of metamorphosis. Symp Soc Exp Biol 1972;163-81.
31. OECD. Report of the Validation of the Amphibian Metamorphosis Assay for the detection of thyroid active substances: Phase 1 – Optimisation of the Test Protocol. 2004.
32. OECD. Amphibian Metamorphosis Assay Histopathology Part 2 - Approach to reading studies, diagnostic criteria, severity grading and atlas. 2007.

33. OECD. Amphibian Metamorphosis Assay: Part 1 - Technical guidance for morphologic sampling and histological preparation. 2007.
34. Kahl MD, Russom CL, DeFoe DL, Hammermeister DE. Saturation units for use in aquatic bioassays. Chemosphere 1999;539-51.
35. OECD. Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. 23. 2000.
36. OECD. Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. No. 54, ENV/JM/MONO(2006)18. 2006. OECD Series on Testing and Assessment.
37. Krain LP, Denver RJ. Developmental expression and hormonal regulation of glucocorticoid and thyroid hormone receptors during metamorphosis in *Xenopus laevis*. J Endocrinol 2004;91-104.
38. Regard E. Cytophysiology of the amphibian thyroid gland through larval development and metamorphosis. Int Rev Cytol 1978;81-118.
39. Buckbinder L, Brown DD. Expression of the *Xenopus laevis* prolactin and thyrotropin genes during metamorphosis. Proc Natl Acad Sci 1993;3820-4.
40. Manzon RG, Denver RJ. Regulation of pituitary thyrotropin gene expression during *Xenopus* metamorphosis: negative feedback is functional throughout metamorphosis. J Endocrinol 2004;273-85.
41. Degitz SJ, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, Tietge JE. Progress towards development of an amphibian-based thyroid screening assay using *Xenopus laevis*. Organismal and thyroidal responses to the model compounds 6-propylthiouracil, methimazole, and thyroxine. Toxicol Sci 2005;353-64.
42. Opitz R, Hartmann S, Blank T, Braunbeck T, Lutz I, Kloas W. Evaluation of histological and molecular endpoints for enhanced detection of thyroid system disruption in *Xenopus laevis* tadpoles. Toxicol Sci 2006;337-48.
43. Tietge JE, Holcombe GW, Flynn KM, Kosian PA, Korte JJ, Anderson LE, Wolf DC, Degitz SJ. Metamorphic inhibition of *Xenopus laevis* by sodium perchlorate: effects on development and thyroid histology. Environ Toxicol Chem 2005;926-33.
44. Goleman WL, Urquidi LJ, Anderson TA, Smith EE, Kendall RJ, Carr JA. Environmentally relevant concentrations of ammonium perchlorate inhibit development and metamorphosis in *Xenopus laevis*. Environ Toxicol Chem 2002;424-30.
45. Becker KB, Stephens KC, Davey JC, Schneider MJ, Galton VA. The type 2 and type 3 iodothyronine deiodinases play important roles in coordinating development in *Rana catesbeiana* tadpoles. Endocrinol 1997;2989-97.
46. Goleman WL, Carr JA. Contribution of ammonium ions to the lethality and antimetamorphic effects of ammonium perchlorate. Environ Toxicol Chem 2006;1060-7.
47. Hu F, Sharma B, Mukhi S, Patino R, Carr JA. The colloidal thyroxine (T4) ring as a novel biomarker of perchlorate exposure in the African clawed frog *Xenopus laevis*. Toxicol Sci 2006;268-77.
48. Brown DD. The role of deiodinases in amphibian metamorphosis. Thyroid 2005;815-21.

49. Brown DD, Cai L, Das B, Marsh-Armstrong N, Schreiber AM, Juste R. Thyroid hormone controls multiple independent programs required for limb development in *Xenopus laevis* metamorphosis. Proc Natl Acad Sci U.S.A 2005;12455-8.
50. Cai L, D.D.Brown. Expression of type II iodothyronine deiodinase marks the time that a tissue responds to thyroid hormone-induced metamophosis in *Xenopus laevis*. Dev Biol 2004;87-95.
51. Brucker-Davis F. Effects of environmental synthetic chemicals on thyroid function. Thyroid 1998;827-56.
52. Jarry H, Christoffel J, Rimoldi G, Koch L, Wuttke W. Multi-organic endocrine disrupting activity of the UV screen benzophenone 2 (BP2) in ovariectomized adult rats after 5 days treatment. Toxicol 2004;87-93.
53. Seidlova-Wuttke D, Jarry H, Christoffel J, Rimoldi G, Wuttke W. Effects of bisphenol-A (BPA), dibutylphtalate (DBP), benzophenone-2 (BP2), procymidone (Proc), and linurone (Lin) on fat tissue, a variety of hormones and metabolic parameters: a 3 months comparison with effects of estradiol (E2) in ovariectomized (ovx) rats. Toxicol 2005;13-24.
54. Opitz R, Trubiroha A, Lorenz C, Lutz I, Hartmann S, Blank T, Braunbeck T, Kloas W. Expression of sodium-iodide symporter mRNA in the thyroid gland of *Xenopus laevis* tadpoles: developmental expression, effects of antithyroidal compounds, and regulation by TSH. J Endocrinol 2006;157-70.
55. Opitz R, Hartmann S, Blank T, Braunbeck T, Lutz I, Kloas W. Evaluation of histological and molecular endpoints for enhanced detection of thyroid system disruption in *Xenopus laevis* tadpoles. Toxicol Sci 2006;337-48.
56. Opitz R, Braunbeck T, Bogi C, Pickford DB, Nentwig G, Oehlmann J, Tooi O, Lutz I, Kloas W. Description and initial evaluation of a *Xenopus* metamorphosis assay for detection of thyroid system-disrupting activities of environmental compounds. Environ Toxicol Chem 2005;653-64.
57. Carlsson G, Norrgren L. The impact of the goitrogen 6-propylthiouracil (PTU) on West-African clawed frog (*Xenopus tropicalis*) exposed during metamorphosis. Aquat Toxicol 2007;55-62.
58. Mitsui N, Fujii T, Miyahara M, Oka T, Kashiwagi A, Kashiwagi K, Hanada H, Urushitani H, Santo N, Tooi O, Iguchi T. Development of metamorphosis assay using *Silurana tropicalis* for the detection of thyroid system-disrupting chemicals. Ecotoxicol Environ Saf 2006;281-7.
59. Opitz R, Hartmann S, Blank T, Braunbeck T, Lutz I, Kloas W. Evaluation of histological and molecular endpoints for enhanced detection of thyroid system disruption in *Xenopus laevis* tadpoles. Toxicol Sci 2006;337-48.
60. Gutleb AC, Schriks M, Mossink L, Berg JH, Murk AJ. A synchronized amphibian metamorphosis assay as an improved tool to detect thyroid hormone disturbance by endocrine disruptors and apolar sediment extracts. Chemosphere 2007.